SAPIENZA
UNIVERSITÀ DI ROMA

Bias correction of survey-data on household wealth components

Dottorato in Scuola di Scienze Statistiche
Curriculum Methodological Statistics (XXXVII cycle)

Matteo Spuri
ID number 1768787

Advisor
Prof. Pierpaolo Brutti

Academic Year 2024/2025

Thesis defended on 27/01/2026
in front of a Board of Examiners composed by:

Prof. Pierfrancesco Alaimo Di Loro (chairman, LUMSA University)
Prof.ssa Maura Mezzetti (University of Rome Tor Vergata)
Prof.ssa Daniela Marella (Sapienza University of Rome)

Bias correction of survey-data on household wealth components
PhD thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: matteo.spuri@uniroma1.it

Abstract

Recent initiatives integrate household surveys with supervisory and national accounts aggregates to produce timely, reliable distributional statistics on household income and wealth. This thesis develops a framework that aligns Italian survey microdata with supervisory targets for deposits, bonds, listed shares, and investment fund shares, with the objective of delivering Distributional Wealth Accounts that are reconciled at both the aggregate and distributional levels.

Three building blocks precede the main calibration. First, we develop methods to estimate instrument-specific target distributions from aggregates, motivate Pareto-lognormal splice margins over simpler lognormal forms, provide parameter-recovery procedures, and introduce a tail test to discriminate between the two. Second, we reconcile Banking Supervisory Reports (BSR) with the Household Finance and Consumption Survey (HFCS) by mapping deposit accounts to households and disentangling securities accounts reported as range sums, thereby placing administrative targets on a household footing. Third, we formalise a benchmark univariate calibration – already used operationally for deposits – that iteratively matches class totals with regularisation to limit class changes; we extend this benchmark to all four instruments as a reference.

Building on these elements, we introduce a distribution-aware univariate calibration with a composite L_2 loss that balances a Cramér-von Mises shape term on the empirical CDF and a relative cumulative-amount term, while keeping survey weights fixed. The method attains the benchmark's aggregate alignment and produces smoother, more credible within-class adjustments, substantially reducing boundary piling. We then extend the composite-loss calibration to the multivariate case to account explicitly for cross-instrument dependence. The multivariate layer maps adjusted values through parametric probability-integral transforms to Gaussian scores and aligns regime-specific correlations across instruments, while preserving the marginal fit; the composite loss retains the Cramér-von Mises and relative-amount terms and adds a dependence penalty on the Gaussian-score scale.

Applications show that the univariate implementation closes aggregate gaps to supervisory totals and reduces distributional divergence from target margins, and that the multivariate calibration maintains those gains while improving cross-instrument coherence and correcting the top tail more accurately. On the marginals and in terms of coverage, the multivariate procedure performs on a par with the univariate distribution-aware calibration, while concentration indicators decrease as tail alignment improves. End-to-end indicators remain stable after proportional allocation, indicating that distributional improvements survive the full DWA pipeline. The framework is modular and auditable, conditional on HFCS-BSR harmonisation and splice-based margins, and it provides a production-ready, distribution-aware alternative to current practice without altering survey weights.

Contents

Chapter 1

Introduction and Background on the Distributional Wealth Accounts Framework

1.1 Motivation and Policy Relevance

In recent years, several initiatives have been launched to combine household-level survey data with macroeconomic aggregates from national accounts to produce more reliable, timely, and policy-relevant statistics on the distribution of household income and wealth.[1] These efforts have led to the development of what are commonly referred to as "*Distributional National Accounts*". These systems integrate information from the System of National Accounts (SNA) – which provides comprehensive, coherent, and internationally comparable figures for the household sector at the macroeconomic level – with the information on the distribution of economic resources among households captured by survey data. They build on the latter to bridge differences and produce estimates of household income, wealth, and consumption that are consistent with national accounts aggregates.

The growing demand for such distributional statistics has been emphasised repeatedly at the international level. The G-20, for example, underscored the need for timely and granular data on income and wealth distribution alongside aggregate estimates, as part of its Data Gaps Initiative (DGI), to support inclusive growth and informed policymaking.[2] Distributional statistics enable policymakers to assess the impact of specific policies and shocks on different household groups, enhancing the targeting and effectiveness of economic and social interventions.

For central banks in particular, distributional data are essential to analyse the heterogeneous effects of monetary policy and inflation. Household wealth distribution

[1] Examples include the OECD Income and Wealth Distribution Database (IDD – WDD), the Distributional National Accounts (DINA) developed by the World Inequality Database, and the Distributional National Accounts (DNA) developed jointly by OECD and Eurostat.

[2] See G-20 Data Gaps Initiative Recommendations 8 and 9: https://www.imf.org/.../DGI/g20-dgi-recommendations.

influences monetary transmission channels, such as consumption responses, credit constraints, and asset price sensitivity, and has implications for both price stability and financial stability. The ECB's 2021 monetary policy strategy review explicitly recognised that the distribution of income and wealth affects how monetary policy is transmitted and may have side effects that influence inequality.[3] Recent analyses based on distributional wealth data have shown that inflation and monetary policy shocks can have markedly unequal effects across the income and wealth distribution (see, e.g., Infante et al., 2023; Kaplan et al., 2018; Bobasu et al., 2023). This empirical evidence reinforces the need for distributional statistics to support monetary policy formulation.

To address persistent gaps in the measurement of household wealth distribution, the European Central Bank has developed the Distributional Wealth Accounts (DWA)[4] - a harmonised set of experimental statistics providing quarterly estimates aligned with macroeconomic aggregates.

1.2 Overview of the Distributional Wealth Accounts Framework[5]

In 2015, the European Central Bank established an expert group with the mandate to understand, quantify, and explain the main differences between the Household Finance and Consumption Survey (HFCS) and the System of National Accounts, and to develop distributional information on household wealth. This initiative builds upon the earlier work of the ECB Expert Group on Linking Macro and Micro Data for the Household Sector (EG–LMM). In 2019, the ECB created the Expert Group on Distributional Financial Accounts (EG–DFA). As shown by Engel et al. (2022), this group designed and implemented an estimation method that produces the experimental quarterly Distributional Wealth Accounts for the euro area and for several member countries. The DWA were first published in January 2024. They are compiled quarterly and disseminated within five months after the reference quarter and are currently classified as *experimental statistics*. The DWA follow the European System of Accounts' quarterly frequency, ensuring consistency with the Quarterly Sector Accounts (QSA) framework.[6] The framework is designed to evolve, incorporating improvements in survey methodology and administrative data sources, subject to ECB validation.

[3] See ECB, *An overview of the ECB's monetary policy strategy*, https://www.ecb.europa.eu/.../strategy_overview.

[4] See ECB (2024), *Introducing the Distributional Wealth Accounts for the euro area*, ECB Economic Bulletin, Issue 5/2024, https://www.ecb.europa.eu/.../ecb.ebart202405_02.

[5] For a detailed and technical description of the DWA methodology, see European Central Bank (2024), *Distributional Wealth Accounts: Methodological Note*, available at https://data.ecb.europa.eu/.../DWA_Methodological_note.pdf.

[6] See the ECB DWA dataset portal: https://data.ecb.europa.eu/data/datasets/DWA/data-information.

1.2.1 Differences between Household Finance and Consumption Survey and Quarterly Sector Accounts

Although both HFCS and QSA measure household wealth, they differ substantially due to their distinct objectives and data-collection frameworks.[7] Two broad sources of discrepancy can be identified: conceptual differences and measurement and reporting errors.

Conceptual differences

Conceptual differences can be grouped into four categories:

Population In the HFCS, the target population comprises private households, defined as "*a person living alone or a group of people who live together in the same private dwelling and share expenditures, including the joint provision of the essentials of living*", meaning institutional households are excluded. Conversely, in the QSA, the household sector covers the entire resident population, defining a unit as resident in a given economic territory if it maintains the strongest economic connection there.

Valuation In the HFCS, assets and liabilities are self-reported and may not be aligned with market values, especially during periods of rapid price change. In the QSA, all financial and non-financial instruments are valued at market prices.

Timing HFCS data are collected over several months, whereas QSA data refer to specific quarter-end dates. To harmonise timing, the QSA quarter closest to the midpoint of the HFCS fieldwork period is selected, unless the HFCS explicitly uses a single reference date.

Coverage Coverage differences arise both from the omission of certain instruments in the HFCS and from divergent terminology between HFCS and QSA.

These conceptual gaps require ad hoc adjustments, which contribute to the overall misalignment between HFCS totals and their corresponding macroeconomic figures. They are addressed in the reconciliation procedure discussed in Section 1.2.2. In practice, conceptual inconsistencies and measurement issues are closely interrelated and are handled jointly during reconciliation and adjustment. Consequently, the estimation process begins by reconciling definitions.

Measurement error and reporting biases

Discrepancies between HFCS and QSA also reflect the measurement issues of both sources.[8]

[7] See European Central Bank (2020), *Report on Linking Macro and Micro Data for the Household Sector*, Statistics Paper Series No. 37, https://www.ecb.europa.eu/.../ecb.sps37.

[8] See Section 2.1, "Measurement issues", in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

The HFCS faces the typical problems of household surveys: sampling variance, non-response bias, and reporting bias. Sampling variance stems from the probabilistic design and is usually assessed through confidence intervals that consider both sampling and imputation.

Unit non-response is systematically higher among the very wealthy. Weight adjustments can partially correct this, but oversampling of affluent households remains necessary to capture the top tail of the distribution. Item non-response occurs when households fail to provide answers to specific questions. In such cases, the HFCS imputes missing values to yield a complete dataset for all balance-sheet and income variables.[9]

Reporting bias arises when households provide systematically inaccurate values, either through intentional misreporting or recall limitations. This can generate under-reporting (households declare less than they hold) or zero-reporting (households mistakenly report no holdings of an instrument).

The QSA aggregates administrative and financial sources. Compilation errors may arise because underlying statistics follow concepts different from ESA 2010 or because the gap between source data and QSA concepts is misestimated. Most QSA household-sector data rely on counterpart reporting from financial institutions such as banks and pension funds. While generally high quality, these sources can misallocate holdings across sectors (e.g., resident vs non-resident misclassification).[10]

Certain asset types in the QSA – such as unlisted equity and non-financial assets like land – remain difficult to measure due to infrequent or incomplete reporting. Residual balancing adjustments used to preserve sectoral consistency may also allocate residuals to the household sector when other sectors lack direct data, particularly for unlisted shares where valuation and attribution are imprecise.

General impact and implications of the discrepancies

Because QSA totals typically exceed HFCS totals for comparable instruments, adjustments to HFCS microdata raise aggregate asset values and therefore affect inequality estimates. The direction of change depends on the adjustment type:

- Adjustments addressing the under-representation of the wealthy increase inequality measures;
- Proportional-allocation adjustments reduce inequality when applied to broadly held assets (e.g., housing, deposits) and increase it for concentrated assets (e.g., listed shares).

[9] The HFCS imputation procedure combines hot-deck and model-based techniques to ensure internal consistency and coherence across variables. See Section 2.1 “Measurement issues” in the DWA Methodological Note https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

[10] For example, a deposit may be misclassified by a reporting bank as belonging to a resident household when it actually belongs to a non-resident. See Section 2.1 “Measurement issues” in the DWA Methodological Note https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

1.2.2 The reconciliation procedure

The DWA estimation process comprises three main stages. First, wealth concepts are reconciled to resolve definitional inconsistencies between the HFCS and QSA, enabling direct comparison across wealth components. Second, statistical adjustments align survey data with macroeconomic figures. Third, interpolation and extrapolation techniques yield quarterly distributional estimates up to the latest QSA quarter. These steps are summarised in Figure 1.1.

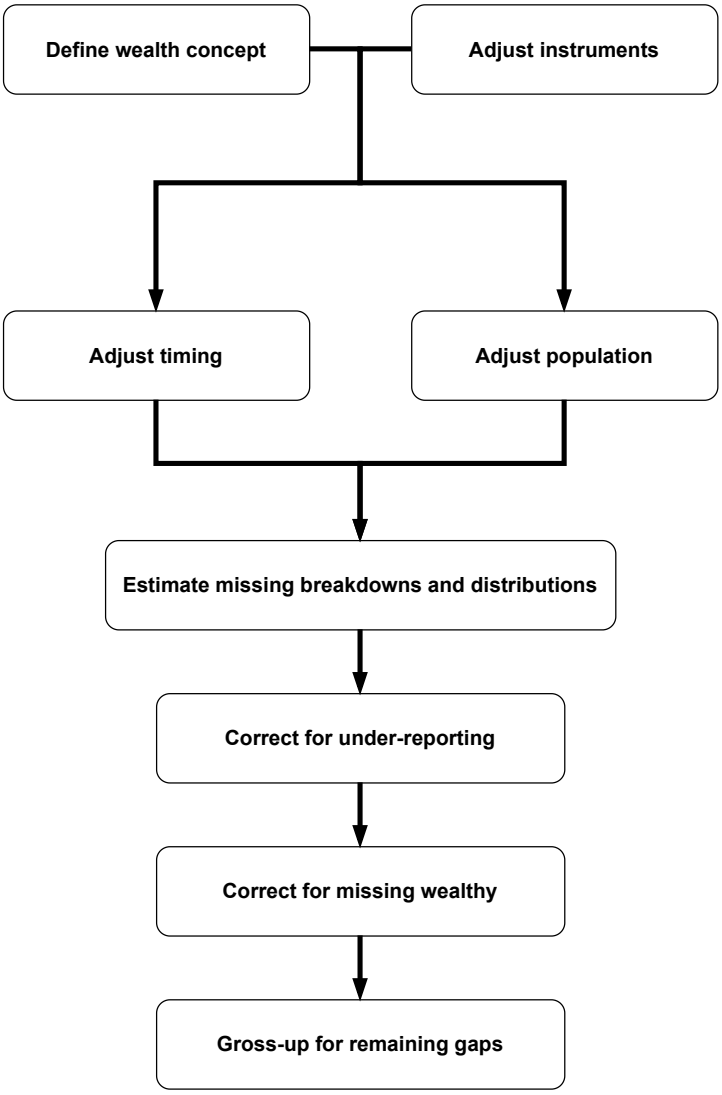

Figure 1.1. Schematic representation of the DWA methodology.
Notes: Arrows indicate processing order. The top panels reconcile concepts and instruments; the middle panels align timing and population; the bottom block implements distributional estimation and gap-closing adjustments. Adapted from the *DWA methodological note*.

Wealth concepts reconciliation and coverage

The first step of the DWA procedure consists of the reconciliation of definitions between the HFCS and the QSA. As mentioned above, some concepts have slightly different definitions, thus requiring some adjustments in order to be comparable, and some components are not collected directly in the survey. Nonetheless, the adjusted wealth concepts included in the DWA capture around 90 per cent of the total household assets and liabilities as recorded in the wealth concept of the national accounts. Table 1.1 shows the instruments included and excluded by asset category and liability, as well as their relative importance on total assets and liabilities. Efforts are ongoing to incorporate some of the excluded categories, namely currency holdings, private pension entitlements, and non-life insurance technical reserves.

Table 1.1. Financial and non-financial wealth in QSA and their representation in DWA (euro area, mid–2023).

Code	Instruments (QSA terminology)	Assets		Liabilities	
		Included in DWA	Not included	Included in DWA	Not included
F21	Banknotes and coins		1%		
F2M	Deposits	13%			
F3	Debt securities	1%			
F4	Loans		0%	88%	
F5	Equity	12%			
F62	Life insurance entitlements	6%			
F6O	Non-life insurance reserves		0%		
F6M	Pension entitlements[a]		5%		
F8	Trade credits and other accounts receivable		1%		12%
NUN	Housing wealth[b]	54%			
NUN	Business assets[c]	7%			
Sub-total		**93%**	**7%**	**88%**	**12%**
	Total wealth	**100%**		**100%**	

Notes: Shares refer to the QSA wealth concept (euro area, mid–2023). This table reproduces Table 1 in the *DWA Methodological Note*.
[a]Pension claims (excluding social security pensions), generally involving pay-as-you-go systems.
[b]Dwellings and land underlying dwellings, in QSA terminology.
[c]Fixed assets minus dwellings and land underlying dwellings.

In addition, the DWA applies a population adjustment to reconcile differences in coverage: the QSA includes institutional households (such as those in retirement or long-term care facilities) while the HFCS does not. To ensure comparability, the QSA aggregates are adjusted downward proportionally.[11]

[11]See Section 2.2.1 in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note.pdf.

Although several instruments can be directly mapped between HFCS and QSA, others require substantial adjustments to account for classification differences or data limitations. These adjustments address classification inconsistencies and gaps in the scope of available micro and macro data. While these adjustments are largely technical, they are essential for ensuring consistency between household-level micro-data and macroeconomic financial accounts and, therefore, form the foundation on which distributional estimates are built.

Concerning business wealth, the HFCS and the QSA differ substantially. In the HFCS, business wealth is split into self-employment and non-self-employment components, defining the former as businesses in which a household member is either self-employed or has an active role (e.g., like a managing partner), and the latter as "passive' investment only. In the QSA, however, there is no stand-alone "business wealth" concept. Instead, businesses owned by households are classified either as corporations or quasi-corporations (i.e. unincorporated enterprises functioning like corporations) or as part of the household sector, depending on specific national criteria.

Depending on this classification, the mapping of business wealth between HFCS and QSA differs: if businesses owned by the household falls under the first QSA group (i.e., corporations or quasi-corporations), assets identified in the HFCS as business wealth would be recorded as unlisted shares or other equity; if it falls under the second QSA group (i.e., not recognised separately from the household), the assets identified as business wealth in the HFCS are shown in the QSA as financial and non-financial assets of households and are not distinguishable from non-business assets and liabilities of the household in its function as a consumer.[12]

To reconcile these differences, the first step is to separate assets included in self-employed business wealth in the HFCS, which corresponds to the QSA aggregated item "unlisted shares and other equity". This is first achieved by assuming that all legal forms other than sole proprietors and partnerships in the HFCS issue "unlisted shares and other equity", then by applying country-specific criteria to the categories of sole proprietors and partnerships in the HFCS to make a distinction between those falling under the QSA definition of households or under the definition of corporation/quasi-corporation.

Concerning housing wealth, the QSA do not contain directly a variable for this instrument. However, it can be derived by summing the recorded concepts of dwellings and land underlying dwellings.[13]

Regarding mortgage and other debt, QSA data in most compiling countries do not provide a breakdown between mortgage debt and non-mortgage debt. To derive this division, the Monetary Financial Institutions (MFIs) balance sheet statis-

[12]Moreover, the QSA record this wealth in gross terms, while the underlying concept of the HFCS is to record the expected value of the business if sold, which means at net value.

[13]See Section 2.2.5 "Adjustments to sources on housing wealth and business wealth", in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note.pdf.

tics are used, which provide the distinction of loans granted to households by MFIs.[14]

After reconciliation, it is possible to make a preliminary comparison between the aggregates derived from the HFCS and their corresponding macroeconomic figures. Figure 1.2 shows the coverage, namely the ratio of HFCS totals to QSA totals, of the instruments included in this thesis for the Italian data of the five available HFCS waves.

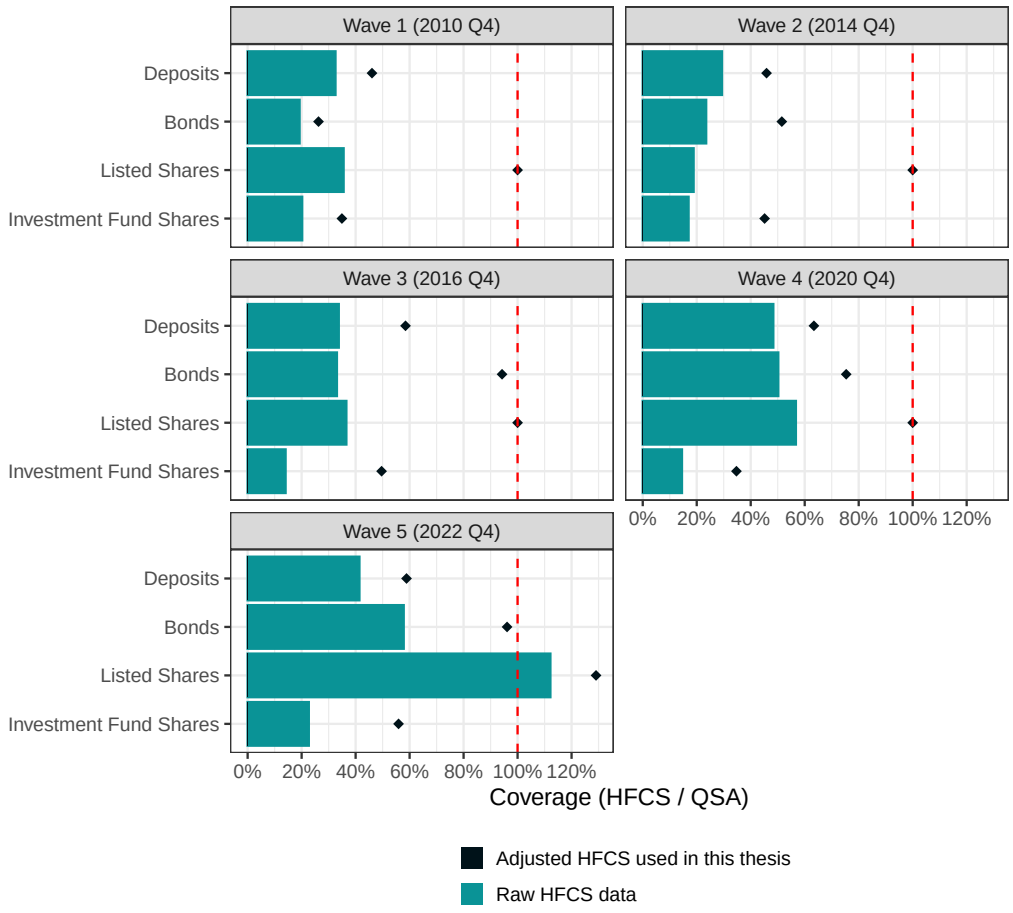

Figure 1.2. Coverage gap of wealth components (Italy).

Notes: Coverage is defined as the ratio of HFCS totals to QSA totals (HFCS/QSA). Bars report the main instruments included in the DWA; the horizontal dashed line marks full coverage (100%). HFCS amounts are survey-weighted; QSA are the national accounts totals for the household sector.

The latest releases show major improvements in coverage, thanks to refinements in the sampling plan from 2020 onwards, such as adjustments to the oversampling of wealthy households and improvements in regional stratification (see Loschiavo et al., 2024). However, some instruments still display a great difference between

[14] See ECB (2023), *MFI balance sheet statistics - methodological notes*, available at https://www.ecb.europa.eu/.../mfi_balance_sheets.

survey-based and macro-aggregate totals. Some of these differences are addressed in the methodological steps.

Reconciling coverage gaps through classification and adjustments

The second step of the reconciliation procedure addresses the residual gap between survey-based aggregates and macroeconomic figures, after adjustments for timing, population, and conceptual differences have been applied. This phase represents the methodological core of the DWA procedure from a statistical point of view. It is also the stage at which participating countries have the greatest flexibility, as they may exploit country-specific data and integrate alternative or additional solutions within the general methodological framework. Any such enhancement must receive prior approval from the ECB Expert Group before being incorporated into the national DWA compilation processes, to ensure that overall cross-country comparability is preserved.

The ECB's reference methodology comprises three key adjustment steps: (*i*) the correction for the under-reporting and zero-reporting of deposits; (*ii*) the integration of additional information and the subsequent estimation of a Pareto distribution to address the under-representation of the wealthiest households; and (*iii*) a final proportional allocation procedure that addresses the remaining gaps and ensures full consistency between survey-based and macroeconomic totals.

Deposits correction As shown in Figure 1.2, deposits are greatly under-estimated across all surveys. This holds not only for Italy, but also for the other participating countries, as under-reporting of deposits is a widespread issue in the HFCS. Combined with their relative importance, as they constitute the largest share of households' financial wealth, the correction of this instrument represents a key step in the DWA reconciliation procedure.

This discrepancy is typically attributed to several factors, such as limited knowledge of the respondent, reluctance to disclose financial holdings, and timing differences between survey and national accounts reference dates. To address this, the DWA methodology applies targeted adjustments to the HFCS data by identifying implausible values of deposits using three diagnostic criteria:[15]

- Households are unlikely to hold deposits significantly below their monthly income;
- At least some of the household portfolio should be in the form of deposits, even if the share is very small;
- Self-employed households report business wealth in net terms and therefore may omit business-related deposits; the QSA, by contrast, includes them.

Based on these criteria, a set of rules is determined to detect outliers:

[15] See Section 2.2.6 "Corrections for under-reporting of deposits" in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

Income criterion: Outliers are defined as households with deposits of less than 10% of their monthly income. However, households with an annual gross income below €10,000 or with credit card debt are excluded from adjustment. Among those subject to adjustment, if deposits fall below 1% of their monthly income, the reported amount is replaced by the average of households within €2,500 of the outlier's annual income.

Assets criterion: Outliers are defined as households whose portfolio consists of less than 1% of deposits. Adjustments are not applied to households holding mortgages or credit card debt.

Self-employment criterion: Self-employed households are compared to employed households with similar income (within €2,500 annually). If their deposits fall below the average of the comparable employed group, the reported amount is replaced by that average.

As households can meet more than one criterion at the same time, only the adjustment that results in the largest increase is applied.

Pareto estimation Household surveys, including the HFCS, typically struggle to capture the wealthiest households adequately. These households are either insufficiently represented in survey samples or systematically refuse to participate (unit non-response). Although oversampling strategies have been introduced to mitigate this issue,[16] the upper tail of the wealth distribution remains largely under-covered in most countries. To address this, the DWA introduces a specific adjustment by fitting a Pareto distribution to the observed upper net wealth tail and simulating synthetic households from this estimated distribution. The assumption that the top tail of net wealth follows a Pareto distribution is widely supported in the literature. It constitutes the key hypothesis of this step (see Section 3.2 for detailed discussion).[17]

The process is implemented in four main steps. First, the shape parameter of the Pareto distribution is estimated using HFCS data above a threshold wealth level and, where available, additional information on the wealthiest households from external sources.[18] Studies, such as Bach et al. (2019), have shown that even the inclusion of a small number of wealthy observations can considerably improve the precision of Pareto estimates.

Second, the interval to be filled with the estimated rich households, usually referred to as "synthetic households", and the number of required additional observations are determined. The wealth distribution is divided into three intervals: (i) a

[16] See, for example, the overall increase in the coverage in Figure 1.2, which was mainly due to the implementation of a different strategy to capture the wealthiest households.

[17] See Section 2.2.7 "Adding rich households" in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

[18] These datasets are typically referred to as "rich list". For many countries, including Italy, the Forbes World's Billionaire List is used. Where better-quality data are available, DWA uses them.

lower range well covered by HFCS; (*ii*) a gap poorly captured by either the HFCS or external sources; and (*iii*) a top range covered by external "rich list" data. The width of the second gap depends on how effectively the HFCS captures the top tail. Figure 1.3 provides a schematic illustration (not to scale) of these intervals.

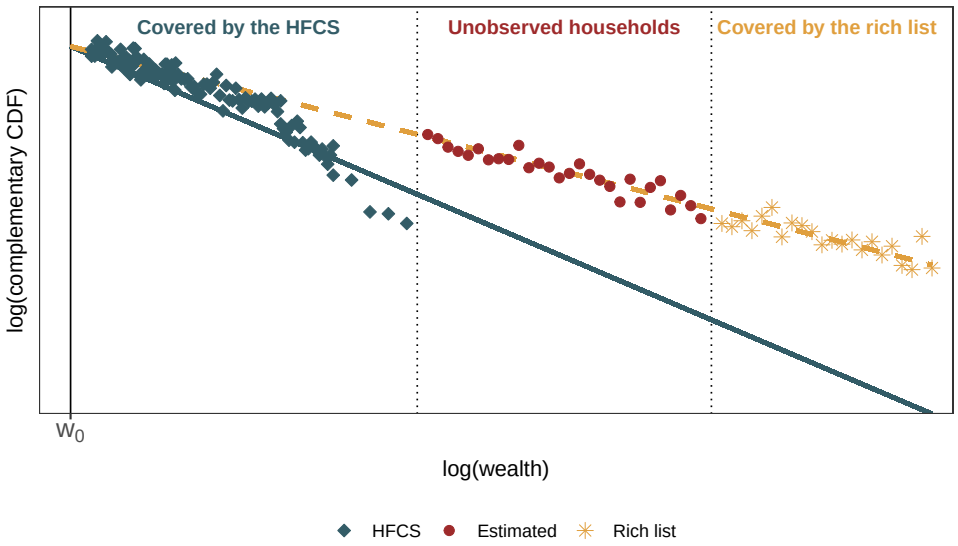

Figure 1.3. Stylised representation of the upper tail of the net wealth distribution. Notes: This figure is schematic (not to scale). Axes are on a log-log scale for wealth and the complementary CDF; axis tick values are intentionally omitted to avoid implying specific monetary magnitudes. The distribution is partitioned into three intervals: covered by HFCS; unobserved households (filled by synthetic households); and covered by rich-list data. Vertical line marks the HFCS imputation threshold w_0. In the Italian application, w_0 = EUR 1 000 000. Adapted from the DWA methodological note.

To estimate the total number of observations across the three intervals, denoted $\hat{m}$, we rely on the number of HFCS households in the first interval, $\hat{m}_1$, and the probability of falling within that interval, $P(x \in I_1)$, assuming all the observations are independent and identically distributed:

$$\begin{aligned}\hat{m}_1 &= E\left[\sum_{i=1}^{\hat{m}} \mathbb{1}_{(x_i \in I_1)}\right] \\ &= \hat{m} \cdot E\left[\mathbb{1}_{(x_i \in I_1)}\right] \\ &= \hat{m} \cdot P\left(x \in I_1\right)\end{aligned}$$

Rearranging gives:

$$\hat{m} = \frac{\hat{m}_1}{P(x \in I_1)}$$

The number of synthetic households to add in the second interval is then:

$$\hat{m}_2 = \hat{m} \cdot P(x \in I_2) = \frac{\hat{m}_1}{P(x \in I_1)} \cdot P(x \in I_2)$$

In the third step, $\hat{m}_2$ synthetic households are drawn using inverse transform sampling. Specifically, $\hat{m}_2$ draws are taken from a uniform distribution defined between the cumulative density function computed on the lower and upper bounds of the second interval. Then, the corresponding net wealth level is derived by applying the inverse of the Pareto cumulative distribution function on the $\hat{m}_2$ random draws.

The fourth and final step allocates liabilities and estimates the portfolio allocation of the synthetic households and the added rich. Since these households are given in terms of their net wealth, it is necessary to first assign liabilities based on country-specific debt-to-asset ratios, which typically range between 5 and 10 per cent for the richest deciles. Then, the total assets are calculated as the sum of the estimated net wealth and assigned liabilities from the previous step. Finally, the assets are split between the different instruments using results from a 2018 survey conducted by UBS/Campden ("Global Family Office Report 2018"),[19] which estimated average portfolios for over 300 households rich enough to have their wealth managed by a family office. This information, combined with the size of the gaps between QSA and HFCS for each instrument in each country, is used to estimate the allocation of the total assets of the wealthiest households. If national sources with better information exist, these may be used instead, subject to ECB approval.

Proportional allocation After applying deposit corrections and incorporating synthetic households through Pareto estimation, residual discrepancies often remain between the survey-based aggregates and the QSA totals. To ensure complete alignment with macroeconomic aggregates, the DWA applies a final adjustment based on proportional allocation.[20]

This approach ensures that, for each financial instrument, the sum of household holdings reported in the survey exactly matches the corresponding QSA total. It operates under the implicit assumption that under-reporting is proportionally distributed across all households for a given instrument and that the extent of it is uncorrelated across instruments. Let $h = 1, \ldots, H$ index households in the microdata used at this stage (including any synthetic households), where H denotes the corresponding total number of households, and let w_h denote the survey weights. Specifically, each household's holdings of instrument x are rescaled by a factor ς_x calculated from the inverse coverage ratio:

$$\varsigma_x = \frac{x^{QSA}}{\sum_{h=1}^{H} w_h \, x_h^{HFCS}},$$

and the new value x'_h is then computed to ensure the total of HFCS holdings of

[19] UBS & Campden Wealth (2018), *The Global Family Office Report 2018*. PDF available at https://www.ubs.com/.../gfo-report-2018-pdf-for-download-final.pdf.

[20] See Section 2.2.8 in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

instrument x equals the respective QSA total:

$$x'_h = \varsigma_x \, x_h^{HFCS},$$
$$x^{QSA} = \sum_{h=1}^{H} w_h x'_h.$$

Although the proportional allocation does not affect within-instrument inequality, since the relative positions of households are preserved, it does alter the joint distribution across instruments. Consequently, inequality indicators for total assets, total liabilities, and net wealth may change, even if those for individual instruments remain stable.

An alternative approach considered is multivariate calibration, which adjusts household weights rather than values by solving a system of constraints while minimising deviations from the original weights. This method was not adopted as the baseline due to several limitations. First, it tends to assign disproportionately large weights to wealthy households already corrected via the Pareto step, which can distort the overall wealth distribution.[21] Second, it implicitly assumes that under-reporting occurs uniformly across all instruments in a household's portfolio, an assumption that may not hold in practice. Third, convergence issues are common when dealing with skewed distributions or extreme initial weights, particularly in the upper tail of wealth.

Solutions based on instrument-specific calibration weights may alleviate some of these problems, but they require several auxiliary constraints to ensure consistency in net wealth distributions, which increases implementation complexity.

Moreover, altering survey weights to meet wealth targets can reduce the comparability of the survey population with the actual demographic structure of the country. Since HFCS weights are typically constructed to reproduce known population totals along demographic dimensions (e.g., age, sex, region), modifying them to correct for wealth aggregates may introduce biases in the population composition.

A specific modification is introduced for liabilities to avoid the artificial amplification of negative net wealth. In particular:

1. Proportional allocation is first applied to all liabilities.
2. If any household ends up with negative net wealth due to liabilities increasing more than assets, the liabilities are capped to match the increase in assets.
3. The remaining gap is reallocated only among households not affected by the cap.
4. The process is iterated until no further inconsistencies arise.

[21] See: *Estimating the Distribution of Household Wealth: Methods for Adjusting Survey Data Estimates Using National Accounts and Rich List Data*, Review of Income and Wealth, 2024. https://onlinelibrary.wiley.com/doi/10.1111/roiw.12657

This iterative adjustment ensures plausibility and coherence in the final net wealth distribution, particularly among households with high leverage or low asset holdings. This is especially important in the lower part of the net wealth distribution, where unadjusted liability increases may otherwise generate implausible negative wealth values.

Interpolation and Extrapolation of Quarterly Time Series

The third and final step of the DWA procedure transforms the reconciled and adjusted survey data into a quarterly time series aligned with the QSA timeline. Since the Household Finance and Consumption Survey (HFCS) is conducted only every few years, interpolation and extrapolation techniques are required to fill in the missing quarters and to extend the distributional estimates beyond the survey reference year.

Interpolation is applied to the period between two HFCS waves and involves estimating quarterly distributions by combining the structure of the latest available HFCS microdata with growth rates observed in the corresponding QSA aggregates. For each quarter, instrument-level totals from the QSA are distributed across households by preserving the relative ranking and shares observed in the HFCS data. This process assumes that, within each instrument, the distribution remains stable between waves except for the proportional growth observed at the macro level.

For extrapolation beyond the last available HFCS wave, the same structure-preserving logic is adopted. Shares from the last wave are projected forward using the growth dynamics of the QSA. Thus, both interpolation and extrapolation rest on the assumption that changes in wealth aggregates over time primarily affect levels rather than distributions. [22]

1.3 Objective and Structure of the Thesis

Although the Distributional Wealth Accounts represent a significant advancement in reconciling household-level survey data with macroeconomic wealth aggregates, several methodological challenges remain. One key limitation concerns the reliance on a harmonised, country-independent methodology. While such standardisation is crucial for ensuring cross-country comparability and fostering a coordinated improvement process, it may also inadvertently neglect relevant country-specific features. Notably, the DWA framework explicitly allows room for national enhancements, subject to ECB approval, recognising the potential value of such refinements.

These features include heterogeneity in misreporting behaviours that a uniform approach may poorly capture, the availability of national data sources that could substantially improve instrument-level correction (in both distributional structure and aggregate coverage), and refined methodologies that better exploit these sources.

[22] See Section 4 "Computing time-series" in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf.

Appropriately leveraging these national specificities through targeted solutions can enhance the accuracy and plausibility of the adjustment procedures, contributing to the methodological development of distributional wealth statistics and thereby enriching the rapidly evolving literature on the subject. Moreover, cross-country comparability is preserved, as the basic economic definitions and the standardised survey data remain the same.

This thesis aims to address these limitations, providing alternative adjustment strategies for selected wealth components within the DWA framework, with a focus on the Italian case and its available data sources, while maintaining overall comparability with the standard methodology.

In particular, the thesis is structured as follows: Chapter 2 describes the datasets used in the analysis and provides an overview of the manipulations needed for the different adjustments proposed in the thesis. Chapter 3 reviews the literature on lognormality and heavy tails of the wealth distribution. It then develops a methodology to estimate parameters and to assess the most suitable distributional form given outstanding amounts by account range – which correspond to the bin probabilities of the size-biased wealth distribution. Chapter 4 formalises a methodology to aggregate the observed range-sum statistics from account to household level. Chapter 5 provides the final methodology needed to ensure comparability between supervisory administrative data (i.e. Bank Supervisory Reports, BSR henceforth) and HFCS data. Leveraging copula theory, it disentangles the marginal distributions of listed shares, deposits, and investment fund shares, given outstanding amounts by overall securities-account range. Chapter 6 formalises in more depth the three calibration approaches considered in this thesis. It starts from the iterative, bounded value-based calibration to BSR class totals already embedded in the Italian DWA framework and published with Andrea Neri and Francesco Vercelli (Neri et al., 2023). It then introduces a univariate, distribution-aware calibration based on a composite loss that targets both marginal shape and the relative cumulative-amount path, delivering smoother micro-adjustments beyond matching bracket totals alone. Finally, it extends the composite criterion to a multivariate setting by adding an explicit dependence penalty on Gaussian-score correlations, so that cross-instrument dependence is adjusted while preserving BSR-consistent marginal benchmarks. Chapter 7 concludes.

Chapter 2

Data Sources and Preliminary Comparisons

2.1 Introduction

The empirical analysis in this thesis relies on three complementary data sources: the Household Finance and Consumption Survey (HFCS), the Quarterly Sector Accounts (QSA), and the Banking Supervisory Reports (BSR). As described in Chapter 1, the HFCS and QSA constitute the fundamental components of the Distributional Wealth Accounts (DWA) framework and are used by all participating countries. In this work, the additional BSR dataset is incorporated to improve the correction of several financial instruments – namely deposits, shares, debt securities (hereafter, "bonds"[1]), and investment fund shares – within the Italian DWA methodology.

The HFCS microdata used here have already undergone a subset of the DWA adjustments described in Section 1.2.2, namely all the adjustments applied prior to the final proportional allocation. This is done to reflect accurately the stage at which the proposed methods would intervene in the production process. Throughout this chapter, we refer to harmonised HFCS and QSA when the data follow the conceptual harmonisations discussed in Section 1.2.1.

The BSR data consist of administrative aggregates collected for supervisory purposes and made available to the Bank of Italy. summarise the population of accounts by value range. For each reporting date, the BSR provide (*i*) the number of accounts and (*ii*) the corresponding outstanding amounts within each account-value range. This information is available for deposit accounts and for securities accounts.

Securities accounts are portfolio accounts that may include shares, bonds, and investment fund shares. Importantly, the BSR range classification for securities accounts is defined on the total portfolio value (i.e., the sum of these components within the account). Within each portfolio-value range, however, the BSR report outstanding amounts separately for shares, bonds, and investment fund shares. This

[1]For brevity, throughout the thesis we use "bonds" as shorthand for debt securities, even though bonds are formally a subset of debt securities in standard instrument taxonomies.

disaggregated-by-instrument information, conditional on a common portfolio-value bin, is exploited in later chapters to disentangle the three components.

Although the BSR lack household identifiers, their information can be used to improve the quality of the adjustments applied to the HFCS data in the DWA methodology prior to the final proportional allocation. Their use requires specific transformations to ensure comparability with the HFCS and QSA data.

This chapter first presents an operational description of each dataset, focusing on the Italian composition for HFCS and QSA, as they were broadly described in Chapter 1. The aim is to clarify coverage, frequency, and relevant variables, as well as their strengths and limitations, for the purposes of the calibrations undertaken in subsequent chapters. Details on dataset vintages, coverage, and frequency are provided in the dataset subsections of this chapter (see, e.g., Section 2.4.2). In particular, the differences between the BSR and the other two datasets will be described in detail. The preliminary aggregate comparisons between the three sources are then reported, highlighting systematic discrepancies and motivating the methodological developments that follow. The chapter concludes by situating the analysis within the DWA production process and clarifying the specific point of intervention of the methods developed in this thesis.

2.2 Quarterly Sector Accounts (QSA)

The Quarterly Sector Accounts (QSA) provide aggregated statistics on the financial and non-financial assets and liabilities of the main institutional sectors of the economy. All EU Member States compile such accounts in a harmonised manner, following the European System of Accounts (ESA 2010) methodology.[2] Time series are available from the first quarter of 1999 onwards, with dissemination typically occurring around four months after the end of the reference period.[3] Concerning the series used in the DWA process, they are mandatory for all member states and without missing values. In Italy, the compilation of QSA is split between two institutions: the Bank of Italy is responsible for the financial accounts, while the National Institute of Statistics (Istituto nazionale di statistica, ISTAT) compiles the non-financial accounts.

2.2.1 Financial accounts

The Italian Financial Accounts are compiled quarterly by the Bank of Italy and cover both transactions (flows) and positions (stocks) in financial assets and liabilities for all institutional sectors.[4] They are disseminated with a lag of around three months after the reference quarter. The compilation draws on multiple sources, including statistics obtained from supervisory reports from banks, investment funds,

[2] See Eurostat, *European System of Accounts (ESA 2010)*, https://ec.europa.eu/eurostat/web/esa-2010.

[3] See ECB, *Quarterly Sector Accounts (QSA)*, https://data.ecb.europa.eu/data/datasets/QSA/data-information.

[4] See Bank of Italy, *Financial Accounts*, https://www.bancaditalia.it/statistiche/tematiche/conti-patrimoniali/conti-finanza/index.html?com.dotmarketing.htmlpage.language=1.

insurance companies, and pension funds; securities databases such as the ECB's Centralised Securities Database (CSDB);[5] government finance statistics; and balance of payments data. For this thesis, QSA are central as they provide totals for deposits, bonds, listed shares, and investment fund shares held by the household sector.

2.2.2 Non-financial accounts

The Italian Non-Financial Accounts are compiled and released quarterly by ISTAT in a framework consistent with annual national accounts.[6] They cover production, income, consumption, saving, and accumulation flows. For the household sector, they also estimate non-financial assets such as dwellings, land, and consumer durables. These components account for a large share of total household wealth, particularly for lower- and middle-wealth households. While non-financial assets are not the focus of the calibration exercises in this thesis, they are essential for completeness of the balance sheet.

2.2.3 Strengths and limitations

The main strengths of the QSA are: (*i*) harmonisation across EU countries via ESA 2010, which ensures international comparability across sectors and instruments; (*ii*) timeliness, with quarterly releases and relatively short lags; and (*iii*) comprehensive coverage of the national economy. In particular, reliance on administrative and supervisory data ensures high accuracy for deposits, listed securities, and investment fund shares.

At the same time, QSA present some limitations. They provide no distributional detail within the household sector, as they are aggregate statistics only. The household sector is partly residual in the accounting system and may therefore absorb statistical discrepancies when other sectors are measured more precisely. Misclassification may also occur (e.g., residents vs non-residents), and certain instruments – such as unlisted shares or household business wealth – are inherently difficult to measure. Finally, figures are subject to frequent revisions, reflecting benchmarking and the incorporation of new and revised sources, which can affect long-term comparability.

2.2.4 Role in the DWA

Within the DWA framework, QSA serve as the macroeconomic benchmark. Adjusted HFCS microdata are aligned with QSA totals for each financial instrument, ensuring consistency between micro and macro data. Moreover, QSA aggregates provide the reference for interpolating between HFCS waves and extrapolating beyond them, enabling the derivation of quarterly distributional estimates. In this thesis, the financial accounts are of primary importance, as they provide benchmarks for deposits, listed shares, bonds, and investment fund shares. Non-financial accounts

[5] https://www.ecb.europa.eu/stats/financial$_m arkets_a nd_i nterest_r ates/securities/html/index.en.html$

[6] See ISTAT, *Quarterly Non-Financial Accounts by Institutional Sector*, https://www.istat.it/en/archive/quarterly-accounts.

are less directly involved in calibration but remain important for understanding the full composition of household wealth.

2.2.5 Preliminary descriptives

Concerning the instruments covered in this thesis, they account for a large share of household wealth as defined in the DWA. Table 2.1 reports the QSA Italian household-sector outstanding amounts, the share of total financial wealth, and the share of total wealth for each instrument. Values refer to the 2025 Q1 release (published 16 July 2025).[7] Figures reflect DWA-concept harmonisations applied to ensure comparability across sources; totals are taken from QSA household-sector aggregates. Values are expressed in € millions.

Table 2.1. QSA household sector totals, share over financial wealth, share over total wealth for deposits, listed shares, bonds, and investment funds shares.

Quarter	Instrument	Total	% of Financial Wealth	% of Total Wealth
2021 Q4	Deposits	1,383,355	29.79	12.94
	Listed Shares	143,884	3.10	1.35
	Bonds	229,066	4.93	2.14
	Fund Shares	765,494	16.48	7.16
2022 Q4	Deposits	1,389,714	30.78	13.00
	Listed Shares	126,603	2.80	1.18
	Bonds	259,996	5.76	2.43
	Fund Shares	674,924	14.95	6.31
2023 Q4	Deposits	1,335,807	27.12	11.93
	Listed Shares	148,508	3.02	1.33
	Bonds	424,231	8.61	3.79
	Fund Shares	715,052	14.52	6.39
2024 Q4	Deposits	1,351,812	26.12	11.72
	Listed Shares	166,744	3.22	1.45
	Bonds	484,743	9.37	4.20
	Fund Shares	841,153	16.26	7.29

Notes: Values in € millions; shares in per cent. Total Wealth and Financial Wealth are defined following the DWA definition of household wealth (for further insights, see Table 1, Section 2.2.1 in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note_0.pdf). Source: Bank of Italy, Quarterly Sector Accounts (release 16 July 2025); author's computations; DWA concept harmonisations applied.

As shown, these four instruments account for roughly 55% of the households' total financial wealth and roughly 25% of the total wealth, thus comprising a very relevant share. Improving their coverage, even just reducing the relevance of the

[7] https://www.bancaditalia.it/pubblicazioni/conti-finanziari/

proportional allocation applied in the DWA last step, could yield significant changes and improve considerably the representativeness of the DWA macro and micro data.

2.3 Banking Supervisory Reports (BSR)

2.3.1 Operational description

The Italian Banking Supervisory Reports (BSR) collect, at end-June and end-December, outstanding amounts and the number of client-accounts by bracket for deposits and for securities held in custody (bonds, listed shares, and investment fund shares).[8] Reporting is at the intermediary level and clients are, prior to the aggregation, identified by fiscal code. Multiple accounts held by the same person at a given bank are consolidated for range assignment; joint accounts are not split across holders. Cross-bank consolidation is not possible: the same fiscal code may therefore appear in multiple intermediaries and, consequently, in different brackets across banks. In this thesis, end-December vintages are used to align with the 2010Q4, 2014Q4, 2016Q4, 2020Q4, and 2022Q4 reference quarters adopted for HFCS and QSA (Sections 2.4 and 2.2).

2.3.2 Securities accounts and coverage

Beyond deposits, the BSR report outstanding amounts for securities accounts by portfolio-value range (i.e., ranges defined on the total securities-account value), and within each range they provide amounts separately for bonds, listed shares, and investment fund shares. These series underpin the disentanglement procedure developed later to recover instrument-specific marginals from the securities-account aggregates (Chapter 5). Coverage with respect to sector-account totals differs by instrument and is summarised here using typical values over the analysis period (author's calculations on BSR and QSA). For deposits, coverage is around 90 per cent and totals are reconciled to the QSA by proportional rescaling.[9] For bonds, coverage is typically about 99 per cent (mean ~ 0.99, median ~ 0.99 over the years covered in this thesis) and above 95 per cent in every year. For listed shares, coverage is around 85–90 per cent (mean ~ 0.86, median ~ 0.89, minimum ~ 0.73). For investment fund shares, coverage is lower and more variable (mean ~ 0.66, median ~ 0.66, range ~ 0.53–0.72). The shortfall for listed shares and investment fund shares largely reflects cross-border custody and portfolio-management arrangements that place part of households' portfolios outside domestic custodians; see the Bank of Italy methodological update on the sectoral reallocation of portfolio assets.[10]

[8] Since 2022Q4, they are disseminated quarterly.

[9] For operational use of deposit information and reconciliation choices in Italy, see Neri et al. (2023).

[10] Bank of Italy (2022), *Le attività estere di portafoglio dell'Italia: la nuova ripartizione per settore detentore*, https://www.bancaditalia.it/. . . /Metodi_e_Fonti_Settori_Portafoglio.pdf.

2.3.3 Range definitions and vintages

The reported brackets are right-open intervals, $[a, b)$; the top bracket is open-ended, $[a, \infty)$. Amounts exactly equal to a bracket's upper threshold b fall into the next bracket. For comparability over time, we report the official bracket schedules for deposits and securities accounts in Tables 2.2 and 2.3, respectively; calibration uses the finest available granularity. For clarity, all amounts are in euro; we display interval bounds in compact form using k and mln, where k = 10^3 and mln = 10^6. As shown, from 2022 onwards, the size of the brackets has shrunk, effectively enhancing the informative capacity of the BSR data. Concerning deposits, we have a finer breakdown, especially in the lower and central part of the distribution. Conversely, securities accounts present a finer breakdown of the top open-ended bin. In both cases, these are the areas where the greatest holding share (and thus, information) is.

Table 2.2. BSR deposit-account brackets: pre-2022 vs post-2022

Pre-2022	Post-2022
$[0, 12.5\text{k})$	$[0, 5\text{k})$ $[5\text{k}, 12.5\text{k})$
$[12.5\text{k}, 50\text{k})$	$[12.5\text{k}, 25\text{k})$ $[25\text{k}, 50\text{k})$
$[50\text{k}, 250\text{k})$	$[50\text{k}, 100\text{k})$ $[100\text{k}, 250\text{k})$
$[250\text{k}, 500\text{k})$	$[250\text{k}, 500\text{k})$
$[500\text{k}, \infty)$	$[500\text{k}, 1\text{mln})$ $[1\text{mln}, \infty)$

Table 2.3. BSR securities-account brackets: pre-2022 vs post-2022

Pre-2022	Post-2022
$[0, 50\text{k})$	$[0, 50\text{k})$
$[50\text{k}, 250\text{k})$	$[50\text{k}, 250\text{k})$
$[250\text{k}, 500\text{k})$	$[250\text{k}, 500\text{k})$
$[500\text{k}, \infty)$	$[500\text{k}, 1\text{mln})$ $[1\text{mln}, 5\text{mln})$ $[5\text{mln}, \infty)$

2.3.4 On the use of BSR account counts versus outstanding amounts

In what follows, we focus on the BSR range-sum information on outstanding amounts and deliberately refrain from exploiting the reported numbers of accounts by range. This choice is motivated by conceptual alignment with the targets of interest, data coverage, and identification.

Alignment with macro targets. The reconciliation target throughout the thesis is defined in terms of amounts. In particular, national aggregates (e.g. DWA/QSA) are available and macro-consistent for outstanding balances by instrument, whereas there is no macro benchmark for the number of accounts. Introducing a variable that lacks an external anchor would risk diverting the calibration from the object that ultimately needs to be matched. We exploit this information solely as additional information in the deposit account-based distribution reconstruction, where we augment the procedure with a soft penalty on the estimated raw BSR average

deposit account value. Consequently, using BSR amounts maintains coherence with the ultimate aggregation constraints, while account counts would add an unconstrained dimension.

Differential coverage and unknown bias in counts. For amounts, supervisory sources provide credible coverage rates as shown, together with institutional information indicating the drivers of undercoverage; these elements enable a transparent and credible correction at the level of outstanding totals. By contrast, for the number of accounts, there is no reliable external benchmark on the population total across institutions. Moreover, there is no reason to expect that the selection mechanism for counts mirrors that for amounts. In practice:

- multiplicity: wealthier households tend to hold multiple accounts and sub-accounts across providers;
- institutional practices: omnibus/custodial and nominee arrangements can pool or split positions, altering the count while leaving amounts largely unaffected;
- low-balance activity: dormant or near-zero accounts inflate counts with negligible effect on amounts;
- reporting conventions: bracket redesigns or definitional heterogeneity can shift counts mechanically without commensurate changes in amounts.

Hence, even if the percentage of amounts missing were known or credibly bounded, the share of accounts missing is likely different in both level and pattern. Treating the two biases as comparable would be unwarranted.

Unit mismatch and aggregation level. BSR account counts are recorded at the account level, while the HFCS is at the household level. The mapping from accounts to households is many-to-one and institution-dependent (joint accounts, household portfolios split across providers, nominee relationships). As a result, counts are not directly additive, and their aggregation to the household level requires strong assumptions that are not testable with the available data. By comparison, outstanding amounts can be reconciled to household aggregates via identity constraints and macro totals.

Implications for identification and efficiency. Range-sum amounts identify size-biased bin probabilities that are directly informative for the tail shape and for splice/threshold parameters.[11] Bringing raw account counts (which reflect ordinary rather than size-biased probabilities) into the same set of moments without a reliable error model introduces errors-in-variables into GMM/MD estimation, potentially biasing tail and dependence parameters. Since the calibration objective is to match macro amounts, adding noisy count moments risks overfitting features of the counting

[11] See Chapter 3 for further discussion.

process rather than improving identification for the macro constraint.[12]

Given (*i*) the absence of a macro benchmark for the total number of accounts, (*ii*) the plausibility of a bias process for counts that differs materially from that for amounts, (*iii*) the account-to-household unit mismatch, and (*iv*) the risk of introducing unmodelled measurement error that does not aid identification of the macro objectives, we exclude BSR account counts from the baseline estimation and place exclusive weight on the range-sum amounts.

2.3.5 Pre-processing and reconciliation to QSA

Throughout the thesis, "BSR data" refer to amounts after reconciliation. BSR amounts by range are aligned to sector-account totals to ensure macro consistency, in line with the DWA workflow.[13] Before reconciliation, securities-account brackets (bonds, listed shares, and investment fund shares) are normalised as follows: all amounts below €50,000 are proportionally reallocated across the remaining brackets, and the lower tail is collapsed into a single <€250,000 class; deposits are unaffected. For 2011 and 2010, securities-account bracket totals are backcast by fixing the 2012 bracket shares and scaling them by the corresponding QSA instrument totals.

These first choices follow the evidence highlighted in Vercelli (2024), where the author notes – and supports with further analysis available within our institution – that lower-bracket values in BSR are likely connected to secondary accounts of wealthier households. This adjustment improves the representativeness of BSR securities accounts and further improves the account-household matching between BSR and HFCS, which is further discussed in Section 2.4.3.

After this normalisation, we apply corrections to reconcile data to their corresponding QSA macro-figures. For deposits, we leverage information on other administrative sources to allocate proportionally postal deposits (*libretti postali*), which are not captured by BSR data but are included in the QSA deposit aggregate.[14] We then assign the residual component treating it as foreign-held component of deposits proportionally to the top tail (i.e., to the classes with $\geq$ 250k values), consistent with evidence that offshore financial assets are concentrated among the wealthiest households (Alstadsæter et al., 2019).

For example, let's consider the 2020 values. BSR totals amount to 1,281,023 milion euros, QSA total deposits held by the household sector on the same reference period amount to 1,327,852, of which 61,182 postal deposits. We use BSR

[12] Counts could, in principle, be incorporated with an explicit measurement-error model and additional structure on multiplicity; however, this would require external information on account proliferation, dormant incidence, and institutional pooling practices, which is currently unavailable.

[13] ECB (2024), *Experimental Distributional Wealth Accounts for the household sector: Methodological note*, https://data.ecb.europa.eu/... /DWA_Methodological_note.pdf.

[14] Postal savings comprise deposit accounts (*libretti*, classified as deposits) and postal savings bonds (*buoni fruttiferi*, classified as securities); only the former are allocated within deposits here. See Osservatorio CPI (Università Cattolica), *CDP e i conti pubblici*, p. 6, https://osservatoriocpi.unicatt.it/... /CDP_conti_pubblici_OssCPI.pdf.

additional data to proportionally split the latter on the different classes. Following the postal deposit distribution, the residual component is treated as foreign-held and is allocated proportionally to all classes above 250k of values. This pushes the BSR implied totals to exactly match the QSA implied ones. Table 2.4 reports the different values added.

Table 2.4. BSR raw and revised values for deposits in 2020.

Class	Deposits			
	BSR Raw	**Postal Savings**	**Residual (foreign)**	**Revised**
[0k, 12.5k)	137,162	8,379	0	145,541
[12.5k, 50k)	319,930	16,586	0	336,515
[50k, 250k)	529,119	25,069	0	554,188
[250k, 500k)	116,930	6,464	24,539	147,933
[500k, ∞)	114,881	4,684	24,109	143,674
Total	1,218,023	61,182	48,648	1,327,852

Notes: Values in € millions.

Concerning bonds, since they are almost perfectly aligned – reflecting the fact that Italian households tend to hold them in Italian accounts – we apply a simple proportional allocation of the gap. For listed shares and investment fund shares, we adopt an instrument-specific reconciliation reflecting the channels highlighted in a Bank of Italy in-depth analysis on foreign holdings of Italian households.[15]

One half of the BSR–QSA gap is allocated proportionally across all normalised ranges (capturing foreign custody via asset management), and one half is allocated to balances at or above €250,000 (capturing directly foreign-held assets concentrated in the upper tail). Years with raw coverage above 100 per cent for listed shares are downscaled proportionally to 100 per cent. These choices preserve the distributional content of the normalised supervisory brackets while enforcing aggregate consistency with the sector accounts. Table 2.5 illustrates the effect for securities accounts in 2022.

Throughout this thesis, we will refer to the reconciled BSR data simply as BSR for the sake of simplicity.

[15]Bank of Italy (2022), *Le attività estere di portafoglio dell'Italia: la nuova ripartizione per settore detentore*, https://www.bancaditalia.it/. . . /Metodi_e_Fonti_Settori_Portafoglio.pdf.

Table 2.5. BSR raw and revised values for listed shares, bonds, and investment fund shares in 2022.

Class		Listed Shares		Bonds		Investment Fund Shares	
Raw	**Revised**	**Raw**	**Revised**	**Raw**	**Revised**	**Raw**	**Revised**
$[0, 50k)$	$[0, 250k)$	9,595	26,172	21,633	82,149	51,413	221,323
$[50k, 250k)$		19,770		70,790		157,078	
$[250k, 500k)$		11,956	15,828	41,101	47,696	99,837	140,671
$[500k, 1mln)$		11,743	21,270	36,253	42,070	53,383	112,548
$[1mln, 5mln)$		20,089	32,319	49,048	56,918	62,578	125,504
$[5mln, \infty)$		19,103	31,014	26,854	31,163	26,648	74,878
Total		92,256	126,603	245,679	259,996	450,937	674,924

Notes: Values in € millions.

2.4 Household Finance and Consumption Survey (HFCS)

2.4.1 Operational description

The Household Finance and Consumption Survey (HFCS) is a harmonised sample survey conducted across all Eurosystem countries and a few other European Union countries that have not yet adopted the euro. It is coordinated by the European Central Bank (ECB) and implemented in each participating country by the respective National Central Bank and/or National Statistical Institute. The main purpose of the survey is to collect comparable micro-level data on household wealth, income, and consumption for both policy and research purposes.

The ECB provides methodological harmonisation and common requirements to achieve representativeness at the euro area and national level, while sampling designs are implemented nationally under these guidelines. Nevertheless, deviations may occur due to differences in collection modes or other country-specific reasons and are corrected ex post through weighting adjustments, imputation, and calibration.[16]

For Italy, the unit of observation is the private household residing in the country of reference; institutional households are excluded by design. Alongside variables on wealth, income, and consumption at the household level, the dataset includes individual-level demographic information. In Italy, the survey is conducted by the Bank of Italy[17], and the national data correspond to those collected for the Survey on Household Income and Wealth (SHIW) after appropriate harmonisation.

[16] See HFCN (2023), *Household Finance and Consumption Survey: Methodological report for the 2021 wave*, ECB Statistics Paper Series No. 45, https://www.ecb.europa.eu/.../ecb.sps45. For Italy-specific documentation, see Bank of Italy (2024), *SHIW 2022 Methodological Notes*, https://www.bancaditalia.it/.../Legen22_eng.pdf.

[17] See: *Eurosystem Survey data (HFCS)*, https://www.bancaditalia.it/statistiche/.../dati-indagine-europea/index.html.

Between Wave 3 and Wave 4, the Italian sample design was updated to improve statistical coverage of indebted and high-income households. Although this reduced longitudinal comparability with previous waves, it substantially improved alignment with macroeconomic figures. To allow comparability with earlier waves, a dedicated set of bridging weights was produced for Wave 4.

Alongside the variables collected for HFCS purposes, given the shared origin of the sample with SHIW, it is possible to link the HFCS microdata to a broader set of variables collected for the latter. This is particularly relevant for Wave 4, which is the first and only survey to include information on the share of total deposits held in the main account, alongside the number of deposit accounts held by each household as all the other surveys. As discussed in Chapter 4, these variables are crucial for the manipulations required to ensure comparability between the HFCS and the Banking Supervisory Reports (BSR) data.

This thesis uses HFCS microdata for Italy from Waves 1–5. For Italy, we map each wave to the end-of-year reference quarter used for the macro sources (QSA and BSR): Wave 1 refers to 2010Q4; Wave 2 refers to 2014Q4; Wave 3 refers to 2016Q4; Wave 4 refers to 2020Q4; Wave 5 refers to 2022Q4. All results reported here refer to these reference periods.

The Italian sample sizes for Waves 1–5, as well as their reference period, are reported in Table 2.6.

Table 2.6. HFCS waves for Italy: reference periods and sample sizes used in this thesis.

Wave	Reference period (Italy)	Sample size
1	2010 Q4	7,951
2	2014 Q4	8,156
3	2016 Q4	7,420
4	2020 Q4	6,239
5	2022 Q4	9,641

The HFCS microdata are confidential. Researchers can request access to anonymised HFCS microdata via the official application channels.[18] The HFCS follows probabilistic, stratified sampling designs implemented by national authorities under common Eurosystem standards, with targeted oversampling of high-wealth strata in several countries. Given the focus of this chapter on reconciliation with macro aggregates, we do not reproduce the full sampling documentation here.[19]

[18] https://www.ecb.europa.eu/stats/ecb_surveys/hfcs/html/index.en.html

[19] See HFCN (2023), *Methodological report for the 2021 wave*, SPS No. 45, and Bank of Italy (2024), *SHIW 2022 Methodological Notes* https://www.bancaditalia.it/.../Legen22_eng.pdf.

2.4.2 Pre-processing within DWA before use in this thesis

The HFCS data used in this thesis are not raw survey microdata. As described in Section 1.2.2, the DWA production process is composed of several subsequent adjustments that produce a final microdata set consistent with national QSA aggregates and disseminated in aggregated form. The objective of this thesis, introduced in Section 1.3, is to develop a multivariate calibration that intervenes prior to the final proportional allocation step of the DWA procedure.

In particular, the HFCS considered here has already undergone the following steps: (*i*) conceptual reconciliations for population coverage, timing, and wealth concepts; (*ii*) correction of deposit under-reporting through imputation, as implemented in the Italian DWA methodology (Neri et al., 2023); and (*iii*) the inclusion of Pareto-distributed synthetic households to address the under-representation of the wealthiest. By contrast, two subsequent steps are excluded: (*i*) the Italian-specific deposit calibration, since the current method coincides with the first methodology presented – and extended – in Chapter 6 and is the method this thesis proposed to enhance in Section 6.1.2–6.1.3; and (*ii*) the final proportional allocation ensuring exact consistency with QSA totals.

It should be noted that, as introduced in Section 2.1, HFCS and QSA data referenced in this thesis also undergo limited adjustments within the DWA framework to ensure comparability between the two sources (namely, the conceptual reconciliations). These adjustments are documented in detail in the DWA Methodological Note published by the ECB, but official codes are not publicly available.[20]

2.4.3 Comparison with BSR and QSA

As discussed, HFCS aggregates differ significantly from reconciled QSA macro-aggregates. Although improved sampling has mitigated the gap (Figure 1.2), it remains material for the instruments considered here. Before comparing these values with BSR data, we first harmonise the unit of observation (account vs household). For securities accounts, the SHIW (mergeable with the HFCS for Distributional Financial Accounts working group users) records the number of securities accounts held by each household. This allows us to document both the average number of declared accounts and its pattern across instrument quantiles. A similar pattern emerges for deposit accounts: the number of accounts per household tends to rise with the deposit quantile. Table 2.7 reports the average and median numbers of securities and deposit accounts, excluding zero holdings, by quintile group of instrument holdings (five groups defined by the 20th, 40th, 60th, and 80th percentiles).[21]

To further put these values into perspective, over the five waves, the share of households with more than one deposit account – given a positive holding of deposits – is $\sim 43\%$, while the share of households with more than one securities account – given a positive holding of at least one between listed shares, bonds, and investment

[20] https://data.ecb.europa.eu/.../DWA_Methodological_note.pdf.

[21] To improve the quality of the estimation, we considered only households with a non-zero reported value of each instrument and excluded those that reported zero number of accounts.

Table 2.7. Raw HFCS number of accounts by quintile group of instrument holdings.

Wave	Quintile group	Mean Value		Median Value	
		Deposits	Securities	Deposits	Securities
1	Q1 (0–20%)	1.37	1.29	1	1
	Q2 (20–40%)	1.29	1.28	1	1
	Q3 (40–60%)	1.51	1.29	1	1
	Q4 (60–80%)	1.69	1.29	1	1
	Q5 (80–100%)	1.91	1.32	2	1
2	Q1 (0–20%)	1.27	1.16	1	1
	Q2 (20–40%)	1.37	1.17	1	1
	Q3 (40–60%)	1.50	1.18	1	1
	Q4 (60–80%)	1.73	1.20	2	1
	Q5 (80–100%)	2.02	1.33	2	1
3	Q1 (0–20%)	1.31	1.16	1	1
	Q2 (20–40%)	1.36	1.01	1	1
	Q3 (40–60%)	1.43	1.06	1	1
	Q4 (60–80%)	1.75	1.17	1	1
	Q5 (80–100%)	2.01	1.17	2	1
4	Q1 (0–20%)	1.42	1.28	1	1
	Q2 (20–40%)	1.48	1.15	1	1
	Q3 (40–60%)	1.65	1.27	1	1
	Q4 (60–80%)	1.98	1.29	2	1
	Q5 (80–100%)	2.23	1.33	2	1
5	Q1 (0–20%)	1.32	1.20	1	1
	Q2 (20–40%)	1.47	1.16	1	1
	Q3 (40–60%)	1.65	1.13	1	1
	Q4 (60–80%)	1.87	1.34	2	1
	Q5 (80–100%)	2.16	1.65	2	1

Notes: "Quintile group" refers to the five bins obtained by splitting the holder distribution into five equal-weight parts; the underlying cut-points are the 20th, 40th, 60th, and 80th percentiles, and are computed within wave and instrument, using survey weights and restricting to holders (positive reported amounts). Households reporting zero accounts for the relevant instrument are excluded. "Securities" aggregates listed shares, bonds, and investment fund shares; the securities-account count is taken from the SHIW auxiliary variable, mergeable with HFCS. HFCS raw microdata (Italy), waves 1–5; no added rich or synthetic households.

fund shares – is $\sim 14.5\%$.[22] These numbers further diverge once we also account for a basic imputation of the deposit accounts and securities accounts. To avoid issues of false-zero reporting on the number of accounts, we consider a basic imputation where to households with holdings of the given instruments (deposits or the com-

[22]Further evidence of this is reported in Section 5.5.3

bination of listed shares, bonds, and investment fund shares) and a zero reported number of accounts, we impute the median number of accounts in their respective holding decile. Conversely, we also assign deposits or securities account values to households reporting a positive number of accounts. After these imputations over the five waves, the share of households with more than one deposit account is still $\sim 43\%$, confirming that the number of accounts declared is credible. Conversely, the share of households with strictly more than one securities account drops to $\sim 7.5\%$, though the overall share of households with at least one account increases. This reduction highlights not only how small the share of households with more than one securities account is (backing up even more the claim that account-level and household-level coincide), but also the need to consider an ad-hoc imputation process for the households considered in the securities account.

In sum, households report non-zero holdings of listed shares, bonds, and investment fund shares less frequently, whereas deposits are more widespread. As said, there is no direct official source to verify this, so preliminary evidence is based on HFCS data. However, it is important to note that HFCS could still be biased with false-zero claims, and results should be regarded as a lower bound. An indirect way of measuring the zero-reporting bias is introduced leveraging the results of methodology in Section 6.1.2. Table 2.8 shows the percentages of non-zero values in HFCS, including added rich and synthetic households (see Section 1.2.2).

Table 2.8. HFCS percentage of non-zero reporting households (weighted) by wave.

Wave	Reference period (Italy)	Non-zero Percentage			
		Deposits	Listed Shares	Bonds	Investment Fund Shares
1	2010Q4	91.18	4.60	14.59	6.30
2	2014Q4	90.48	3.77	13.11	6.03
3	2016Q4	92.76	3.32	10.22	6.25
4	2020Q4	96.02	5.71	9.59	7.39
5	2022Q4	96.54	4.36	9.76	9.26

Notes: Values in per cent. Source: HFCS following DWA reconciliation steps up to the proportional allocation.

As the table shows, holdings of listed shares, bonds, and investment fund shares are very scarce, while almost all households possess at least one deposit account. In particular, if we consider the combined holding of the three instruments contained in the securities accounts, the share of households which still reports zero-value holdings is roughly 82.5% over the five HFCS waves – ranging from 80% in Wave 1 to 85% in Wave 3.

For securities accounts, the near coincidence between account and household-level objects allows for a direct HFCS-BSR comparison. Conversely, deposits require reconciliation, as households often hold multiple accounts, especially in the upper tail. We therefore align units by splitting household deposits across accounts. The

Italian Wave 4 provides an additional variable on the share of deposits held in the main account (when multiple accounts are reported), which we use to allocate deposits across accounts, following Chapter 4 and Section 3.2 of (Neri et al., 2023).

After this transformation, HFCS and BSR-based totals are comparable. For each wave – under the relevant BSR bracket design – the results are reported in the Appendix B, Table B.1, In this case, we report HFCS raw values (as we did when comparing HFCS and QSA outstanding amounts).

For the other instruments, single instrument comparison is more complex in interpretation, as the BSR report amounts by securities-account ranges. Reconstructing instrument-specific marginals is non-trivial and is addressed in Chapter 5. As a preliminary step, we compare single-instrument aggregates at the household-level securities-accounts breaks, reconstructed in the HFCS by summing listed shares, bonds, and investment fund shares. Results for each wave after this reconciliation are reported in the Appendix B, Table B.2, Table B.3, and Table B.4, for listed shares, bonds, and investment fund shares, respectively.

The tables indicate substantial under-reporting for all instruments across the entire bracket distribution. Coverage improves after Wave 4, yet the shortfall remains sizeable and consistent across all the distributions. However, some features deserve emphasis. First, concerning deposits, Wave 4 HFCS displays a markedly heavier upper tail than earlier waves and, crucially, than the corresponding BSR distribution. Second, listed shares exhibit extremely high under-reporting up to Wave 4, which turns into over-reporting in Wave 5 due to the abrupt rise in HFCS totals in the top open-ended range for this instrument. Although counterintuitive, this pattern is consistent with internal validation by the SHIW working group and with the sampling redesign introduced in Wave 4, which disproportionately increased the inclusion probability of high-wealth households. Consequently, survey coverage rose strongly and mostly for instruments concentrated at the top of the wealth distribution – notably listed shares and investment fund shares – while the impact on broadly held instruments such as deposits and bonds was tighter, though still visible. The absence of an equally abrupt shift in Wave 4 is plausibly attributable to first-time respondent effects that dampen reporting changes, and additionally, for listed shares, to the heightened market volatility in 2020 that complicated accurate valuation by households with large holdings.

Taken together, these patterns reinforce the case for an administrative-data-dependent adjustment strategy in place of – or at least substantially down-weighting – proportional allocation. Calibrating survey values to administrative aggregates can accommodate wave-specific and period-specific biases, such as those affecting listed shares in Waves 4–5. Extending this approach to a multivariate setting, while accounting for survey-based dependence across instruments – assuming the rank-based dependence is at least partially preserved – should further enhance the credibility and precision of the bias correction.

Chapter 3

Parameter estimation from BSR range data

3.1 Introduction

As introduced in Section 2.3, the Banking Supervisory Reports (BSR) report outstanding amounts of deposits, bonds, listed shares, and investment fund shares by account value. Concretely, for each j-th account range $[a_{j-1}, a_j)$, with values from 1 to J where J is the total number of classes, the BSR reports provide the outstanding amounts of euros held in accounts with balances in that interval. Formally, letting X_i denote the balance of account i, the bracket total, with i varying from 1 to N where N is the total number of accounts recorded in the BSR, is:

$$T_j \;\equiv\; \sum_{i=1}^{N} X_i \, \mathbb{1}_{[a_{j-1},a_j)}\left(X_i\right).$$

Under a parametric density $f(x;\boldsymbol{\theta})$ for account-level balances, this can be rewritten in expectation as:

$$\mathbb{E}\left[T_j\right] \;=\; N\mathbb{E}\left[X\,\mathbb{1}_{[a_{j-1},a_j)}\left(X\right)\right] \;=\; N\int_{a_{j-1}}^{a_j} x\, f\left(x;\boldsymbol{\theta}\right)\, dx. \tag{3.1}$$

Equation (3.1) corresponds to the (unnormalised) first raw moment of X over $[a_{j-1}, a_j)$ – i.e., the partial first moment multiplied by N. It differs from the more familiar grouped-data setting based on bracket frequencies, $\int_{a_{j-1}}^{a_j} f(x;\boldsymbol{\theta})\, dx$, and motivates methods that explicitly exploit information on the first moment by range.

This chapter sets out the first methodological step towards multivariate calibration. As outlined in Section 1.3, the overarching goal is to improve the correction of survey-based measures of household wealth components within the DWA framework by incorporating as much distributional information as possible from administrative sources available for Italy, in particular the BSR. The expected benefits are twofold: (*i*) more reliable aggregate estimates of household wealth components; and (*ii*) refined microdata underlying the DWA procedure, thereby laying the groundwork for the dissemination of DWA-style microdata. Moreover, recovering the distributional

form of each component allows for intra-instrument concentration analyses that do not require sample surveys.

Specifically, Section 3.2 states the pivotal distributional assumption used to encode within-bracket shapes – namely, a lognormal body with a possible Pareto tail. Section 3.3 develops an estimator of $\boldsymbol{\theta}$ from BSR amount shares. Section 3.4 outlines a testing strategy to determine whether the Pareto tail component is required, thereby establishing whether a simple lognormal specification suffices. Section 3.5 presents results on both simulated and real data, assessing (*i*) model selection and (*ii*) parameter recovery. We stress that the literature offers little direct guidance on parametric laws for single-instrument household holdings; most formal results concern income or total wealth – typically a lognormal body with Pareto-type tails (see, e.g., Kleiber and Kotz, 2003; Hajargasht et al., 2012; Chotikapanich et al., 2018). Accordingly, the LN/PLN class is adopted as a parsimonious, testable working specification rather than a maintained truth.

3.2 Lognormality and Pareto tails in economic size distributions

This section sets out the working distributional assumption supporting all the methodologies developed in this thesis: instrument-level holdings are well described by a lognormal (LN) body, with a possible Pareto upper tail accommodated through a lognormal-Pareto (PLN) composite. The assumption is instrumental for two reasons. First, it simplifies parameter estimation from BSR data – thereby improving the distributional information extracted. Second, it provides a coherent parametric family to be used in all the later developed methodologies, and in particular in the L_2 calibration with composite distributional loss, and in the multivariate composite calibration developed in Section 6.1.2 and 6.1.3.

Although HFCS microdata could, in principle, be used to test the need for a Pareto tail and to assess the overall LN functional form, the bias that characterises these data – in particular, differential non-response that censors top-wealth households – undermines the possibility of using raw HFCS microdata to test for a Pareto tail across instruments. Moreover, the adjusted HFCS microdata used in this thesis already include a correction for the upper tail of the distribution based on the assumption that the upper tail of the wealth distribution follows a Pareto law.[1]

The combination of a lognormal body with a Pareto tail is consistent with classical results on size distributions and with recent evidence on the wealth distribution's top tail. The lognormal has a long tradition as a model for economic sizes (Aitchison and Brown, 1957), while empirical work on wealth demonstrates Paretian upper tails (Vermeulen, 2018). Moreover, recent operational work adopts Pareto-based tail corrections when reconciling survey wealth with macro aggregates for total wealth, reinforcing the case for a splice specification in applied settings (Cantarella

[1] See Section 1.2.2, paragraph "Pareto estimation".

et al., 2024). In parallel, alternative parametric families that embed a lognormal-like body with Pareto-type tails (e.g., double Pareto-lognormal) are available and will be discussed as robustness checks (Reed and Jorgensen, 2004). These elements provide the foundations for the parameter recovery and model testing developed in the following sections.

3.2.1 Literature review

The lognormal distribution's pervasiveness across sciences is well documented (Limpert et al., 2001; Aitchison and Brown, 1957). In economics, lognormality for the body of income and wealth distributions is often explained via multiplicative growth ("Gibrat's law"), whereby growth results from the accumulation of many small percentage changes (Kleiber and Kotz, 2003, Paragraph 4.2). In practice, a random variable X is lognormal if its logarithm $\log X$ is Gaussian. Formally, X is lognormal if $\log X \sim \mathcal{N}(\mu, \sigma^2)$, with mean $\mu \in R$ and variance $\sigma^2 > 0$, and its density function f_X and cumulative distribution function (CDF, henceforth) F_X are:

$$f_X(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \quad x > 0,\ \mu \in R,\ \sigma^2 > 0, \tag{3.2}$$

$$F_X(x; \mu, \sigma^2) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad x > 0,\ \mu \in R,\ \sigma^2 > 0, \tag{3.3}$$

where $\Phi(\cdot)$ represents the CDF of the standard normal distribution.

However, many economic size variables display heavier upper tails that a standard LN does not adequately capture. This empirical regularity traces back to Pareto's studies of income and wealth, where he observed a decreasing linear relationship between the logarithm of income and the logarithm of the number of observations – in other words, a very top tail that holds a disproportionately large share of income (or wealth) relative to an LN benchmark.[2] A simple model that captures such behaviour is the Pareto Type I distribution. A random variable Y is Pareto Type I above a threshold $\tau > 0$ with shape $\alpha > 0$ if its density function f_Y and its CDF F_Y are:

$$f_Y(y; \alpha, \tau) = \frac{\alpha}{\tau}\left(\frac{\tau}{y}\right)^{\alpha+1}, \quad y \geq \tau, \tag{3.4}$$

$$F_Y(y; \alpha, \tau) = 1 - \left(\frac{\tau}{y}\right)^{\alpha}, \quad y \geq \tau. \tag{3.5}$$

There is a hierarchy of Pareto-type families (Pareto Type I, II, III, IV, and Feller-Pareto distributions). In applied works, the Type I distribution already provides an effective approximation to heavy upper tails in many economic contexts, including wealth (Davies and Shorrocks, 2000; Gabaix, 2009). Accordingly, and

[2] Pareto, V. (1896), *Cours d'économie politique*, Vol. I, Lausanne: F. Rouge. Archive: `https://archive.org/details/fp-0148-1`.

given that BSR data are totals by brackets (which limits identification of additional parameters), we focus on the Pareto Type I distribution for two reasons: (*i*) multiple works show it is a good approximation for heavy upper tails in economic phenomena (e.g. Davies and Shorrocks, 2000; Gabaix, 2009; Vermeulen, 2018); and (*ii*) using more general forms risks over-parameterisation relative to the information content of BSR aggregates. Henceforth, the Pareto Type I distribution will be referred to as the Pareto distribution.

Specifically on wealth, a large literature has consolidated the usefulness of a Pareto tail. For Europe, Vermeulen (2018) combine survey data with rich lists to fit a Pareto tail for top wealth, significantly raising the estimated wealth share of the top 1% (with a methodology similar to that later adopted in the DWA). This addresses a well-known survey bias: standard household surveys (like HFCS) tend to under-represent the richest due to non-response and sampling limits.[3] As a result, a LN or other thin-tailed fit to raw survey data underestimates tail concentration. Similarly, Bach et al. (2019) report notable increases in top 1% shares for France, Germany, and Spain after integrating a Pareto tail. These findings underscore the empirical relevance of a lognormal body with a Pareto tail for wealth components. In the actuarial splicing literature – see, for instance, Scollnik (2007) and Cooray and Ananda (2005) – composite laws are constructed by joining conditional pieces at a threshold to ensure CDF continuity, with optional smoothness constraints and, in some variants, smooth transition weights around the splice. Following this general approach, we define the PLN density by joining a lognormal body and a Pareto tail at a threshold $\tau > 0$ using conditional components:

$$f_{\mathrm{PLN}}(z;\mu,\sigma,\alpha,\tau,c) \;=\; \begin{cases} c\,\dfrac{f_{\mathrm{LN}}(z;\mu,\sigma)}{F_{\mathrm{LN}}(\tau;\mu,\sigma)} & 0 < z \le \tau, \\ (1-c)\,f_{\mathrm{PA}}(z;\alpha,\tau) & z > \tau, \end{cases} \tag{3.6}$$

where $c \in (0,1)$ is a weighting constant and $F_{\mathrm{LN}}(\tau;\mu,\sigma)$ is the lognormal CDF at τ. Since the Pareto component is supported on (τ,∞), it is already normalised on its domain, and no additional truncation factor is required. CDF continuity at τ holds by construction. If density continuity at the splice is desired – a common choice in composite-modelling practice – it pins down c through

$$c\,\frac{f_{\mathrm{LN}}(\tau;\mu,\sigma)}{F_{\mathrm{LN}}(\tau;\mu,\sigma)} \;=\; (1-c)\,f_{\mathrm{PA}}(\tau;\alpha,\tau),$$

which can be used to eliminate c from the parameter vector. Solving for c yields the closed-form expression:

$$c(\mu,\sigma,\alpha,\tau) \;=\; \frac{f_{\mathrm{PA}}(\tau;\alpha,\tau)}{f_{\mathrm{PA}}(\tau;\alpha,\tau) + \dfrac{f_{\mathrm{LN}}(\tau;\mu,\sigma)}{F_{\mathrm{LN}}(\tau;\mu,\sigma)}}\,. \tag{3.7}$$

Moment finiteness for BSR amount shares is ensured by requiring $\alpha > 1$ for a finite first moment in the tail.

[3] See Section 1.2.1 for further discussion.

In the thesis, we adopt this smooth-fit splice – CDF continuity and density continuity at τ – as the reference definition. This choice yields an explicit and interpretable split between body parameters (μ, σ) and tail parameters (α, τ), preserves normalisation on $(0, \infty)$, and produces closed-form bin integrals under (3.6), which are crucial for estimation with BSR range-sum shares. It also aligns with the downstream L_2 composite calibration, where a distribution-distance penalty compares the empirical CDF with the PLN CDF: the explicit threshold τ governs where the loss transitions from LN to Pareto behaviour without introducing artificial curvature at the splice. By contrast, (*i*) fixing $c = F_{\mathrm{LN}}(\tau)$ – effectively taking the unscaled LN below τ – generally introduces a density kink at the splice; and (*ii*) soft-merge variants with smooth transition weights (as sometimes considered in composite settings) add tuning parameters that are weakly identified with exogenous binning and an open-top bracket. Relative to Scollnik (2007) and Cooray and Ananda (2005), this formulation follows their conditional-piece construction but specifically imposes density continuity to determine $c = c(\mu, \sigma, \alpha, \tau)$, thereby retaining a four-parameter model tailored to identification with BSR range-sum inputs.

Before progressing to a more extensive review of the alternative distributional models used in economic wealth and income-related analysis, it must be noted that almost all literature focuses on aggregate wealth holdings of households, and far less evidence is available for the specific financial instruments covered in this thesis (deposits, bonds, listed shares, and investment fund shares). It is then worth dedicating a few other words to these instruments, in particular to deposits.

Specific wealth instruments While much of the literature establishes a LN body with a Pareto upper tail for total wealth and income, there is comparatively little direct evidence on the parametric form of the marginal distribution of individual instruments (in our case, deposits, bonds, listed shares, and investment fund shares). In general, balances in such instruments grow via the accumulation of many small inflows and interest accruals – a multiplicative-type process that naturally yields LN outcomes.

Concerning the last three instruments (bonds, listed shares, and investment fund shares), it is reasonable in the Italian context to anticipate heavy-tailed behaviour. As shown in Table 2.8, holdings of these instruments are relatively scarce, with approximately 82.5% of households having zero holdings for each, over the five HFCS waves. Moreover, aggregate HFCS statistics show that these assets become increasingly relevant as household wealth rises – i.e., households with high net wealth tend to hold a larger portion of their portfolios in these instruments, see Tables B.2–B.4. This points to a disproportionate-holdings effect, whereby the majority of outstanding amounts are concentrated among high-wealth households. These patterns suggest a two-part model comprising a point mass at zero and a continuous positive component, for which a plain LN or a PLN specification may be appropriate.

Conversely, deposits appear more diffuse along the wealth distribution and con-

stitute an important part of financial wealth for low-wealth households. Several theoretical considerations suggest that the distribution of deposits may be well approximated by a plain LN. First, households' precautionary saving behaviour tends to scale with income in percentage terms, aligning with Gibrat's law. Second, there are institutional and behavioural frictions that can attenuate extremely fat upper tails at the single-account level. For example, deposit insurance schemes (such as the ∼€100,000 guarantee in the EU) reduce the incentive to keep far above that threshold in one account.[4] Third, as shown from our study on the preliminary findings on Italian DWA aggregated results (Neri et al., 2024), the share of deposits in the total portfolio declines with wealth, consistent with deposits increasing at a slower pace than wealth itself (which exhibits a Pareto tail). Taken together, these facts suggest that, conditional on participation, the positive part of household deposits could be adequately captured by a thin-tailed form such as a LN, rather than a power-law tail.

Nonetheless, given the literature gap on the distribution of single-instrument holdings, this motivates instrument-by-instrument tests of a PLN rather than a LN specification using BSR range-sum data. This specific issue is addressed in Section 3.4, which sets out a testing procedure after the identification and estimation of the underlying parameters for a given functional form using binned (range-sum) data.

Alternative distributional models It is worth mentioning that a number of alternative parametric families have been proposed for income and wealth data modelling. The Generalised Beta of the Second Kind (GB2) family (encompassing the Singh-Maddala and Dagum models) is particularly popular due to its flexibility – being defined with four parameters controlling for peakedness, shape, and two for scale, respectively – nesting LN, gamma, Weibull and Pareto laws as special or limiting cases (McDonald and Xu, 1995).[5,6,7] Comparative studies show that GB2 or Dagum-type fits can perform very well for income distributions in many settings, and comprehensive reviews emphasise their empirical success (Kleiber and Kotz, 2003).

In this thesis, however, we deemed the choice of the PLN distribution more adequate. With only aggregated inputs such as BSR range-sums, four-parameter specifications are more demanding to identify and stabilise than two or three-parameter constructions (see, e.g., grouped-data estimation discussions in Chotikapanich et al., 2018 and related GMM treatments for grouped shares/means in Hajargasht and Griffiths, 2013). Moreover, GB2 does not deliver a clean body–tail separation. By

[4] European Banking Authority, Deposit Guarantee Schemes data: https://www.eba.europa.eu/. . . /deposit-guarantee-schemes-data

[5] For further details on the Dagum distribution, see Kleiber, Christian. "A guide to the Dagum distributions." *Modeling income distributions and Lorenz curves.* New York, NY: Springer New York, 2008. 97–117.

[6] For further details on the Singh-Maddala distribution, see Singh, Surendra K., and Gary S. Maddala. "A function for size distribution of incomes." *Modeling income distributions and Lorenz curves.* New York, NY: Springer New York, 2008. 27–35.

[7] For further details on the GB2 distribution, see Chotikapanich, Duangkamon, et al. *Using the GB2 income distribution. Econometrics* 6.2 (2018): 21.

contrast, a PLN splice isolates tail parameters (threshold τ and tail index α) from body parameters (μ, σ), which is advantageous both for testing whether a simple LN suffices and for interpreting tail behaviour. From a literature perspective on binned inputs, range-sum statistics correspond to features of a size-biased law rather than the original density, so identification hinges on how well a family maps into bracketed amount shares; this makes estimation sensitive to bin design, especially in the upper tail (Patil and Rao, 1978; Chotikapanich et al., 2018; Kleiber and Kotz, 2003). Reviews and grouped-data applications further note that richer four-parameter families (GB2; double Pareto-lognormal, dPLN) and mixtures can be fragile under coarse aggregation and typically require many, well-spaced bins to stabilise tail parameters, whereas parsimonious bodies (LN) and spliced forms with an explicit threshold (PLN) tend to be more robustly identified and interpretable in practice (Hajargasht and Griffiths, 2013; Scollnik, 2007; Reed and Jorgensen, 2004). A related practical point is that bracket shares are compositional objects (they lie on the simplex), and methods that operate directly on shares are well aligned with this data structure (Aitchison, 1982).

Another option is the family of skewed lognormal variants (e.g., the log-skew-normal, LSN), which introduces a skewness parameter on the log scale and can capture asymmetry beyond a plain LN.[8] However, tails in the LSN model remain lighter than Pareto, so these models typically do not address heavy-tail phenomena if present.

Mixtures of LN provide another flexible approach: a two-component mixture can approximate heavy right tails (a high-variance component generating large values) and accommodate heterogeneity or mild multimodality. This strategy has a long tradition in the LN literature (see, e.g., Aitchison and Brown, 1957). Mixtures, however, raise practical questions on the number of components and can overfit with limited (binned) information. A related but conceptually distinct route is to use dynamic mixtures where the mixing weight varies with x to estimate tails without an explicit threshold (e.g., Frigessi et al., 2002); these methods are elegant, but were developed for micro data and rely on features (e.g., local likelihood) that are not directly exploitable with aggregated range-sums.

Finally, the dPLN is a single-law alternative with a LN core and Pareto-type tails on both sides, grounded in proportional growth mechanics, and often competitive with GB2 in empirical work (Reed and Jorgensen, 2004). Its tractability is attractive, but, as with GB2, dPLN does not deliver an explicit body-tail separation at a data-informed threshold, which is central to the present application.

To make the trade-offs more transparent, Figure 3.1 provides a graphical comparison between the PLN specification adopted in this chapter and a set of commonly used alternatives (a skewed lognormal and GB2). We keep the LN baseline fixed at $(\mu, \sigma) = (8.5, 1.5)$ and calibrate the other families to deliver broadly comparable

[8] See Azzalini, Adelchi, Thomas Del Cappello, and Samuel Kotz. *Log-skew-normal and log-skew-t distributions as models for family income data. Journal of Income Distribution* 11.3–4 (2002).

levels and dispersion. Panel A shows the full density on a log-x scale. LN and PLN are nearly indistinguishable over most of the support – the LN curve is effectively hidden because it closely overlays the PLN curve in the body. By contrast, GB2 implies a materially different body shape, with relatively more density at very small balances and a flatter mid-range, while the skewed lognormal concentrates mass more tightly around its mode. Panel B zooms into the far right tail and highlights the substantive difference that matters for our application: PLN delivers a markedly thicker right tail than LN and, in this calibration, tail decay that is very close to GB2. This motivates the PLN choice as a compromise between parsimony and tail flexibility: it preserves a lognormal body while allowing the upper tail to depart in a controlled and interpretable way.

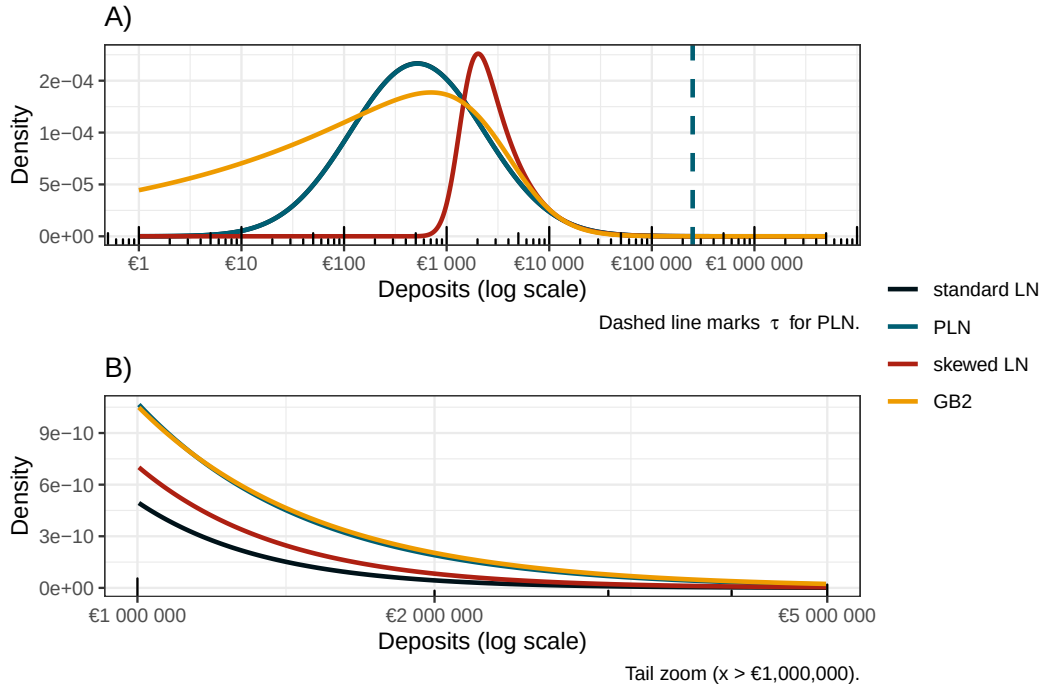

Figure 3.1. Comparison of candidate parametric families used to model deposits. Panel A shows the full density on a log-x scale; Panel B zooms into the far right tail. The LN baseline is fixed at $(\mu, \sigma) = (8.5, 1.5)$; PLN uses $\alpha = 1.5$ and $\tau = 250{,}000$.

In sum, while GB2, skewed lognormal variants, mixtures, and dPLN offer useful benchmarks, the PLN splice is preferred here for three reasons: (*i*) it aligns with the observed heavy-tail vs. lognormal-body dichotomy; (*ii*) it separates body and tail parameters, facilitating targeted testing of tail relevance; and (*iii*) it is parsimonious and tractable with BSR range-sum inputs.

3.3 Parameter estimation from BSR range-sum data

This section addresses the recovery of parameters under a given distributional form, using only grouped (binned) data. In particular, the aim is to identify a suitable approach to estimate parameters within an assumed family, focusing on the lognor-

mal (LN, henceforth) and the Pareto-lognormal (PLN, henceforth). As discussed in Section 3.1, BSR provide outstanding amounts by account range which correspond, in expectation, to size-biased truncated first moments; by contrast, typical grouped (binned) data provide counts by range (or, equivalently, frequency shares).

The goal is to recover the underlying distribution implied by the estimated parameters. The section proceeds as follows. First, it reviews the state of the art, outlining alternative methodological solutions and motivating the chosen approach. Second, it presents the theoretical framework for the selected method. Finally, it reports results on simulated and real data. It is worth noting that the real-data application is, for the time being, limited to deposits: BSR currently report the other three instruments (bonds, listed shares, and investment fund shares) in aggregated form (see Section 2.3), so these components must first be disentangled – this is undertaken in Chapter 5.

The key idea is to choose parameters that minimise discrepancies between observed bin statistics (share of total amount held in each range, $\widehat{s}_j$) and those predicted by the candidate distribution ($s_j(\boldsymbol{\theta})$). The approach can be viewed as a specialised method-of-moments or minimum-distance estimator: in practice, we equate the empirical aggregated moments over each bin to the theoretical ones derived from the target distribution. By constructing an objective function that penalises differences between observed and theoretical bin shares and truncated means, we solve for the parameters that best align the two.

In this section, the focus is purely on evaluating the ability of this methodology to recover the underlying parameters, assuming the true functional form is chosen. In Section 3.4, we later develop a specific methodology to test for the presence of a heavy right tail given BSR data.

3.3.1 Literature review and candidate estimators

The literature concerning parametric income and wealth distributions from grouped information is extensive, also due to the prevalence of grouped data availability in these domains. When only grouped shares (e.g., decile or bracket shares) or grouped means are available, efficient procedures have been proposed for GB2, LN, and PLN/dPLN (see, e.g., Hajargasht and Griffiths, 2013; Griffiths and Hajargasht, 2012; Chotikapanich et al., 2018).

It is possible to summarise the proposed approaches in three main categories:

- Generalised Method of Moments/moment-based methods;
- Quasi/composite-likelihood methods;
- Minimum-divergence methods;

with the first two as the main focus of the current literature. However, most contributions work with one (or more) of the following grouped inputs:

- Bracket counts or frequencies, where bin edges are either exogenous (fixed thresholds) or endogenous (deciles/quantiles);
- Group means, with exogenous or endogenous bins;
- Income/wealth shares by population groups (e.g., top p%, deciles).

By contrast, as discussed in Section 3.1, BSR provide outstanding amounts by account range which correspond to size-biased truncated first moments by exogenous bins. Normalising these values by the total amount yields wealth shares

$$\widehat{s}_j = \frac{T_j}{T}, \tag{3.8}$$

for bin j, where $T = \sum_j T_j$ is the total outstanding amounts from BSR. These are the bin probabilities of the size-biased (length-biased) distribution:

$$\begin{aligned} g_\theta(x) &= \frac{x f(x;\boldsymbol{\theta})}{\mu_1(\boldsymbol{\theta})}, \\ \mu_1(\boldsymbol{\theta}) &\equiv \int_{\mathbb{R}_+} x\, f(x;\boldsymbol{\theta})\, dx, \end{aligned}$$

rather than of the original f (on weighted/size-biased laws see Patil and Rao, 1978; Vardi, 1982).

In particular, with bracket edges $0 = a_0 < a_1 < \cdots < a_{J-1} < a_J = \infty$,

$$s_j(\boldsymbol{\theta}) = \frac{1}{\mu_1(\boldsymbol{\theta})} \int_{a_{j-1}}^{a_j} x\, f(x;\boldsymbol{\theta})\, dx = \Pr_{g_\theta}\{a_{j-1} \le X < a_j\}, \qquad j = 1, \ldots, J, \tag{3.9}$$

and, under correct specification, the empirical shares satisfy $\widehat{s}_j \xrightarrow{p} s_j(\boldsymbol{\theta})$ as $N \to \infty$. Indeed,

$$\widehat{s}_j = \frac{\frac{1}{N}\sum_{i=1}^N X_i\, \mathbb{1}\{a_{j-1} \le X_i < a_j\}}{\frac{1}{N}\sum_{i=1}^N X_i} \xrightarrow{p} \frac{\mathbb{E}[X\, \mathbb{1}\{a_{j-1} \le X < a_j\}]}{\mathbb{E}[X]} = s_j(\boldsymbol{\theta}),$$

where the convergence follows from the law of large numbers and the continuous mapping theorem.

This difference has several methodological implications:

- **Identification.** With wealth shares $s_j(\boldsymbol{\theta})$, the scale $\mu_1(\boldsymbol{\theta})$ cancels, as the PLN truncated first moment over (a_{j-1}, a_j) – abbreviated as $\mu_{\mathrm{PLN}}(a_{j-1}, a_j; \mu, \sigma, \alpha, \tau)$ – can be written piecewise as

$$\mu_{\mathrm{PLN}}(a_{j-1}, a_j; \boldsymbol{\theta}) = \begin{cases} c\, \dfrac{\mu_{\mathrm{LN}}(a_{j-1}, a_j; \boldsymbol{\theta}_{\mathrm{B}})}{F_{\mathrm{LN}}(\tau; \boldsymbol{\theta}_{\mathrm{B}})}, & a_j \le \tau, \\ (1-c)\ \mu_{\mathrm{PA}}(a_{j-1}, a_j; \alpha, \tau), & a_{j-1} \ge \tau, \\ c\, \dfrac{\mu_{\mathrm{LN}}(a_{j-1}, \tau; \boldsymbol{\theta}_{\mathrm{B}})}{F_{\mathrm{LN}}(\tau; \boldsymbol{\theta}_{\mathrm{B}})} + (1-c)\ \mu_{\mathrm{PA}}(\tau, a_j; \alpha, \tau), & a_{j-1} < \tau < a_j, \end{cases} \tag{3.10}$$

with $\boldsymbol{\theta} = \{\mu, \sigma, \alpha, \tau\}$, $\boldsymbol{\theta}_{\mathrm{B}} = \{\mu, \sigma\}$, c is the splice weight implied by density continuity at τ as defined in Equation (3.7), $\mu_{\mathrm{LN}}(L, U; \boldsymbol{\theta}_B)$ and $\mu_{\mathrm{PA}}(L, U; \alpha, \tau)$ are the the LN and Pareto truncated first moment over $[L, U]$, respectively, and can be rewritten as

$$\mu_{\mathrm{LN}}(L, U; \boldsymbol{\theta}_{\mathrm{B}}) = e^{\mu+\sigma^2/2}\left[\Phi\left(\frac{\ln U - \mu - \sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln L - \mu - \sigma^2}{\sigma}\right)\right],$$

$$\mu_{\mathrm{PA}}(L, U; \alpha, \tau) = \begin{cases} \dfrac{\alpha\,\tau^{\alpha}}{1-\alpha}\left(U^{\,1-\alpha} - L^{\,1-\alpha}\right), & U < \infty,\ \alpha \neq 1, \\ \dfrac{\alpha\,\tau^{\alpha}}{\alpha-1}\,L^{\,1-\alpha}, & U = \infty,\ \alpha > 1. \end{cases}$$

Hence, the wealth share can be written as the ratio of a partial to a total first moment,

$$s_j(\boldsymbol{\theta}) = \frac{\int_{a_{j-1}}^{a_j} x\, f_{\mathrm{PLN}}(x; \boldsymbol{\theta})\, dx}{\int_0^{\infty} x\, f_{\mathrm{PLN}}(x; \boldsymbol{\theta})\, dx} = \frac{\mu_{\mathrm{PLN}}(a_{j-1}, a_j; \boldsymbol{\theta})}{\mu_1(\boldsymbol{\theta})}.$$

Therefore, when working with shares $s_j(\boldsymbol{\theta})$, the overall scale $\mu_1(\boldsymbol{\theta})$ enters only through this normalisation. Identification still relies on shape parameters and the placement of exogenous cut-offs $\{a_j\}$. For LN, $s_j(\boldsymbol{\theta})$ depends on (μ, σ) via transformed limits; for PLN, (α, τ) enter through the tail block and the matching at τ. If τ lies deep inside a wide bracket, local identification weakens; in such cases, profiling τ on a grid (as in splicing practice) is natural and aligns with Scollnik (2007).

- **Likelihood vs moments.** Grouped-data MLE in the literature treats frequencies as (multinomial) outcomes. In our case, we do not observe counts, but range-sum shares; a multinomial likelihood is therefore a misspecification for the observed object. Moment-based methods (GMM) match the exact bin-level identities implied by g_θ, avoiding pseudo-likelihood assumptions on an artificial sampling law for $\widehat{\mathbf{s}} = (\widehat{s}_1, \ldots, \widehat{s}_J)^{\mathsf{T}}$.

- **Sampling variability and weighting.** Many studies rely on survey microdata to compute grouped inputs and hence have well-defined sampling variances. The BSR are administrative aggregates; $\widehat{\boldsymbol{s}}$ is effectively noise-free at a point in time but subject to structural misspecification. A conservative approach is to use robust covariance for extremum estimators and, when feasible, to exploit over-identifying information (e.g., repeated cross-sections over time) to estimate an empirical weighting matrix.

- **Exogenous vs endogenous bins.** Decile-based inputs (endogenous bins) simplify some identification problems but require solving for quantiles at each iteration. The BSR use fixed currency thresholds (exogenous bins): simpler numerically, but the information content depends on how brackets overlap with the lognormal body and the potential Pareto tail. Given that the top bin has a lower bound of €500,000 up to 2022, and of €1,000,000 from 2022 onwards for deposits and €5,000,000 from 2022 onwards for securities accounts, it is

reasonable to suppose the Pareto threshold is enclosed in the previous bin or somewhere around the lower bound of the top tail bin.[9]

- **Open-top bin and tail testing.** Many administrative tabulations, including BSR, have an open-top bracket. Tail parameters are primarily informed by the mass and first moment in the top one or two brackets; methods that explicitly focus on binned-tail fit (e.g., LR/KL comparisons of Pareto vs lognormal in the upper bins) complement global fit criteria and connect to the tail-testing literature on binned data (Virkar and Clauset, 2014).

Generalised Method of Moments. The first methodology considered is the Generalised Method of Moments (GMM, henceforth). GMM estimates parameters by equating sample and model moments implied by the postulated distribution (Hansen, 1982; Newey and McFadden, 1994).[10]

In general, suppose there is a vector $h(x;\boldsymbol{\theta})$ such that its expectation satisfies $\mathbb{E}[h(x;\boldsymbol{\theta}_0)] = 0$. The GMM estimator is the one that minimises the squared Euclidean distance of sample moments of the function h with a suitable positive-definite weighting matrix (for further discussion, see Section 3.3.2).

For grouped income distributions, Hajargasht and Griffiths (2013) derive moment conditions for PLN and double Pareto-lognormal (dPLN) using shares or grouped means and estimate parameters by GMM across a large cross-country sample (Luxembourg Income Study). They report small-sample robustness and good fit, and show how identification weakens when bins are wide or the tail threshold lies within a bin – points that translate directly to BSR, where bracket widths are exogenous and an open-top bin is present, though from 2022 bin sizes shrink noticeably and it can be argued that the Pareto tail threshold lies before the lower bound of the open-top bin.

Griffiths and Hajargasht (2012) propose efficient GMM constructions for grouped moments (including shares and means), showing precision gains over ad hoc minimum-distance criteria in simulations calibrated to empirical deciles; their weighting-matrix discussion is relevant here because BSR provide high-fidelity aggregates but limited information on sampling variability. Earlier and companion pieces (Hajargasht et al., 2012) set out share and mean-based moment conditions for mixtures and PLN, discuss over-identification (more moments than parameters), and provide guidance for open-ended top bins – all features mirrored in BSR range-sum data.

The actuarial splicing literature provides complementary guidance: Scollnik (2007) splice a lognormal body with a Pareto tail and emphasise practical estimation via grid search over the splice threshold (profiling), continuity constraints at the threshold, and the role of the top bin – tools that are directly applicable when PLN

[9] For the majority of EU countries in the DWA, the threshold for the Pareto tail on net wealth is set at €1 million. See Section 2.2.7 "Adding rich households" in the *DWA Methodological Note*, https://data.ecb.europa.eu/.../DWA_Methodological_note.pdf.

[10] Further details on the GMM theory are provided in Section 3.3.2.

is fitted to BSR shares.

For BSR, GMM is well-suited. The moment equations needed for identification are observed (up to normalisation) and are exact functions of $\boldsymbol{\theta}$ under the assumed $f(\cdot;\boldsymbol{\theta})$. GMM neither requires a sampling model for counts nor an "effective sample size". In practice, we use as weighting matrix either (i) the identity weight as a baseline, or (ii) an empirical weight estimated from BSR time slices to capture cross-bin co-movement.

Maximum Composite likelihood method. In the absence of counts, one can treat the vector of wealth shares $\widehat{\mathbf{s}} = (\widehat{s}_1, \ldots, \widehat{s}_J)^\mathsf{T}$ as "proportions" and maximise a composite likelihood (also referred to as pseudolikelihood, approximate-likelihood, and quasi-likelihood) as if they arose from a multinomial model with cell probabilities $s_j(\boldsymbol{\theta})$:

$$\ell_c(\boldsymbol{\theta}) \;=\; \sum_{j=1}^{J} \widehat{s}_j \log s_j(\boldsymbol{\theta}), \tag{3.11}$$

which is equivalent to minimising the cross-entropy (Kullback-Leibler divergence) from $\widehat{\mathbf{s}}$ to $\mathbf{s}(\boldsymbol{\theta}) = (s_1(\boldsymbol{\theta}), \ldots, s_J(\boldsymbol{\theta}))^\mathsf{T}$. Since BSR are administrative aggregates rather than random counts, (3.11) is a working likelihood.

Composite likelihood may be seen as misspecified likelihoods, consequently statistical inference should be based on the Godambe information matrix (also known as the sandwich information matrix), rather than the Fisher information (Lindsay, 1988; Varin et al., 2011):

$$\mathrm{Var}\Big(\widehat{\boldsymbol{\theta}}_{\mathrm{CL}}\Big) \;\approx\; \mathrm{G}\,(\boldsymbol{\theta})^{-1} = \Big(H(\boldsymbol{\theta})\, J(\boldsymbol{\theta})^{-1}\, H(\boldsymbol{\theta})\Big)^{-1}\,, \tag{3.12}$$

where:

$$H(\boldsymbol{\theta}) = \mathbb{E}_\theta\left[-\nabla^2 \ell_c(\boldsymbol{\theta})\right],$$
$$J(\boldsymbol{\theta}) = \mathrm{var}_\theta\left\{\nabla \ell_c(\boldsymbol{\theta})\right\},$$

with $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ estimated at the maximum composite likelihood estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{CL}}$. In the PLN case, continuity of the spliced density at τ and the open-top bin mean constraint enter only through the smooth integrals defining $s_j(\boldsymbol{\theta})$, so (3.11) remains straightforward to evaluate.

Composite likelihood criteria are widely used to fit parametric size distributions to grouped inputs when sampling models are uncertain. Within the GB2 and related literatures, empirical studies often approximate decile shares by a multinomial and maximise (3.11) (or, equivalently, minimise cross-entropy) to obtain stable estimates, then correct uncertainty using sandwich variance; Graf and Nedyalkova (2014) build on such practice and provide estimated parameters and indicators assuming a GB2 distribution, although they use microdata rather than aggregates.

Although not tailored to BSR, the methodology carries over because $s_j(\boldsymbol{\theta})$ are smooth functions of $\boldsymbol{\theta}$ even under PLN, and $\widehat{\mathbf{s}}$ are compositional. Compared with GMM applications such as Hajargasht and Griffiths (2013) and Griffiths and Hajargasht (2012), composite likelihood typically attains similar in-sample fit but may understate uncertainty if one treats (3.11) as a true likelihood with a fictitious sample size n; this is remedied by the Godambe variance in (3.12) and, where available, by exploiting repeated BSR releases to estimate $J(\boldsymbol{\theta})$ empirically.

For BSR range-sum shares, quasi-likelihood via (3.11) is attractive computationally (smooth, concave in many cases, amenable to standard optimisation) and is useful for initialisation and benchmarking. However, because BSR do not provide a sampling variance for $\widehat{\boldsymbol{s}}$, estimates of an adequate approximation of $J(\boldsymbol{\theta})$ might be non-trivial. In short, composite likelihood is feasible and pragmatic, but for final estimation and testing, we prefer GMM because it matches the exact moment identities implied by the BSR design.

Minimum divergence method. Minimum-divergence (minimum-distance) estimators choose $\boldsymbol{\theta}$ to minimise a discrepancy between $\widehat{\boldsymbol{s}}$ and $\boldsymbol{s}(\boldsymbol{\theta})$. A widely used class is the Cressie-Read power-divergence family (Cressie and Read, 1984), which – for $\lambda \in \mathbb{R}$ – is defined as:

$$D_\lambda(\boldsymbol{\theta}) \;=\; \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{J} \widehat{s}_j \left[\left(\frac{\widehat{s}_j}{s_j(\boldsymbol{\theta})} \right)^{\lambda} - 1 \right], \tag{3.13}$$

with limits D_0 (Kullback-Leibler) as $\lambda \to 0$, D_1 (Pearson χ^2) at $\lambda = 1$, and $D_{-1/2}$ related to the Hellinger distance (see Basu et al., 2011; Chen et al., 2004). The minimum-divergence estimator solves:

$$\widehat{\boldsymbol{\theta}}_\lambda \;=\; \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\theta}} \; D_\lambda(\boldsymbol{\theta})\,. \tag{3.14}$$

Weights can be introduced to emphasise particular bins (e.g., the open-top bracket), and composite constraints (e.g., continuity at a PLN splice) enter only through $s_j(\boldsymbol{\theta})$.

Power-divergence criteria have a long tradition in grouped-data inference for income and wealth. Empirical studies such as McDonald and Xu (1995) fit GB2/LN-type distributions to decile shares using D_0 (KL) and D_1 (Pearson), often obtaining fits comparable to more elaborate likelihood-based procedures when the number of bins is moderate. Hajargasht and Griffiths (2013) compare minimum-distance and GMM for PLN/dPLN using grouped shares and show that, while carefully chosen λ can yield competitive performance, GMM tends to be more efficient when valid identifying moments are available. From a robustness standpoint, Chen et al. (2004) discuss how the choice of λ trades off efficiency and insensitivity to model misspecification or local anomalies (e.g., heaping in particular brackets). For the present application, the ability to (*i*) re-weight the open-top bin and (*ii*) run sensitivity checks across $\lambda \in \{0, 1, -\frac{1}{2}\}$ are practical advantages given that BSR bins are exogenous and top-heavy.

Minimum-divergence is directly applicable to BSR wealth shares: it requires no sampling model, handles PLN/LN seamlessly through $s(\boldsymbol{\theta})$, and admits transparent robustness analysis through λ and bin weights. Its main limitation is efficiency when the empirical variability of $\widehat{\boldsymbol{s}}$ is unknown; for this reason, we still deem GMM-based methods preferable.

Final remarks on method choice. All three approaches – GMM, composite likelihood, and minimum divergence – can be implemented with BSR range-sum data as the matched object is the vector of wealth shares by exogenous bracket, i.e. the bin probabilities of the size-biased distribution g_θ in (3.22). In view of the literature and the nature of the data, GMM is the most suitable primary method: it targets the exact bin-level identities induced by truncated first moments and handles PLN splicing by profiling the threshold τ when it falls within a bin (cf. Scollnik, 2007). Quasi-likelihood via (3.11) is computationally convenient and useful for initialisation and benchmarking, but credible uncertainty quantification must rely on the Godambe form (3.12) rather than a fictitious sample size, and the single-dissemination format of BSR data makes variance estimation particularly problematic. Finally, minimum divergence via (3.13) is an effective robustness check, allowing one to prioritise tail fit (open-top bin) and to report sensitivity across divergence choices. Consistent with grouped-data evidence (Hajargasht et al., 2012; Griffiths and Hajargasht, 2012; Chotikapanich et al., 2018), we therefore adopt a GMM/minimum-distance estimator that matches $\widehat{\boldsymbol{s}}$ to $s(\boldsymbol{\theta})$.

3.3.2 GMM – Theoretical framework

This section presents a more in-depth review of the theory behind the Generalised Method of Moments (GMM, henceforth), as it will be the main method of estimation. GMM is typically applied in the context of semi-parametric models, where the parameter of interest is finite-dimensional and the full shape of the data-generating distribution may be unknown; maximum likelihood is therefore not directly applicable. A necessary order condition for identification is that the number of moment functions weakly exceeds the number of parameters, and parameters are then estimated by equating sample and model moments implied by the postulated distribution Hansen (1982); Newey and McFadden (1994).

In general, let $x \in \mathcal{X}$ denote the data, $\boldsymbol{\theta} \in \boldsymbol{\theta} \subset \mathbb{R}^p$ the parameter vector, and $h(x;\boldsymbol{\theta}) \in \mathbb{R}^q$ a vector of moment functions such that

$$\mathbb{E}[h(X;\boldsymbol{\theta}_0)] \;=\; 0\,, \qquad q \geq p, \tag{3.15}$$

for the true parameter $\boldsymbol{\theta}_0$.

Given i.i.d. observations $\{x_i\}_{i=1}^N$, define the sample moments:

$$\bar{h}_N(\boldsymbol{\theta}) \;=\; \frac{1}{N}\sum_{i=1}^{N} h(x_i;\boldsymbol{\theta})\,. \tag{3.16}$$

The GMM estimator minimises a quadratic form in (3.16):

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_N &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\theta}} Q_N(\boldsymbol{\theta}), \\ Q_N(\boldsymbol{\theta}) &= \bar{h}_N(\boldsymbol{\theta})^{\mathsf{T}}\, \widehat{\Omega}_N\, \bar{h}_N(\boldsymbol{\theta}), \end{aligned} \tag{3.17}$$

where $\widehat{\Omega}_N$ is a symmetric positive-definite weight matrix.

Under standard regularity and a rank condition on the Jacobian $G = \mathbb{E}[\frac{\partial}{\partial\boldsymbol{\theta}} h(X;\boldsymbol{\theta}_0)^{\mathsf{T}}]$ (full column rank p), $\widehat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}_0$ and:

$$\begin{aligned} \sqrt{N}\,(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) &\Rightarrow \mathcal{N}(0,\ \mathcal{V}(\Omega)), \\ \mathcal{V}(\Omega) &= (G^{\mathsf{T}}\Omega G)^{-1}\, G^{\mathsf{T}}\Omega S \Omega G\, (G^{\mathsf{T}}\Omega G)^{-1}, \end{aligned} \tag{3.18}$$

with $S = \mathbb{E}[h(X;\boldsymbol{\theta}_0)h(X;\boldsymbol{\theta}_0)^{\mathsf{T}}]$ and $\Omega = \widehat{\Omega}_N$ (Newey and McFadden, 1994, Ch. 36).

The optimal weight is $\Omega = S^{-1}$, in which case $\mathcal{V}(S^{-1}) = (G^{\mathsf{T}}S^{-1}G)^{-1}$; in practice, a two-step GMM sets $\widehat{\Omega}_N = I$ to obtain a preliminary estimate and then replaces Ω by a consistent estimate of S^{-1}. With $q > p$, the over-identifying restrictions can be tested via the J-statistic

$$J_N = N\, \bar{h}_N(\widehat{\boldsymbol{\theta}}_N)^{\mathsf{T}}\, \widehat{\Omega}_N\, \bar{h}_N(\widehat{\boldsymbol{\theta}}_N) \xrightarrow{d} \chi^2_{q-p}. \tag{3.19}$$

In the present setting, let $f(x;\boldsymbol{\theta})$ denote the density of account-level balances with finite first moment:

$$\mu_1(\boldsymbol{\theta}) = \int_{\mathbb{R}_+} x\, f(x;\boldsymbol{\theta})\, dx, \tag{3.20}$$

and let $0 = a_0 < a_1 < \cdots < a_{J-1} < a_J = \infty$ be the BSR bracket edges. The BSR deliver range-sum totals:

$$T_j = \sum_{i=1}^{N} X_i\, \mathbb{1}\{a_{j-1} \le X_i < a_j\}, \qquad j = 1, \ldots, J, \tag{3.21}$$

which, after normalisation by $T = \sum_{j=1}^{J} T_j$, yield wealth shares $\widehat{s}_j = T_j/T$.

In expectation, these correspond to the bin probabilities of the size-biased density $g_\theta(x) = xf(x;\boldsymbol{\theta})/\mu_1(\boldsymbol{\theta})$:

$$s_j(\boldsymbol{\theta}) = \Pr_{g_\theta}\{a_{j-1} \le X < a_j\} = \frac{1}{\mu_1(\boldsymbol{\theta})} \int_{a_{j-1}}^{a_j} x\, f(x;\boldsymbol{\theta})\, dx, \qquad j = 1, \ldots, J. \tag{3.22}$$

Collect $\widehat{\boldsymbol{s}} = (\widehat{s}_1, \ldots, \widehat{s}_J)^{\mathsf{T}}$ and $\boldsymbol{s}(\boldsymbol{\theta}) = (s_1(\boldsymbol{\theta}), \ldots, s_J(\boldsymbol{\theta}))^{\mathsf{T}}$, and define the moment vector

$$h(\boldsymbol{\theta}) = \widehat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{\theta}). \tag{3.23}$$

Since $\mathbb{E}[h(\boldsymbol{\theta}_0)] = 0$ by (3.22), a GMM estimator based on (3.17) identifies $\boldsymbol{\theta}_0$ provided a rank condition holds. Because $\sum_{j=1}^{J} \widehat{s}_j = 1$ and $\sum_{j=1}^{J} s_j(\boldsymbol{\theta}) = 1$, one

component is redundant; numerically, one may drop the last component or work with log-ratio contrasts on the simplex (Aitchison transforms) to improve conditioning without altering the target moments.

Two practical aspects follow from (3.23) – (3.19). First, identification rests on shape parameters (the overall scale cancels in the normalisation), on the placement of the exogenous cut-offs $\{a_j\}$, and – under Pareto-lognormal (PLN) – on the tail threshold τ when it falls inside a bin. If τ lies deep within a wide bracket, local sensitivity of $s_j(\boldsymbol{\theta})$ to tail parameters weakens; profiling τ on a grid is therefore natural and consistent with splicing practice. Second, the Jacobian that enters G is available in closed form by differentiation under the integral sign:

$$\begin{aligned} \frac{\partial s_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{1}{\mu_1(\boldsymbol{\theta})^2}\left(\mu_1(\boldsymbol{\theta})\int_{a_{j-1}}^{a_j} x\,\frac{\partial f(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,dx \;-\; \Big(\int_{a_{j-1}}^{a_j} x\,f(x;\boldsymbol{\theta})\,dx\Big)\,\frac{\partial \mu_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right), \\ \frac{\partial \mu_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \int_{\mathbb{R}_+} x\,\frac{\partial f(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,dx, \end{aligned} \tag{3.24}$$

with analogous expressions when the parameter vector includes a splice threshold τ that is profiled rather than estimated jointly. The rank condition for GMM reduces to rank $G = p$ for $G = -\partial \boldsymbol{s}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}^{\mathsf{T}}$ after removing the redundant component.

For estimation, the ideal setting would be a two-step criterion

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\; Q(\boldsymbol{\theta})\,, \\ Q(\boldsymbol{\theta}) &= h(\boldsymbol{\theta})^{\mathsf{T}}\,\widehat{\Omega}\,h(\boldsymbol{\theta}), \end{aligned} \tag{3.25}$$

with $\widehat{\Omega} = I$ at step 1 and an estimated optimal weight matrix at step 2. In practice, BSR are administrative totals that cover the entire population and – particularly after reconciliation to QSA totals in Section 2.3 – exhibit no sampling variance by construction; any rounding or residual coverage error is plausibly small. Moreover, each BSR release provides a single point estimate per half-year, and inter-release movements reflect portfolio reallocation rather than sampling noise, so the variance of the BSR moments cannot be identified from the data. Consequently, an empirical optimal weighting matrix is infeasible, and we implement GMM with $\widehat{\Omega} = I$ throughout.

Two model-specific remarks are in order. For LN, $s_j(\boldsymbol{\theta})$ in (3.22) admit closed-form expressions via truncated lognormal first moments, which improves numerical accuracy. For PLN, continuity at the splice point and finiteness of the first moment require $\alpha > 1$ and $\tau > 0$; with these constraints, (3.22) decomposes into a lognormal contribution for $[a_{j-1},\, \min\{a_j, \tau\}]$ and a Pareto contribution for $[\max\{a_{j-1}, \tau\},\, a_j]$, both available in closed form, and the Jacobian in (3.24) follows by piecewise differentiation.

3.3.3 Oracle recovery design (known DGP)

We first validate the numerical performance of the GMM estimator in an "oracle" setting where the data-generating process (DGP) coincides with the family used

in estimation. Specifically, we consider either a LN or a PLN distribution for the account total X, with an unknown splice threshold u under PLN. Synthetic BSR range-sum data are created by binning X on the grid $[a_{j-1}, a_j)$ that matches the BSR design, and computing the bin shares s_j and (when needed) bracketed first moments.

Across scenarios, we vary: (*i*) the LN parameters (μ, σ); (*ii*) for PLN, the tail exponent and the threshold u too (reported as both level and bin index); and (*iii*) the brackets used to generate the s_j shares, mimicking pre-2022 and post-2022 deposit brackets. For each scenario, we run B Monte Carlo replications, estimate $\boldsymbol{\theta}$ under correct specification, and summarise bias, variance, root-mean-squared error (RMSE), and nominal credible-interval coverage. The simulation grid is documented in Section 3.5.1; optimiser settings, starting-value grids, and stopping criteria match those used in the real data application to ensure comparability.

For a parameter component (or vector) $\boldsymbol{\theta}$ with truth $\boldsymbol{\theta}_0$, we report

$$\mathrm{RMSE}(\hat{\boldsymbol{\theta}}) \;=\; \Big\{ \frac{1}{B} \sum_{r=1}^{B} \|\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_0\|^2 \Big\}^{1/2},$$

alongside Monte Carlo bias and variance. Central intervals are computed as the empirical 0.025 and 0.975 quantiles of $\{\hat{\boldsymbol{\theta}}_b\}_{b=1}^{B}$, yielding a central 95% interval for the Monte Carlo distribution of the estimator. These diagnostics isolate numerical and identification features of the estimator by construction, since model selection and threshold detection are not involved in the oracle setting.

3.3.4 Final remarks

To conclude, we regard GMM as a natural and effective choice for this setting: it targets the exact bin-level identities implied by BSR range-sum shares and accommodates the Pareto-lognormal splice via profiling when needed. We postpone numerical results on an "oracle" setting in Section 3.5.1, and report diagnostic on simulated data only, as model choice is crucial in the real data setting. In the following section, we develop a data-driven procedure to assess whether a pure lognormal (LN) or a Pareto-lognormal (PLN) provides the appropriate marginal specification, and report full results – combining model testing and parameter estimation – on both simulated data and real data. As noted earlier, the real-data application is currently limited to deposits at the account level; results for deposits at the household level and securities-account components are deferred until after Chapters 4–5.

3.4 Testing Pareto tail presence

Following the parameter estimation, this section addresses the issue of whether a Pareto upper tail is required for each instrument and reference period, given the BSR range-sum inputs. As discussed in Section 3.2, the presence of a Pareto tail seems plausible overall, though there is no direct evidence from literature for

the single instrument components – literature focus mostly on aggregated wealth or income. Moreover, some instrument holdings (e.g., deposits) might be diluted enough to be adequately captured by a plain lognormal (LN, henceforth) distribution.

The aim is to assess, in a replicable manner, whether a simple LN suffices or a lognormal-Pareto (PLN, henceforth) specification is necessary. The BSR provide exogenous brackets and, for each bracket, totals of outstanding amounts; normalised by the overall total, these yield wealth shares by range (see Section 3.3.2 for further insights). In expectation, such shares correspond to bin probabilities under the size-biased law implied by the postulated distribution in (3.22). This feature is central to the present exercise: evidence for a heavy upper tail is amplified in wealth shares and is concentrated in the upper bins, often including an open-top bracket.

In practice, the question is formulated as follows. For a given instrument, suppose LN and PLN have been fitted as in Section 3.3. Under LN, the body and the upper tail are jointly governed by the same parameter vector. Under PLN, the body is lognormal and the upper tail beyond a threshold $\tau > 0$ follows a Pareto law with tail index $\alpha > 1$, with continuity enforced at the splice point. The threshold τ is unknown and, in applications, is profiled over a grid aligned with the BSR brackets. The decision task is to determine whether the additional flexibility afforded by the Pareto tail is empirically required once the model is constrained to reproduce the BSR wealth shares. The procedure developed below uses the same loss function employed for estimation, thereby avoiding discrepancies between estimation and testing frameworks and ensuring that any model improvement is measured on the same object as the one used to estimate the parameters.

The section proceeds as follows. First, it reviews the relevant literature on testing problems with nuisance parameters that appear only under the alternative, and on tail detection with grouped data, evaluating which approaches are most appropriate for BSR range-sum inputs and selecting a primary method accordingly. Second, it sets out the theoretical framework, clarifying how the decision rule is constructed when the threshold is searched over a grid and how calibration is performed in the absence of a natural sampling variance for administrative aggregates. Finally, it reports results on simulated and real data. In line with Section 3.3, the empirical application is, for the time being, limited to account-level deposits; household-level deposits will be tested following BSR data aggregation in Chapter 4. The remaining instruments (bonds, listed shares, and investment fund shares) are currently bundled in the BSR and will be tested once disentangled following Chapter 5. The outcome of this section is a model choice, period by period, that will be carried forward to the following steps of the thesis.

3.4.1 Literature review

This subsection reviews the main strands of literature relevant to assessing whether a Pareto upper tail is required when only grouped information is available. The focus is on testing frameworks with nuisance parameters that appear only under the alternative, on tail detection with binned data, and on procedures compatible

with estimation based on wealth-share moments from BSR range-sum inputs. In line with the parameter-estimation section, attention is restricted to the comparison between a lognormal (LN) specification and a lognormal-Pareto (PLN) splice with threshold $\tau > 0$ and tail index $\alpha > 1$, where τ is unknown and profiled over a grid.

A first theme concerns hypothesis testing when parameters are unidentified under the null or lie on the boundary of the parameter space. Under H_0 (no Pareto tail), the PLN-specific parameters are either not identified (e.g., the threshold τ) or collapse to a boundary case (e.g., zero tail mass). In such settings, classical likelihood ratio or Wald statistics do not follow standard chi-square limits; non-standard asymptotics arise and require alternative calibration. Foundational contributions propose tests that take the supremum (or suitable averages) of pointwise statistics over the nuisance parameter space, with non-standard limiting distributions (Davies, 1987; Andrews and Ploberger, 1994; Hansen, 1996). In particular, when a threshold is searched, the sup-likelihood ratio and related sup-Wald or sup-LM statistics provide a principled approach; their critical values are not of the usual chi-square type and must be obtained either from tabulations specific to the model or by bootstrap. Closely related results in structural change and threshold models, where the change point is unknown, confirm the need for supremum-type statistics or model-averaged combinations (Andrews, 1993; Hansen, 2000). These insights apply directly to LN versus PLN because the splice threshold plays the role of an unidentified nuisance under the null. When parameters are constrained at the boundary, further non-standard results for likelihood ratio tests reinforce the same message (Self and Liang, 1987).

A second theme relates to the calibration of test statistics by bootstrap in the presence of GMM estimation and non-standard nulls. For tests based on GMM criteria, asymptotic critical values may be poor in finite samples, and the distribution of criterion differences can be complicated by over-identifying restrictions and by the presence of nuisance parameters under the alternative. Bootstrap procedures tailored to GMM offer reliable size control and power, even under heteroskedasticity or weak identification (Hall and Horowitz, 1996). In the present context, a parametric bootstrap under the LN null – simulating wealth-share vectors from the fitted LN model using the exact bracket edges and then re-estimating LN and PLN with the same loss – provides a natural way to obtain empirical p-values for a supremum drop in the GMM objective. This keeps the estimation and testing frameworks aligned and respects the administrative nature of the BSR aggregates.

A third strand addresses tail detection with binned or grouped data. Much of the empirical work on power laws leverages individual observations. It relies on maximum likelihood for the tail index and on goodness-of-fit statistics such as the Kolmogorov-Smirnov distance, with careful selection of the lower tail cutoff $x_{\min}$ (Clauset et al., 2009). For binned data, direct application of unbinned techniques is invalid; dedicated procedures show how to compute likelihoods and discrimination statistics using bin integrals, and how to account for the information loss from binning (Virkar and Clauset, 2014). These methods emphasise two points that are particularly relevant here: first, tests should be calibrated by bootstrap because

non-standard features and binning complicate asymptotics; second, the choice of the threshold (here the splice τ) is central and should be selected by a targeted search rather than imposed ad hoc. While Virkar and Clauset (2014) focus on tail-only fits with counts, the same logic carries over when the matched object is a vector of wealth shares, because these are integrals of $xf(x;\boldsymbol{\theta})$ over the brackets and thus bin probabilities under the size-biased law.

Complementary guidance comes from the extremes literature. The theory of regular variation and extreme value methods specifies conditions under which Pareto-type tails arise and provides estimators and diagnostics for tail indices and thresholds (Embrechts et al., 2003; Beirlant et al., 2004). A considerable portion of that theory presumes access to individual exceedances and focuses on estimators such as Hill's index and threshold selection diagnostics (Dekkers et al., 1989; Guillou and Hall, 2001). Even though these methods are not directly applicable to BSR aggregates, they inform the expected behaviour of upper-bin summaries under LN versus Pareto tails: Pareto tails imply that tail probabilities and truncated first moments decay as power laws, whereas a lognormal tail decays much faster. Because the BSR shares are under the size-biased law, this contrast is magnified in the upper bins. In practice, this suggests simple upper-bin diagnostics – such as the ratio of the top-bin share to the model-implied lognormal value – that can be used as supporting evidence alongside a formal criterion-based test.

A fourth element concerns goodness-of-fit and model discrimination with grouped inputs. Minimum-distance procedures based on power-divergence criteria (including Kullback-Leibler and Pearson's chi-square) are widely used to fit parametric size distributions to decile or bracket shares (Cressie and Read, 1984; Basu et al., 2011; Chen et al., 2004). In the present case, those divergences can be used as alternative model-comparison criteria between LN and PLN, particularly if one wishes to emphasise the upper bins through appropriate weights. However, the divergences are not tied to a unique sampling model for the BSR shares and lack a simple null distribution when the threshold is searched; hence, bootstrap calibration is again the most practical route. Likelihood-based discrimination between non-nested models (e.g., Vuong-type tests) is less attractive here because it relies on pseudo-likelihoods for proportions, presumes i.i.d. units, and does not accommodate nuisance parameters present only under the alternative without further adjustment (Vuong, 1989). Composite-likelihood information criteria such as CLIC can be reported as supplementary evidence when quasi-likelihood is used on shares, but they are not decisive in the face of the non-standard null and are secondary in a GMM-based workflow (Varin et al., 2011).

A fifth line of work emphasises the role of size-biasing and the geometry of the bins. When the matched object is the share of total amounts by bracket rather than frequencies, the relevant probabilities belong to the size-biased distribution $g_\theta(x) = xf(x;\boldsymbol{\theta})/\mu_1(\boldsymbol{\theta})$. Classical results on weighted distributions clarify that larger values receive proportionally greater weight, which increases sensitivity to tail mis-specification (Patil and Rao, 1978). This has two practical implications for testing. First, the top bin (often open) carries disproportionate information for tail

parameters such as α; its model-implied share under LN is a natural benchmark and departures are readily interpretable. Second, identification improves when the splice threshold lies below the top-bin edge so that more than one bin contains tail mass, a point echoed in the binned power-law literature (Virkar and Clauset, 2014). These data features justify the use of tail-focused moment checks as secondary diagnostics and motivate grid-search strategies for the threshold that respect the exogenous bracket edges.

Finally, within the specific domain of spliced lognormal-Pareto models, the actuarial science literature provides constructive estimation strategies that extend naturally to testing. Splicing a lognormal body to a Pareto tail with continuity at the splice point leads to tractable expressions for bracket integrals and to practical threshold profiling schemes (Scollnik, 2007). While the primary focus is estimation, the same architecture underlies a criterion-drop test between LN and PLN: the improvement in fit is directly measured on the bracket integrals that define the loss, and the threshold is chosen by search rather than fixed a priori. In a BSR context, this ensures that the testing procedure is fully aligned with the estimation engine and exploits the exact object that the data identifies.

To summarise, the literature points to three requirements that shape the present approach. First, because the PLN parameters are not identified under the LN null and the threshold is profiled, a supremum-type statistic is required for valid discrimination. Second, because the matched object is a wealth-share vector and no sampling model for counts is available, a parametric bootstrap under the LN null is appropriate for calibration; GMM-based bootstraps have strong support in the econometrics literature (Hall and Horowitz, 1996). Third, because size-biasing concentrates information in the upper bins, it is informative to accompany the global decision with tail-focused diagnostics that can be interpreted by non-specialists and checked across time. In view of these points, the primary method adopted here is an empirical p-value for the supremum drop in the GMM criterion from LN to PLN over a grid of thresholds, calibrated by a parametric bootstrap under the LN null. Alternative criteria based on divergences or composite likelihoods are retained as secondary robustness checks, reported for completeness rather than as the basis for the main decision.

3.4.2 Theoretical framework

This subsection formalises a decision rule, grounded in the same GMM loss used for estimation, to assess whether a Pareto upper tail is required once models are fitted to BSR wealth-share moments. The construction accommodates the nuisance splice threshold τ that is present only under the alternative and relies on a parametric bootstrap under the LN null to calibrate a pseudo-p-value. Let $Q(\boldsymbol{\theta})$ denote the GMM objective in (3.25), with weight $\widehat{\Omega}$ held fixed across all fits. Write

$$\begin{aligned} Q_{\mathrm{LN}} &= \min_{\boldsymbol{\theta}\in\boldsymbol{\theta}_{\mathrm{LN}}} Q(\boldsymbol{\theta}), \\ Q_{\mathrm{PLN}}(\tau) &= \min_{\vartheta\in\boldsymbol{\theta}_{\mathrm{PLN}}(\tau)} Q\big(\vartheta,\tau\big), \end{aligned}$$

where $\boldsymbol{\theta}_{\mathrm{PLN}}(\tau)$ enforces lognormal-Pareto splicing with continuity at $\tau > 0$ and $\alpha > 1$. Let $\mathcal{T} = \{\tau_1, \ldots, \tau_K\}$ be a finite grid aligned with BSR bracket edges (Section 3.3.2). The test problem is

$$\begin{aligned} H_0 &: \text{ LN on the whole support} \\ H_1 &: \exists\, \tau \in \mathcal{T} \text{ such that PLN holds with } \alpha > 1. \end{aligned}$$

Under H_0, the splice threshold is unidentified and Pareto mass is on the boundary (zero). In such non-standard settings, pointwise Wald/LR critical values are invalid and threshold search must be accounted for via supremum-type statistics (Davies, 1987; Andrews and Ploberger, 1994; Hansen, 1996; Andrews, 1993; Hansen, 2000; Self and Liang, 1987). For each $\tau \in \mathcal{T}$ define the criterion drop

$$\Delta(\tau) \;=\; Q_{\mathrm{LN}} - Q_{\mathrm{PLN}}(\tau), \tag{3.26}$$

and aggregate across the threshold grid by the supremum:

$$D_{\sup} \;=\; \sup_{\tau \in \mathcal{T}} \Delta(\tau) \;=\; Q_{\mathrm{LN}} - \inf_{\tau \in \mathcal{T}} Q_{\mathrm{PLN}}(\tau). \tag{3.27}$$

By construction, $D_{\sup} \geq 0$ and equals zero if PLN cannot improve the GMM fit at any grid point. The supremum addresses the "search over τ" multiplicity and mirrors sup-LR/Wald constructions for unidentified thresholds (Davies, 1987; Andrews and Ploberger, 1994; Hansen, 1996).

Because $D_{\sup}$ has a non-standard null law (unidentified τ, boundary parameters), we calibrate it by a parametric bootstrap tailored to GMM (Hall and Horowitz, 1996). Fix the bin edges $\{a_j\}_{j=0}^{J}$ implied by the BSR reporting design (exogenous), the weight $\widehat{\Omega}$ used in the empirical fit, and the threshold grid $\mathcal{T}$. Then, for $b = 1, \ldots, B$:

0. **Initialisation.**
   Set $\{a_j\}_{j=0}^{J}$, $\widehat{\Omega}$, and $\mathcal{T}$ as in the empirical fit, and choose the pseudo sample size N^*.

1. **Simulate pseudo range-sum shares under LN.**
   Generate a pseudo sample of size N^*

   $$\{X_i^{(b)}\}_{i=1}^{N^*}$$

   i.i.d. from the fitted lognormal $f(x; \widehat{\boldsymbol{\theta}}_{\mathrm{LN}})$.[11] For each $j = 1, \ldots, J$, compute pseudo range-sum totals

   $$T_j^{(b)} = \sum_i X_i^{(b)} \mathbb{1}\{a_{j-1} \leq X_i^{(b)} < a_j\}$$

   and pseudo shares

   $$\widehat{s}_j^{(b)} = \frac{T_j^{(b)}}{\sum_j T_j^{(b)}}.$$

[11] The choice of N^* affects only the Monte Carlo noise of $\widehat{\boldsymbol{s}}^{(b)}$; it does not enter the statistic's definition because calibration is fully bootstrap-based.

2. **Re-fit LN and PLN on $\widehat{s}^{(b)}$.**
   Compute
   $$Q_{\mathrm{LN}}^{(b)}$$
   and
   $$Q_{\mathrm{PLN}}^{(b)}(\tau)$$
   using the same objective $Q(\cdot)$, weight $\widehat{\Omega}$, bin edges $\{a_j\}$ and grid $\mathcal{T}$ as in the empirical fit.

3. **Form the bootstrap statistic.**
   Set
   $$D_{\mathrm{sup}}^{(b)} = Q_{\mathrm{LN}}^{(b)} - \inf_{\tau\in\mathcal{T}} Q_{\mathrm{PLN}}^{(b)}(\tau).$$

The bootstrap pseudo-p-value is then

$$\widehat{p}_B \;=\; \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\Big\{D_{\mathrm{sup}}^{(b)} \geq D_{\mathrm{sup}}\Big\}. \tag{3.28}$$

Reject H_0 in favour of a Pareto tail when $\widehat{p}_B \leq \eta$ for a pre-specified level η (e.g., 5% or 10%). By construction, the calibration replicates the full estimation/testing pipeline (objective, weight, bin geometry, threshold search), which is essential when the null is non-standard (Andrews and Ploberger, 1994; Hansen, 1996).

Remarks on implementation and power.

(i) **Alignment with the estimation target.** Using the GMM loss on wealth shares ensures that improvements are measured on the same object that identifies the parameters (size-biased bin probabilities, see Equation (3.22)); no pseudo-likelihood on artificial counts is required.

(ii) **Unidentified threshold under H_0.** The supremum in Equation (3.27) properly accounts for profiling over τ; bootstrap critical values capture the search over $\mathcal{T}$ by construction (Davies, 1987; Andrews and Ploberger, 1994).

(iii) **Size and finite-sample behaviour.** GMM-adapted bootstraps provide reliable size control for criterion-based tests even with heteroskedasticity/weak identification (Hall and Horowitz, 1996).

(iv) **Open-top bin and sensitivity.** Since the matched probabilities are under the size-biased law $g_\theta(x) = xf(x;\boldsymbol{\theta})/\mu_1(\boldsymbol{\theta})$, tail misspecification is magnified in upper-bin shares (Patil and Rao, 1978); power is therefore driven by grid points τ that place non-negligible tail mass below the open-top edge, in line with binned power-law diagnostics (Virkar and Clauset, 2014; Clauset et al., 2009).

(v) **Continuity and constraints.** The PLN fit enforces continuity at τ and $\alpha > 1$; these constraints enter smoothly through the bin integrals (Section 3.3.2).

3.5 Results

This section concludes the parameter estimation and tail testing chapter by assessing how the methods perform under controlled scenarios. We proceed in two steps. First, we report oracle results – the true family (LN or PLN) is known – on simulated BSR range-sum data. The objective is to document numerical performance, focusing on bias, RMSE, and (where available) confidence-interval coverage, abstracting from model selection. Second, we evaluate the full pipeline – tail detection, model choice, and re-estimation under the selected specification – in a blind simulation setting in which the DGP is unknown to the estimator (Section 3.5.1). The section ends with the real-data application to account-level BSR deposits. From a computational perspective, conditional on a chosen specification, parameter estimation is effectively immediate, whereas the bootstrap-based tail test is the main cost driver; in the real-data application it takes roughly 15 minutes per wave under the reference implementation, depending on when the sequential stopping rule triggers.

3.5.1 Simulated data

Oracle recovery (known DGP)

We construct twenty scenarios; the full grid and labels (S 01–S 20) are reported in Appendix B, Table B.5.

Two scenarios are LN with fixed mean and standard deviation, each paired with one of the two bracket designs that mirror BSR deposits before and after 2022 (see Section 2.3, Table 2.2). The remaining eighteen are PLN, combining the same mean and standard deviation with three values of the tail exponent $\alpha \in \{1.5, 2.0, 2.5\}$, three threshold levels $\tau \in \{250{,}000,\ 350{,}000,\ 600{,}000\}$, and the two bracket designs. The choices for α span very heavy to relatively light tails. The τ values are chosen to probe identification under different degrees of top-end information: $\tau_0 = 250{,}000$ lies below the open-ended top class under both bracket schemes; $\tau_0 = 350{,}000$ lies in the upper range and, under the pre-2022 design, falls within the $[250\text{k}, 500\text{k})$ class; $\tau_0 = 600{,}000$ is included as a deliberate stress test, as it lies inside the open top class $[500\text{k}, \infty)$ under the pre-2022 design. In this configuration, range-sum moments carry limited information to localise the splice point within the open-ended interval, so weaker τ recovery should be expected even under correct specification. The post-2022 design, which further splits the upper range, is typically more informative for τ at high thresholds, although weak identification may persist when the tail mass above τ_0 is negligible. The lognormal location and scale are fixed at $\mu = 8.5$ and $\sigma = 1.5$. Values are chosen to mimic the empirical distribution of deposits and to ensure comparability with the real-data application.

Each scenario is evaluated over $B = 100$ Monte Carlo replications. In each replication, BSR-like bracket shares are computed from a random sample of $N^* = 2{,}000{,}000$ draws from the target distribution. This design both mimics the scale of the application and introduces sampling variation – more pronounced under heavier tails – that is informative about finite-sample behaviour in practice. (Deposit accounts

in Italy are on the order of 50–60 million per wave by HFCS estimates; we set the replication sample to be an order of magnitude smaller to induce more pronounced sampling variation.) For μ, σ, and (in PLN) α, we summarise Monte Carlo bias, variance, and RMSE; results are reported in Figure 3.2.

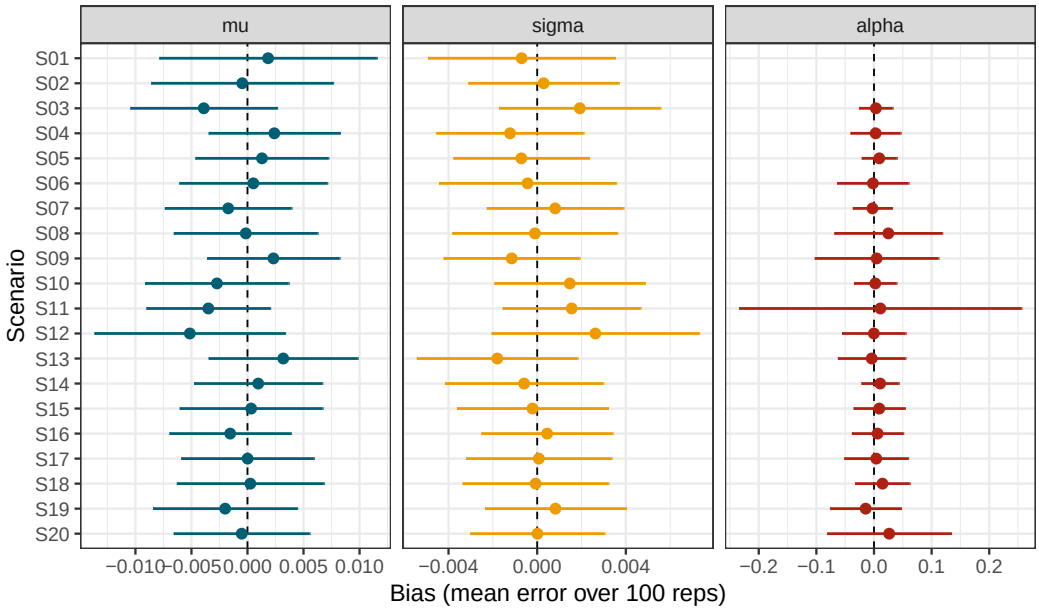

Figure 3.2. Bias and Monte Carlo spread by scenario and parameter (oracle DGP).
Notes: Points show mean estimation error (bias) over $B = 100$ replications; whiskers are ± one standard deviation; the dashed line marks zero bias. LN reports (μ, σ); PLN reports (μ, σ, α). Scenarios refer to Table B.5.

The procedure is highly reliable for the body parameters. They are recovered almost perfectly in all scenarios, with average bias at most 0.005 for μ (Scenario 12) and 0.0026 for σ (Scenario 12). The tail parameter α exhibits slightly larger errors, although the average bias over 100 replications is close to zero in each scenario. Dispersion is higher than for the body parameters but remains limited (the bias magnitude is at most 0.2 times the Monte Carlo standard deviation). Scenarios that perform worse are those with pre-2022 breaks and high τ values – which is reasonable, given that pre-2022 bins carry less information on the top tail and that higher τ both pushes the threshold deep into the open top bin and reduces the tail weight. Overall, the GMM estimator is well-suited to this task, yielding near-unbiased point estimates with as few as $B = 100$ replications. RMSE patterns are summarised in Table 3.1, which reports the relative RMSE (rRMSE, in per cent) by scenario and parameter. We define

$$\mathrm{rRMSE}(\boldsymbol{\theta}) \;=\; 100 \times \frac{\mathrm{RMSE}(\boldsymbol{\theta})}{|\boldsymbol{\theta}_0|},$$

where $\boldsymbol{\theta}_0$ denotes the scenario truth for the relevant parameter (for LN scenarios, rRMSE is reported only for (μ, σ)). Consistent with the identification geometry discussed above, rRMSE remains uniformly small for the body parameters, while tail

accuracy is more sensitive to the bin design and to scenarios with limited upper-tail mass. In particular, rRMSE for α is highest in scenarios where the effective tail weight is smallest, so that tail parameters are weakly informed by range–sum shares even under correct specification. In these scenarios, rRMSE is uniformly small for the body parameters: it is at most 0.12% for μ and 0.36% for σ, confirming near-exact recovery of the lognormal body at the BSR resolution. Tail accuracy is lower but remains moderate: rRMSE for α is below 5% in most scenarios, with larger values concentrated in the weakest-tail cases (notably S 09 and S 11).

Table 3.1. Relative RMSE (per cent) by scenario and parameter (oracle DGP).

Scenario	**rRMSE**(μ)	**rRMSE**(σ)	**rRMSE**(α)
S 01	0.12	0.28	-
S 02	0.10	0.23	-
S 03	0.09	0.27	1.94
S 04	0.07	0.23	2.16
S 05	0.07	0.21	1.26
S 06	0.08	0.27	4.11
S 07	0.07	0.21	1.70
S 08	0.08	0.25	3.85
S 09	0.07	0.22	7.13
S 10	0.08	0.25	1.84
S 11	0.08	0.23	9.76
S 12	0.12	0.36	3.66
S 13	0.09	0.27	2.92
S 14	0.07	0.24	1.35
S 15	0.07	0.23	3.01
S 16	0.07	0.20	2.24
S 17	0.07	0.22	2.21
S 18	0.08	0.22	3.28
S 19	0.08	0.22	3.14
S 20	0.07	0.20	4.40

Notes: rRMSE is defined as $100 \times \mathrm{RMSE}(\boldsymbol{\theta})/|\boldsymbol{\theta}_0|$, where $\boldsymbol{\theta}_0$ is the scenario truth. For LN scenarios (S01–S02) only (μ, σ) are reported. Scenarios refer to Table B.5.

Threshold estimation requires additional care because τ is selected by profiling the GMM objective over a discrete grid. Pointwise RMSE for τ is therefore not the most informative metric. Instead, we report: (*i*) the *hit rate to the snapped truth* (the share of replications in which $\hat{\tau}$ equals the grid point closest to the true threshold τ_0); (*ii*) the mean absolute error to the snapped truth (MAE, snapped); and (*iii*) the hit rates within 5 per cent and 10 per cent of τ_0. Results are summarised in Table 3.2.

Performance varies with α, the location of τ_0 relative to bin boundaries, and the bracket scheme. Identification is strongest when τ_0 falls in interior upper-range classes; it weakens as τ_0 approaches bin edges and, in the pre-2022 design, it deteriorates sharply when τ_0 lies inside the open top class $[500\text{k}, \infty)$. The $\tau_0 = 600{,}000$ pre-2022 scenarios (S 09–S 11) should therefore be interpreted as stress tests that quantify the identification loss induced by open-ended top binning. Post-2022 brackets, which split the top range, typically improve hit rates and reduce MAE, although weak identification may remain when the tail segment above τ_0 carries negligible mass.

Table 3.2. PLN threshold identification: MAE (snapped), hit rate (snapped), and hit rates within 5 per cent and 10 per cent of τ_0, by scenario.

Scenario	MAE (snapped)	Hit rate (per cent)		
		Snapped	Within 5%	Within 10%
S 03	5,300	70.0	86.0	94.0
S 04	14,500	61.0	76.0	85.0
S 05	2,600	77.0	97.0	100.0
S 06	28,400	13.0	34.0	78.0
S 07	95,700	3.0	9.0	25.0
S 08	27,500	15.0	42.0	78.0
S 09	191,000	1.0	2.0	12.0
S 10	156,100	1.0	0.0	0.0
S 11	205,000	1.0	1.0	8.0
S 12	11,000	61.0	76.0	86.0
S 13	53,800	48.0	63.0	83.0
S 14	2,600	76.0	98.0	100.0
S 15	23,100	11.0	37.0	83.0
S 16	156,200	1.0	5.0	8.0
S 17	16,800	20.0	55.0	88.0
S 18	30,300	12.0	63.0	91.0
S 19	168,400	1.0	11.0	14.0
S 20	85,700	6.0	38.0	61.0

Notes: The snapped truth is the τ grid point closest to τ_0. MAE (snapped) is the mean absolute error to the snapped truth over $B = 100$ replications. "Hit rate within 5%/10%" is the share of replications with $|\hat{\tau} - \tau_0| \leq 0.05\,\tau_0$ or $0.10\,\tau_0$, in per cent. Scenarios refer to Table B.5.

To visualise these patterns, Appendix A reports two supplementary figures. Figure A.1 plots the hit rate within 10 per cent of τ_0 against τ_0, with lines for α and panels for the bracket scheme. The display confirms that τ identification weakens near class boundaries and within the open top range, and that the post-2022 partition is systematically more informative at higher thresholds. Figure A.2 provides the complementary mean absolute error to τ_0, showing the same ordering across designs.

One result that may seem counterintuitive is the worse threshold-recovery performance associated with $\alpha_0 = 2$ in Table 3.2. While the worsening of recovery precision as τ_0 increases is expected, one might also expect identification to deteriorate monotonically as the tail becomes lighter (larger α_0). In the present setting, however, the key driver is the BSR bin geometry: for $\alpha_0 = 2$ and the chosen (μ_0, σ_0), a LN and a PLN specification can imply very similar vectors of BSR-style bin shares, so that the profiled GMM objective becomes relatively flat in τ and the splice point is weakly identified.

To make this similarity explicit, let s_j^{LN} denote the implied bin shares under LN with (μ_0, σ_0), and $s_j^{\mathrm{PLN}}(\tau_0)$ the implied bin shares under PLN with the same (μ_0, σ_0), $\alpha_0 = 2$, and splice at τ_0. We summarise the discrepancy between the two implied share vectors by

$$\Delta_{\max}(\tau_0) = \max_{j=1,\dots,J} |s_j^{\mathrm{PLN}}(\tau_0) - s_j^{\mathrm{LN}}|, \qquad \Delta_2(\tau_0) = \Big(\sum_{j=1}^{J} \big(s_j^{\mathrm{PLN}}(\tau_0) - s_j^{\mathrm{LN}}\big)^2 \Big)^{1/2}.$$

Table 3.3 reports $\Delta_{\max}(\tau_0)$ and $\Delta_2(\tau_0)$ for the three τ_0 values and the two bracket schemes. The distances are small across all six cases, indicating that LN and PLN with $\alpha_0 = 2$ are effectively indistinguishable at the resolution of the BSR moments. In such a setting, selecting LN is appropriate by parsimony, and weak identification of τ is expected.

Table 3.3. Distance between LN and PLN implied BSR-style bin-share vectors for $\alpha_0 = 2$.

Breaks	$\boldsymbol{\tau_0}$	$\boldsymbol{\Delta_{\max}(\tau_0)}$	$\boldsymbol{\Delta_2(\tau_0)}$
pre-2022	250,000	0.004	0.015
pre-2022	350,000	0.006	0.020
pre-2022	600,000	0.006	0.019
post-2022	250,000	0.004	0.016
post-2022	350,000	0.003	0.018
post-2022	600,000	0.003	0.018

Notes: Distances are computed between the implied share vectors $\{s_j^{\mathrm{LN}}\}_{j=1}^{J}$ and $\{s_j^{\mathrm{PLN}}(\tau_0)\}_{j=1}^{J}$, where PLN uses (μ_0, σ_0) as in Table B.5, $\alpha_0 = 2$, and density continuity at τ_0. Bracket schemes mirror BSR pre-2022 and post-2022 deposit bins.

Table 3.4 reports the upper-tail probability $\Pr(X > \tau_0)$ for each PLN scenario. In cases where the tail weight is negligible, τ is weakly identified and accuracy metrics deteriorate; by contrast, when the tail segment has non-trivial weight, the profiling step tends to locate τ_0 more precisely. This pattern is consistent with the shape of the profiled loss surface (which is not shown).

Considering the scenarios with $\alpha_0 = 2$, adopting a pure LN specification with the same body parameters yields tail probabilities $1 - F_{\mathrm{LN}}(\tau_0; \mu_0, \sigma_0)$ of approximately

Table 3.4. Upper-tail probability at the threshold τ_0 by PLN scenario: $\Pr(X > \tau_0)$.

Scenario	$\Pr(X > \tau_0)$
S 03 and S 12	0.0057
S 04 and S 13	0.0043
S 05 and S 14	0.0034
S 06 and S 15	0.0031
S 07 and S 16	0.0023
S 08 and S 17	0.0019
S 09 and S 18	0.0010
S 10 and S 19	0.0008
S 11 and S 20	0.0006

Notes: $\Pr(X > \tau_0) = 1 - F_{\mathrm{PLN}}(\tau_0; \mu_0, \sigma_0, \alpha_0, \tau_0)$ under the PLN mixing scheme used here (lognormal body up to τ_0, Pareto tail above). Scenarios refer to Table B.5.

0.004, 0.002, and 0.0007 for $\tau_0 = 250{,}000$, 350,000, and 600,000, respectively. These values are close to those under the PLN specification. By contrast, when $\alpha_0 = 1.5$ (heavier than LN) or $\alpha_0 = 2.5$ (lighter than LN), the implied bin-share vectors differ more materially, and identification improves, with systematic gains under the post-2022 bracket design that splits the upper range.

Full pipeline on simulated data (tail detection and model choice)

In this block, the DGP is known to the researcher but treated as unknown to the estimation procedure, so that tail detection and model choice are part of the experiment. We report model-selection results for the Pareto-tail detection procedure on simulated data. We consider 20 scenarios, matching the design used in the previous section (Table B.5). For each scenario, we apply the decision rule of Section 3.4 with pseudo p-value threshold set at the 5% level, and, conditional on the selected model, estimate the parameters using the same GMM criterion and constraints as in Section 3.3. As expected, parameter-recovery patterns closely mirror those in the previous section once the correct model is selected. Here we focus on model detection performance and on the efficiency gains of the stopping rule used to compute the bootstrap pseudo-p-value.

For each empirical fit, we form the supremum statistic $D_{\sup}$ in (3.27) and calibrate a pseudo-p-value by bootstrap under the LN null as in (3.28). Let k be the number of bootstrap draws (out of B) for which $D_{\sup}^{(b)} \geq D_{\sup}$, so that $\widehat{p} = k/B$. To improve efficiency, we compute after each draw the exact Clopper-Pearson binomial $(1 - \varrho)$ confidence interval $[\underline{p}, \overline{p}]$ for the true tail probability p (Clopper and Pearson, 1934; Brown et al., 2001). We denote the test size (significance level) by $\varrho \in (0, 1)$ and reserve $\alpha > 1$ for the Pareto tail index. Accordingly, Clopper-Pearson confidence

level on the first n iterations is:

$$\underline{p} = \begin{cases} 0, & k = 0, \\ \text{Beta}^{-1}\left(\frac{\varrho}{2};\, k,\, n-k+1\right), & 1 \le k \le n, \end{cases}$$

$$\overline{p} = \begin{cases} 1, & k = n, \\ \text{Beta}^{-1}\left(1-\frac{\varrho}{2};\, k+1,\, n-k\right), & 0 \le k \le n-1, \end{cases}$$

with $\varrho = 0.05$ for a 95% interval. We stop as soon as one of the following conditions is met, with threshold $\eta = 0.05$:

$$\underline{p} > \eta \;\Rightarrow\; \text{select LN (fail to reject)},$$
$$\overline{p} < \eta \;\Rightarrow\; \text{select PLN (reject } H_0).$$

If neither condition is met, sampling continues up to a cap of $B = 250$ bootstrap draws, after which we report $\widehat{p} = k/B$ and its interval and decide by the point estimate relative to η. This design preserves size while reducing computation when evidence is strong in either direction.

Since, conditional on model selection, parameter estimates are already documented in the previous section, we report here only model-selection outcomes and, where relevant, threshold-accuracy metrics for selected-PLN scenarios. Table 3.5 presents, for each scenario, the pseudo-p-value (point estimate and interval), the iteration at which the stopping rule triggered, and the true and selected DGP. Each scenario takes roughly 30 minutes to run, which is greatly reduced if convergence is met before the iterations cap.

The methodology correctly recovers the true DGP in all but two cases (S 11 and S 20), where α_0 is largest and τ_0 is high. In both scenarios, the upper-tail weight is negligible (see Table B.5), implying that a PLN tail yields little improvement over LN. In line with this interpretation, point estimates of μ and σ under the LN selection remain essentially identical to their true values, confirming practical equivalence of the two specifications in these cases: for S 11, $\widehat{\mu} = 8.5$ and $\widehat{\sigma} = 1.5$ with absolute errors 8×10^{-6} and 4×10^{-6}; for S 20, $\widehat{\mu} = 8.5$ and $\widehat{\sigma} = 1.5$ with absolute errors 1×10^{-5} and 4×10^{-6}.

Final remarks

The simulation evidence indicates that the pipeline developed in this chapter – tail detection followed by parameter estimation on BSR range-sum shares – is well suited to the data and decisions required in the application. Across the 20 scenarios of Section 3.5.1, the Clopper-Pearson-based stopping rule delivers size control at $\eta = 0.05$ with sizeable computational savings, and power increases monotonically with upper-tail weight and when τ_0 lies in interior bins. Conditional on the selected model, parameter recovery closely matches the oracle experiments in Section 3.5.1. The only systematic limitation appears when the Pareto segment carries negligible mass or when τ_0 falls effectively in the open top bin, in which case LN and PLN are empirically indistinguishable – as desired for a parsimonious specification.

Table 3.5. Tail detection on simulated data: true and selected model, bootstrap pseudo-p-values, and stopping iteration.

Scenario	Data Generation Process		pseudo-p-value			Iteration
	True	Estimated	Estimate	Lower	Upper	
S 01	LN	LN	1.00	0.82	1.00	19
S 02	LN	LN	1.00	0.82	1.00	19
S 03	PLN	PLN	0.00	0.00	0.05	76
S 04	PLN	PLN	0.02	0.01	0.05	209
S 05	PLN	PLN	0.00	0.00	0.05	76
S 06	PLN	PLN	0.00	0.00	0.05	76
S 07	PLN	PLN	0.02	0.01	0.05	209
S 08	PLN	PLN	0.03	0.01	0.06	250
S 09	PLN	PLN	0.00	0.00	0.05	76
S 10	PLN	PLN	0.03	0.01	0.06	250
S 11	PLN	LN	0.84	0.60	0.97	19
S 12	PLN	PLN	0.00	0.00	0.05	76
S 13	PLN	PLN	0.02	0.01	0.05	250
S 14	PLN	PLN	0.00	0.00	0.05	76
S 15	PLN	PLN	0.00	0.00	0.05	76
S 16	PLN	PLN	0.00	0.00	0.05	76
S 17	PLN	PLN	0.00	0.00	0.05	76
S 18	PLN	PLN	0.00	0.00	0.05	76
S 19	PLN	PLN	0.00	0.00	0.05	76
S 20	PLN	LN	0.53	0.29	0.76	19

Notes: Pseudo-p-values are computed by parametric bootstrap under the LN null; [Lower, Upper] is the exact Clopper-Pearson interval updated sequentially; "Iteration" is the number of bootstrap draws used when the stopping rule triggered (B = 250). Scenario codes as in Table B.5.

In the next section, we apply the same procedure to real data, implementing the full pipeline: reconciliation of concepts, model selection under the LN null, threshold profiling over the bracket-aligned grid, and estimation under the selected specification. We report period-by-period model choices and the associated parameter estimates for the instruments in scope. Moreover, we propose an augmentation of the GMM method, based on BSR available information for deposits.

3.5.2 Real-data extension: augmenting the GMM with a soft mean-of-clients moment

This subsection introduces a real-data augmentation to the bracket-share GMM developed in this chapter. The objective is strengthened by a soft moment on the mean among holders, using administrative counts of deposit clients from the Banking Supervisory Reports (BSR). Because both the aggregate deposits and the number of clients refer to the same reporting perimeter, the resulting anchor is coherent with the bracket shares used in estimation. The moment is imposed softly to account for coverage and definition issues (e.g., exclusion of deposits held abroad and potential multi-bank duplication of clients).

Define, for year t, the administrative anchor for the average amount per holding unit (client):

$$\widehat{\mu}_t^C \;=\; \frac{T_t^{\mathrm{QSA}}}{N_{t,\mathrm{BSR}}^{\mathrm{clients}}}, \tag{3.29}$$

where T_t^{QSA} denotes the sector-account (QSA) total for deposits, equal to the BSR-reconciled aggregate used in Section 2.3.5, and $N_{t,\mathrm{BSR}}^{\mathrm{clients}}$ is the corresponding count of holding clients. Let $g_t(\boldsymbol{\theta}) \in \mathbb{R}^J$ denote the vector of bracket-share moment discrepancies under parameters $\boldsymbol{\theta}$, where J is the number of BSR brackets used in the fit, and let I_J denote the $J \times J$ identity matrix. Let $\mu_t^+(\boldsymbol{\theta})$ be the model-implied mean among positive holders. We then estimate $\boldsymbol{\theta}$ for year t by minimising the following penalised (augmented) GMM loss:

$$\mathcal{L}_t(\boldsymbol{\theta}) \;=\; g_t(\boldsymbol{\theta})^\top I_J \, g_t(\boldsymbol{\theta}) \;+\; \lambda_\mu \left(\frac{\mu_t^+(\boldsymbol{\theta}) \;-\; \widehat{\mu}_t^C}{\widehat{\mu}_t^C} \right)^2, \tag{3.30}$$

with tuning constant $\lambda_\mu > 0$. The relative-error scaling renders the penalty dimensionless and comparable across years. Because T_t^{QSA} and $N_{t,\mathrm{BSR}}^{\mathrm{clients}}$ are administrative aggregates, $\widehat{\mu}_t^C$ does not have a sampling variance in the usual sense; nevertheless, it is subject to conceptual uncertainty stemming from perimeter and definitional differences (e.g., exclusion of deposits held abroad and potential multi-bank duplication of clients). To avoid ad hoc tuning, λ_μ is therefore calibrated to make deviations in the mean anchor and deviations in bracket shares comparable in the objective, reflecting an expert-driven judgement on their relative reliability (we place greater trust in the bracket shares while using the anchor as a stabiliser for the level). The anchor stabilises the level of the distribution while the bracket moments continue to discipline its shape and tail behaviour.

For years aligned with HFCS waves, Table 3.6 reports BSR client counts and the implied average per client used as soft anchors. We also use sector-account totals (equal to the BSR-reconciled totals) to estimate the target mean, as this is a more plausible anchor: missing BSR components are likely concentrated among high-balance units, implying a higher expected value for foreign-held deposits than for domestic holdings. The difference is small in practice but yields a more credible soft target, consistent with the evidence discussed in Section 2.3.5.

Table 3.6. BSR client counts and implied average deposits used as soft anchors.

Year	$\boldsymbol{N}_{t,\mathrm{BSR}}^{\mathrm{clients}}$ (thousands)	$\boldsymbol{T}_{\boldsymbol{t}}^{\mathrm{QSA}}$ (million €)	$\widehat{\boldsymbol{\mu}}_{\boldsymbol{t}}^{\boldsymbol{C}}$ (€)
2010	-	985,657	14,884
2014	70,146	1,092,103	15,569
2016	73,625	1,155,933	15,700
2020	74,742	1,327,853	17,766
2022	71,338	1,389,714	19,481

Notes: Client counts refer to the BSR perimeter; aggregates are QSA totals (equal to the BSR-reconciled totals). Averages are computed as in (3.29). See text for coverage and definition caveats (foreign-held deposits; potential duplication across banking groups).

Since 2010 values for client counts are not reported in BSR, we back-cast the 2010 mean-of-clients anchor using a log-growth rule anchored at 2014. Let $\widehat{\mu}_t^C$ denote the implied average per client as defined in (3.29). For years $y > 2014$ with observed $\widehat{\mu}_y^C$, we compute the annualised log-growth slopes

$$g_y \;=\; \frac{\log \widehat{\mu}_y^C - \log \widehat{\mu}_{2014}^C}{y - 2014}\,,$$

and take their average $\bar{g} = \mathbb{E}[g_y]$. The missing 2010 anchor is then obtained by a log-linear back-cast

$$\widehat{\mu}_{2010}^C \;=\; \widehat{\mu}_{2014}^C \exp(\bar{g}\,(2010 - 2014))\,. \tag{3.31}$$

This approach mirrors multiplicative accumulation in deposits, is robust to level shifts, and is conservative over the short 2010–2014 horizon. Using T_t^{QSA} mitigates numerator-side coverage issues while keeping coherence with the BSR client perimeter.

Two implementation notes follow. First, a residual mismatch may arise because client counts are compiled on the BSR perimeter; we therefore keep λ_μ conservative. Second, if client counts are not fully de-duplicated across banking groups, $N_{t,\mathrm{BSR}}^{\mathrm{clients}}$ may be upward-biased, pushing $\widehat{\mu}_t^C$ downward; this also motivates a soft (rather than hard) implementation of the moment.

In the empirical application, we set $\lambda_\mu = 0.05$ based on a scale-matching rule agreed with BSR-domain experts. Specifically, since the share component equals $\sum_{j=1}^{J}(\widehat{s}_{j,t} - s_j(\boldsymbol{\theta}))^2$, a 10% relative mismatch in the mean anchor contributes $\lambda_\mu \times 0.10^2$

to the loss. We choose λ_μ so that this contribution is comparable to the share loss induced by an average absolute discrepancy of about one percentage point per bracket. This calibration reflects our greater trust in the bracket shares while still discouraging implausible levels for the implied mean among holders. We assess robustness by re-estimating over a small grid around the baseline value and verifying that tail selection and $\widehat{\alpha}$ are stable.

Finally, because the bracket-share system does not hard-enforce the aggregate level, the additional mean moment is non-redundant and improves numerical stability without materially altering the identification of tail parameters. Estimation below adopts (3.30) for the real-data application, while simulation results in the preceding subsections remain based on the baseline criterion to preserve comparability.

3.5.3 Real-data results: deposits (BSR)

We estimate results for the HFCS reference years 2010, 2014, 2016, 2020, and 2022. For each year we apply the full pipeline of Section 3.4: model selection under the LN null with significance $\eta = 0.05$ using a parametric bootstrap calibrated on BSR range-sum shares (threshold grid aligned to bracket edges), followed by parameter estimation under the selected specification with the same GMM objective as in Section 3.3. Relative to simulations, we cap the bootstrap at $B = 500$ replications to tighten the Clopper-Pearson bounds for the pseudo-p-value.[12]

Table 3.7 reports, for each year, the selected DGP (LN vs PLN), the pseudo-p-value $\widehat{p}$, its Clopper-Pearson lower/upper bounds, and the iteration at which stopping occurred. Results indicate a PLN specification in all years at the 5% level. In 2010, the evidence is comparatively weaker: $\widehat{p} < 0.05$ while the upper bound at $B_{\max} = 500$ is 0.06. Nevertheless, the decision still favours PLN at $\eta = 0.05$.

Conditional on the selected model (PLN in every year), we estimate $(\mu, \sigma, \alpha, \tau)$ by GMM for the account-balance random variable X (a generic deposit account drawn from the BSR reporting perimeter) and report the implied mean $\mathbb{E}[X]$, i.e. the model-implied average balance per account. Because, for each year, the BSR supply a single vector of range–sum shares that we treat as fixed administrative aggregates, sampling variability is not defined. Accordingly, we report point estimates only and do not attach standard errors or confidence intervals. Quantifying uncertainty would require additional assumptions on a data-generating mechanism for the aggregates (e.g., measurement-error models or a time-series design), which lie outside the scope of this chapter. Table 3.8 summarises the estimates.

Although the estimated tail index is close to 2 in most years, this does not imply that LN and PLN are observationally equivalent at the BSR resolution. In the simulation discussion, near-equivalence arose for a specific calibration (μ, σ) and bin geometry under which LN and PLN induced almost identical vectors of bracket

[12]The stopping rule is as in Section 3.5.1: after each bootstrap draw the exact Clopper-Pearson interval is updated; stopping occurs when $\underline{p} > \eta$ (select LN) or $\overline{p} < \eta$ (select PLN), or at $B = 500$ otherwise.

Table 3.7. Model selection by year.

Year	Chosen DGP	pseudo-p-value			Iteration
		Estimate	Lower	Upper	
2010	PLN	0.04	0.02	0.06	500
2014	PLN	0.00	0.00	0.05	76
2016	PLN	0.00	0.00	0.05	76
2020	PLN	0.00	0.00	0.05	76
2022	PLN	0.00	0.00	0.05	76

Notes: Pseudo-p-values computed by parametric bootstrap under the LN null with threshold grid aligned to BSR brackets; bounds are exact Clopper-Pearson. The stopping rule and $\eta = 0.05$ decision threshold follow Section 3.4. "Iter." is the number of bootstrap draws used when stopping occurred (B = 500).

Table 3.8. Estimated parameters by year under the selected specification (PLN) and implied mean.

Year	Augmented					Non-augmented				
	μ	σ	α	τ	$\mathbb{E}[X]$	μ	σ	α	τ	$\mathbb{E}[X]$
2010	8.15	1.77	2.09	319,000	15,508	8.61	1.60	2.09	475,000	19,544
2014	8.18	1.77	2.09	331,000	16,120	8.57	1.63	2.07	413,000	19,636
2016	8.09	1.89	2.03	227,000	16,202	8.71	1.63	2.03	281,000	21,719
2020	8.22	1.93	2.03	203,000	18,240	8.71	1.76	1.89	157,000	22,713
2022	8.38	2.19	1.87	106,000	19,582	8.77	1.94	1.87	110,000	23,102

Notes: Parameters estimated by GMM on BSR range–sum shares with fixed weight matrix, density continuity enforced at τ, and $\alpha > 1$. "Implied mean" is computed from the fitted PLN.

shares. In the real-data fits, the estimated body parameters and splice point differ, and LN and PLN generate materially different implied share vectors. For instance, in 2010 the maximum absolute difference between LN- and PLN-implied bracket shares is about 0.05 (five percentage points), and the ℓ_2 distance is about 0.08, compared with values below 0.01 in the near-equivalence simulation calibration.[13]

The pattern of estimates indicates that, as years progress, the implied mean and dispersion increase, and the estimated tail index declines in 2022, suggesting a heavier upper tail in that wave. This pattern also emerges under the non-augmented GMM. For comparison, Table 3.8 also reports estimates obtained without the soft mean anchor. The implied mean is slightly higher without the augmentation. By contrast, the augmented specification introduces a conservative soft anchor on the mean level, which nudges estimates towards a value consistent with sector-account totals and the BSR client perimeter, without materially altering the bracket fit. In

[13]See Table 3.7

light of the perimeter reconciliation and the robustness checks reported above, we view the augmented estimates as the more credible baseline; the non-augmented results remain informative as a sensitivity benchmark.

Figure 3.3 contrasts estimated densities across years under the selected model. The vertical line marks $\widehat{\tau}$ for each year. The visual evidence is consistent with Table 3.8: both the central mass and the tail shift upwards over time, with a more pronounced tail in later years.

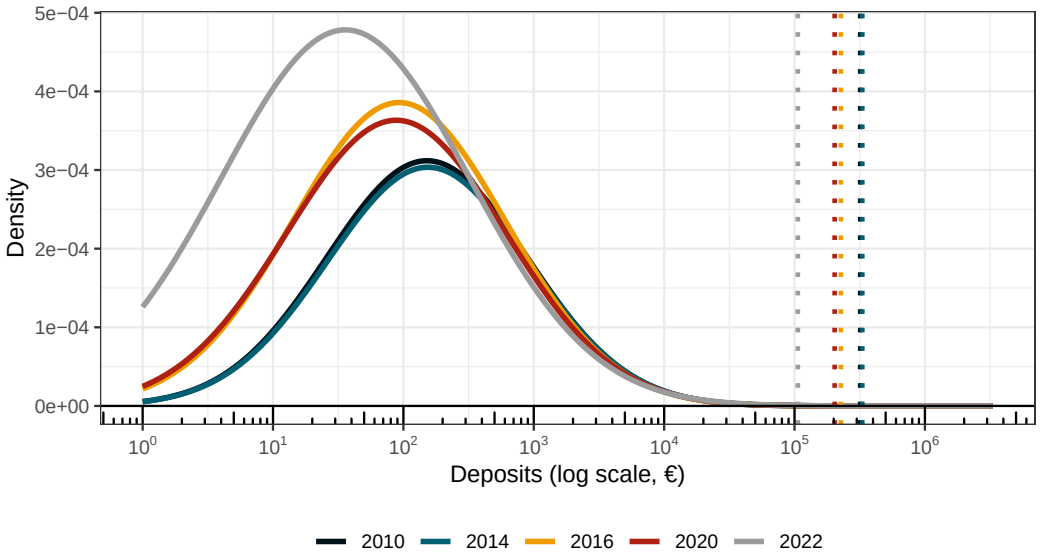

Figure 3.3. Estimated densities by year under the selected specification.
Notes: Each panel shows the fitted density for deposits in the indicated year; the vertical line marks $\widehat{\tau}$. Densities are displayed on log-scale for readability.

To assess fit on the estimation target, Figure 3.4 compares observed BSR wealth shares with model-implied shares by bracket and year. The two series nearly coincide in all years, confirming that the fitted PLN reproduces the range-sum geometry. To visualise the remaining discrepancies without introducing a secondary axis, the figure encodes the absolute bracket-by-bracket difference $|\widehat{s}_j - \widetilde{s}_j|$ through the shading intensity of the observed markers: darker points indicate larger local deviations. As the side-by-side panels show, introducing the mean anchor does not worsen the bracket fit: deviations remain small and local. Under the augmented specification, the largest discrepancy is about 2% and occurs in the lowest bracket; the standard GMM exhibits a mild underfit in the upper brackets, again within 2%. These patterns support the use of the augmented criterion: it preserves the range–sum fit while delivering a more credible level for the mean implied by the fitted distribution.

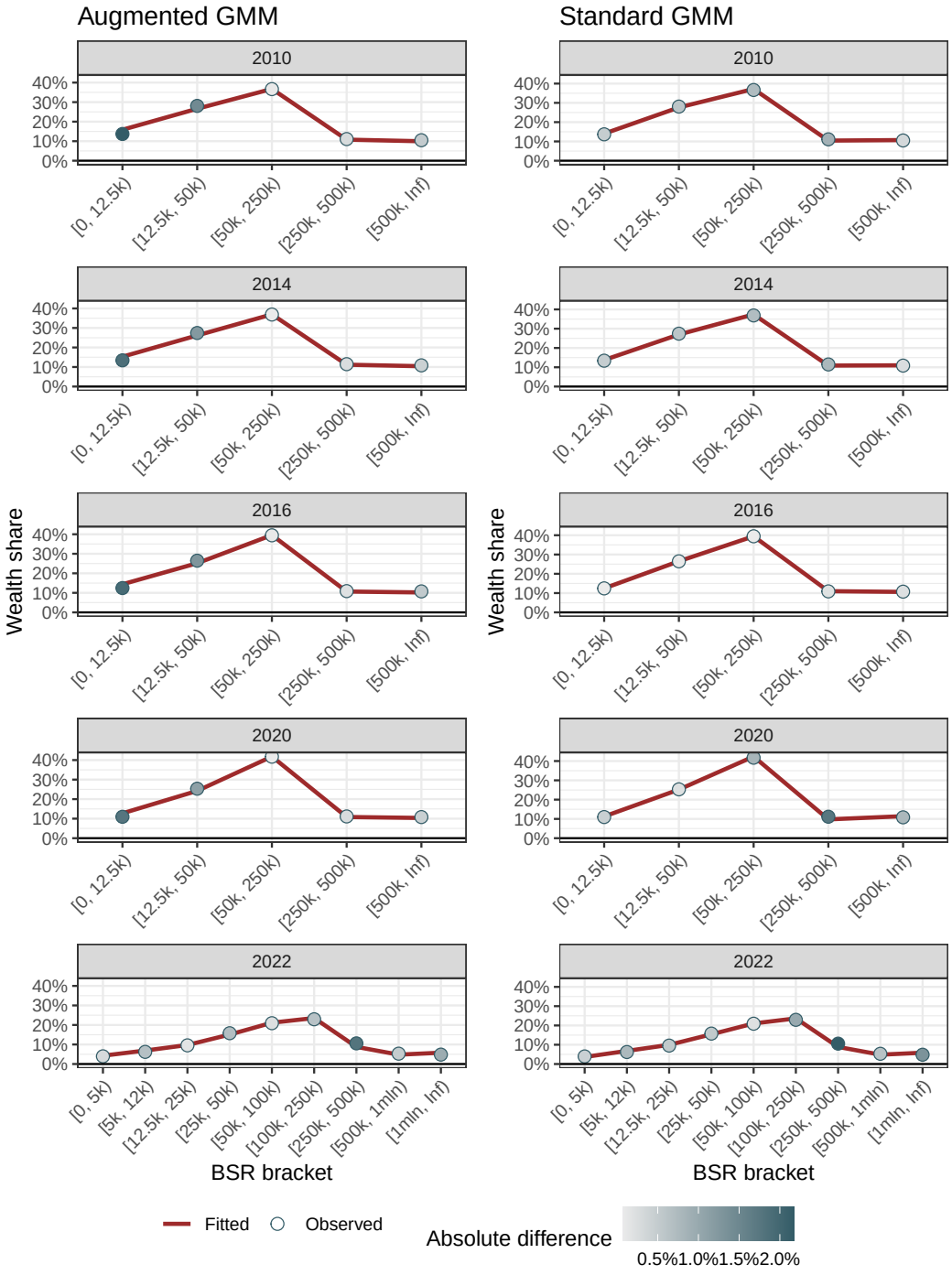

Figure 3.4. Observed vs fitted BSR wealth shares by bracket and year (left: augmented GMM with soft mean-of-clients anchor; right: standard GMM without anchor).
Notes: Points and solid lines denote observed and fitted shares, respectively. Point shading reflects the absolute difference $|\widehat{s}_j - \widetilde{s}_j|$, with darker points indicating larger deviations. Panels are by year; bracket definitions (and totals) follow the BSR perimeter for households and differ in 2022 due to the revised break structure.

3.5.4 Summary and implications

This chapter developed and evaluated a bracket-share GMM for recovering parametric deposit distributions from BSR range-sum data, with a formal test for Pareto tails and a profiling strategy for threshold identification. Evidence from simulations and from real BSR data suggests that, under the implementation choices adopted in this chapter (starting values, constraints, and profiling grid), the estimator behaves stably in the scenarios considered and delivers a practically useful basis for model selection and parameter recovery from bracket shares, while identification of τ weakens when the threshold lies effectively in the open top bin or when upper-tail mass is negligible.

On simulated data with known DGP (oracle recovery), body parameters (μ, σ) are recovered with negligible bias and small dispersion across all scenarios, while the Pareto index α shows modest variability yet remains close to its true value on average (Figure 3.2). RMSE patterns are similarly benign (Table 3.1). Threshold identification is more delicate because τ is selected on a discrete grid: performance depends on the tail weight and on the location of τ_0 relative to bin boundaries. Hit rates and MAE to the snapped truth deteriorate when τ_0 lies effectively in the open top bin and when upper-tail mass is negligible; by contrast, post-2022 brackets, which refine the upper range, improve identification (Table 3.2).

The tail-detection step based on the supremum statistic and a bootstrap pseudo-p-value calibrated under the LN null (with an exact Clopper-Pearson stopping rule) attains the desired size and delivers power that increases with tail mass and interior placement of τ_0. Misclassifications occur only when PLN and LN are practically indistinguishable because the Pareto segment carries vanishing mass, in which case LN and PLN yield almost identical body estimates – an outcome that is both interpretable and desirable for parsimony.

In the real-data application to BSR deposits (HFCS reference years 2010, 2014, 2016, 2020, 2022), the procedure selects PLN in every year at the 5 per cent level (Table 3.7). Estimated parameters display a coherent temporal evolution: location and dispersion increase, and the tail becomes more prominent and heavier (lower α); implied means rise accordingly (Table 3.8). We also introduced a soft moment on the mean among holders, anchored to BSR client counts combined with sector-account totals (with a log-growth back-cast for 2010). This augmentation stabilises the level without compromising the bracket fit: observed and fitted range-sum shares nearly coincide in both anchored and non-anchored specifications, with discrepancies typically within 2 percentage points and concentrated in the lowest brackets (Figure 3.4). The augmented estimates therefore provide a more credible level for average holdings while preserving the range-sum geometry; unanchored results remain a useful sensitivity benchmark (Table 3.8).

Two caveats merit emphasis. First, inference on parameters is reported without sampling errors because BSR range-sum shares are treated as fixed administrative aggregates; uncertainty quantification would require an explicit measurement-error or time-series design. Second, identification of τ is weaker when the threshold lies

near the open top range or when tail mass is vanishingly small; in such cases, LN and PLN are empirically close, and threshold profiling should be interpreted with caution. Nevertheless, across designs and years, the combination of tail testing, grid profiling, and GMM estimation delivers stable, interpretable fits aligned with the BSR bracket structure.

The next chapter addresses the account-to-household aggregation of BSR deposits. The resulting household-level distributions are coherent with the BSR perimeter and provide the inputs required for the methods developed subsequently in the thesis workflow.

Chapter 4

Aggregating BSR accounts to household level

4.1 Introduction

This chapter addresses the unit mismatch between the Banking Supervisory Reports (BSR) and the Distributional Wealth Accounts (DWA) workflow. As discussed in Chapter 3 and Section 2.3, BSR provide aggregates by account and by balance ranges, whereas the DWA calibration operates on household-level holdings. For securities accounts, this difference is limited, as households typically hold at most one account and the account-household distinction is minor. By contrast, deposits display a materially different structure: richer households tend to hold multiple accounts and to diversify balances across them. Consequently, account-based BSR deposit totals cannot be compared directly with survey-based household holdings, nor can they be used as they stand to anchor a multivariate calibration at the household level. In addition, to align survey participation with the BSR scope, the analysis corrects obvious false zeros in HFCS by assigning small positive amounts to households reporting at least one deposit account but zero deposits – this affects only the lower brackets and leaves the distributional shape otherwise unchanged.

One solution for this issue is to split household HFCS deposit holdings into the respective accounts, leveraging the additional information collected in the Italian core survey for HFCS Wave 4. However, despite allowing for a constrained calibration (as presented in our work Neri et al., 2023, and in Section 6.1.1), this solution poses several problems when extending to a multivariate case. Most notably, it is difficult to define the multivariate object of the calibration. Suppose, for example, that a household holds several deposit accounts and a single securities account. Defining the arguments of the constrained calibration is non-trivial, as this household will yield multiple pairs – with unclear optimal decision on the definitions of these couples – or, alternatively, a vector with the dimension of the accounts held by the household – which would imply having vectors of different dimensions depending on the household's number of accounts. This becomes straightforward once households are defined by their total holdings of each instrument, generating one fixed n-dimensional vector per household, with n depending on the number of instruments considered in

the calibration.

In practice, the chapter develops a procedure that reconstructs a household-level deposit distribution coherent with the account-level BSR information while preserving within-household account dependence patterns observed in HFCS/SHIW. Preliminarily, the survey microdata are corrected for false zeros as noted above, so that participation in deposits is measured consistently before reconciliation. The method then proceeds as follows. First, household deposits in the survey are split into accounts by an operational allocation rule, yielding an account-level dataset that retains household identifiers. Second, the account-level target distribution from BSR data is taken from the estimation in Chapter 3 – allowing for a lognormal (LN, henceforth) or a Pareto-lognormal (PLN, henceforth) form. Third, survey account amounts are rank-transformed and mapped to the corresponding quantiles of the estimated distribution, producing debiased account values consistent with the BSR-estimated distribution while preserving the survey rank structure. Fourth, a cluster bootstrap resamples households (and all their accounts) to enlarge the effective sample without breaking intra-household correlation; accounts are then re-aggregated to obtain household-level totals. From these synthetic household totals, wealth shares by BSR brackets are computed directly and, if required, scaled by BSR totals to ensure coherence in levels.

The contribution is twofold. First, the chapter supplies a transparent and implementable bridge from account to household-level quantities for deposits, enabling like-for-like comparisons with HFCS and the construction of household-level targets required by the subsequent multivariate calibration. Second, it isolates the assumptions needed for the translation – most notably the stability of the distribution of the number of accounts per household and of the main-account share – so that their implications can be evaluated systematically and subjected to sensitivity analysis. The preliminary correction of false zeros further reduces spurious non-participation at the bottom of the distribution, improving the coherence of participation rates with the reconciled BSR scope.

The remainder of the chapter is organised as follows. Section 4.2 reviews the relevant literature regarding this method, motivating the proposed approach and providing a solid foundation to develop the methodology. Following, Section 4.3 defines the steps of the applied methodology in more detail. Section 4.4 sets out the theory behind the proposed methodology. Finally, Section 4.6 presents the results on simulated data, real data, and summarises the implications for the calibration framework developed in the following chapters.

4.2 Literature review

This section reviews the strands that motivate the account-to-household reconstruction adopted in this chapter. The organising thread is simple: (*i*) supervisory inputs are range-sum aggregates that identify a size-biased marginal at the account level; (*ii*) we impose that marginal on survey accounts by rank-preserving mapping; (*iii*) we

propagate within-household dependence via cluster resampling; and (*iv*) we exploit Italian survey side-information on account multiplicity and main-account concentration to make the account-to-household aggregation operational. Throughout, parsimony is preferred over richer but weakly identified alternatives, in line with the informational content of range-sum data and with the production constraints of DWA-style pipelines.

BSR aggregates are totals of outstanding amounts by account-value brackets. After normalisation by the total, the resulting shares are amount shares, i.e. bin probabilities under the size-biased (length-biased) law $g(x) \propto xf(x)$ rather than frequencies under $f(\cdot)$, as explained in more detail in Section 3.3. Classical results on weighted and size-biased distributions show that such statistics over-weight large balances and sharpen identification of right-tail behaviour; they also clarify how moment constraints under $g(\cdot)$ relate to those of the underlying $f(\cdot)$ (Patil and Rao, 1978; Vardi, 1982). Related selection-bias analyses further show how to recover features of f from observations drawn under a biased law, a perspective that is directly relevant when the matched objects are truncated first moments by bin (range sums) rather than counts (Vardi, 1985). In practice, two working principles follow for our setting: top-bin shares carry disproportionate information relative to frequency tabulations, and identification of the account-level target distribution should proceed with functionals that are native to size-biased inputs. This perspective underpins Chapter 3, where we estimate the account-level target distribution from range-sums and, in the present chapter, treat it as a marginal constraint to be imposed on micro data.

A natural way to align survey accounts with an external marginal while avoiding parametric modelling of covariates is rank-preserving mapping. Let $U = \widehat{F}_A(X)$ be the (weighted) empirical rank of an account value X in the split survey accounts A, and let F denote the target marginal CDF (here, the BSR-implied marginal). Setting $\widetilde{X} = F^{-1}(U)$ yields a transformed value with distribution F (Rosenblatt, 1952). This monotone quantile mapping is widely used to correct distributional biases where an external, higher-quality marginal is available (e.g., climate downscaling and bias correction; Gudmundsson et al., 2012; Themeßl et al., 2011), and connects to monotone rearrangements in counterfactual analysis (Chernozhukov et al., 2013). Two practical lessons matter here. First, quantile mapping exactly enforces the target marginal and is robust to level mismeasurement in the source data because it operates on ranks. Second, diagnostics should accompany the mapping to detect pathologies in tails or coarse bins; the bias-correction literature stresses careful treatment of extrapolation and open-ended tails (Maraun et al., 2010). In general, mapping should be restricted to the observed support of the target (Hempel et al., 2013). In our case, BSR open top-bin structure avoids off-support extrapolation, as estimated distributions are defined on the same support (non-negative, up to infinity).

Deposits display household-specific multi-account behaviour (number of accounts, concentration in the main account, within-household correlation across accounts). Resampling, therefore, occurs at the household (cluster) level to propagate realistic intra-household dependence into any synthetic micro-population. Survey replication

and bootstrap methods recommend resampling primary sampling units to preserve intra-cluster correlation and design features (Rao and Wu, 1988; Rao et al., 1992; Rust and Rao, 1996; Davison and Hinkley, 1997). When inference targets tail-sensitive functionals or effective sample sizes are modest, bootstrap procedures that respect clustering dominate i.i.d. resampling; wild/cluster variants are recommended for robust uncertainty quantification (Cameron et al., 2008). We adopt that logic directly: each draw selects households and brings along all their accounts; subsequent re-aggregation produces household totals that inherit empirically observed account multiplicity and within-household dependence. In our setting, the bootstrap is primarily for stabilisation (and later for uncertainty display), but its design follows best practice for complex surveys.

A prerequisite for rank mapping is a disciplined, reproducible disaggregation of household deposits into accounts. For Italy, HFCS/SHIW provide both the number of deposit accounts and, since 2020, the share held in the main account. As my co-authors and I have shown in Neri et al. (2023), these variables make the household-to-account split operational, transparent, and easy to benchmark against BSR constraints. Concretely (see Neri et al., 2023, footnote 12 and Section 3.2), we implement:

1. Missing number of accounts and non-zero value of deposits: impute using SHIW patterns by deposit size: 1 account if deposits below €25,000; 2 accounts if €25,000 – €75,000; 3 accounts if above €75,000.
2. One account: assign the full household deposit to a single account.
3. Two accounts: assign the main-account share as estimated with the 2020 survey (mean $\approx 66\%$); allocate the residual to the second account.
4. More than two accounts: allocate 66% to the main account; 20% to the second-largest; split the remaining residual equally across the remaining accounts.

Each synthetic account retains the parent household identifier, ensuring that re-aggregation is well defined and enabling the clustered bootstrap. The imputation for missing counts deliberately trades flexibility for stability; it is in the spirit of constrained, donor-free hot-deck ideas reviewed in Andridge and Little (2010) but anchored to a single predictive dimension (deposit size) that is directly relevant to account multiplicity.

The split is defensible on two grounds. First, it is a constrained disaggregation consistent with compositional data principles (Aitchison, 1982). Second, it respects the identification limits of binned, size-biased aggregates (Patil and Rao, 1978; Vardi, 1982). We favour this design over richer alternatives that are weakly identified in our setting: equal splits conditional on number of accounts are rejected because SHIW documents strong main-account concentration (the 66–20–rest profile), and equalisation would erase this salient feature (Neri et al., 2023); order-statistic partitions require genuine account-level microdata to estimate spacing distributions, which are not identified from household totals and a single main-account share; Dirichlet splits calibrated to the main share add nuisance parameters with limited payoff under

coarse external constraints and, without informative priors or auxiliary margins, risk overfitting the split rather than improving coherence with BSR.

Two design choices follow from the reconciliation aim. First, we test if the account-level distribution after bootstrap matches the BSR-derived distribution. This ensures that household-level bootstrap sampling doesn't introduce biases at the account-level distribution, which could potentially undermine the credibility of the aggregated results. Second, diagnostics are run on the post-mapping account distribution – especially in the top bin – to ensure that the supervisory marginal is matched without generating implausible within-household concentration patterns.

A practical subtlety arises when the BSR-implied target features a heavy right tail (e.g. a Pareto-lognormal splice). Because the upper tail of the quantile function F^{-1} is highly convex under heavy tails, small rank gaps within a household (main vs. minor accounts) can be magnified after rank-based mapping, inflating the main-account share even though the marginal is exactly matched. Similar warnings appear in the bias-correction literature for extremes: monotone quantile mapping preserves marginals but can distort dependence and tail behaviour if used mechanically (Maraun et al., 2010; Gudmundsson et al., 2012; Themeßl et al., 2011). A remedy, consistent with the rearrangement logic of Chernozhukov et al. (2013) and multivariate bias-correction principles (Cannon, 2018), is to introduce a light, rank-scale adjustment within households (down-shifting main-account ranks, up-shifting minors) followed by re-uniformisation (i.e. replacing adjusted ranks by their weighted mid-ranks). This preserves the BSR marginal exactly while controlling the amplification of within-household concentration induced by tail convexity.[1] In practice, we use diagnostics to decide whether the adjustment is needed, and (if so) calibrate a single tilt parameter to align the global main-account share with survey evidence, leaving all macro constraints unchanged.

A natural question is whether a more ambitious micro-synthesis calibrated to macro margins would dominate. Iterative proportional fitting and related techniques (Deming and Stephan, 1940), as well as population-synthesis frameworks from transport and spatial microsimulation (Beckman et al., 1996; Williamson et al., 1998), can be powerful when many accurate auxiliary margins are available. Our context differs: with a single dominant macro constraint (BSR range-sums) and limited auxiliary structure, parsimony buys robustness and transparency. Rank mapping exactly matches the BSR-implied account marginal, cluster resampling preserves the survey's intra-household dependence, and – crucially – identification from BSR (levels) remains cleanly separated from identification from HFCS/SHIW (household clustering and account multiplicity). Within the DWA production philosophy of targeted corrections followed by macro alignment, this division of labour is desirable; it minimises the number of moving parts and keeps the reconciliation traceable (Cantarella et al., 2024).[2]

[1] See also Embrechts et al. (2003) for background on tail convexity and the amplification of differences at extreme quantiles.

[2] See also European Central Bank (2024), *Distributional Wealth Accounts: Methodological Note*, available at https://data.ecb.europa.eu/.../DWA_Methodological_note.pdf

Because range-sum inputs are binned, bin geometry matters. Coarse upper bins magnify the leverage of the size-biased law, improving tail sensitivity but also concentrating identification on fewer aggregates. The mapping step therefore respects the observed bracket structure (including the open-ended top bin), and the household-cluster bootstrap stabilises household totals after re-aggregation. These choices mirror advice from the bias-correction and grouped-data literatures: constrain adjustments to the support identified by the external source, profile tail behaviour using bin-consistent diagnostics, and prefer procedures whose outputs can be explained in terms of the original bins (Maraun et al., 2010). In the Italian application, these considerations are especially salient for deposits, where account multiplicity is non-negligible and where the BSR open top bin concentrates a large fraction of outstanding amounts.

One additional consideration is that these results could potentially be used in assessing instrument-specific disparities in household holdings, as done in works like Chotikapanich et al. (2018) or Shorrocks and Wan (2008), for example. In fact, a credible, continuous distribution of a given instrument allows for detailed studies on concentration. However, in this thesis, given the different goal of improving the adjustment of survey-based wealth components, accounting for their multivariate distribution, single-level instrument concentration is out of scope and is left as a possible additional work outside of this thesis.

In summary, the reconstruction rests on established building blocks – size-biased inference from range-sum data, rank-based quantile mapping, and cluster-preserving bootstrap – augmented by Italian-specific survey side-information on account multiplicity and concentration. Together, these elements provide a transparent, statistically principled, and operationally tractable bridge from account-level BSR to household-level totals that is fit for purpose within the DWA calibration pipeline. The design is intentionally minimal: it targets the single macro constraint that is informative for deposits, preserves micro structures that matter for multivariate calibration, and remains auditable and reproducible across survey waves.

4.3 Proposed procedure (multi-step overview)

This section lays out, in words and in the order implemented, the workflow that turns account-based supervisory information into household-level deposit totals suitable for DWA-style calibration. The emphasis is on what is done and why each step is justified; formal details are deferred to the theoretical framework.

Step 0: Correct false zeros in HFCS participation. Before any reconstruction, the survey microdata are aligned to the BSR household scope by correcting obvious false zeros in deposits. Households reporting at least one deposit account but zero deposits are assigned a positive amount according to a deterministic, class-based rule. Classes are defined by deciles of net wealth excluding deposits and by the number of deposit accounts $\{1, 2, 3+\}$. For each class, we compute the survey-weighted median

Table 4.1. Impact of correcting false zeros in deposits.

Wave	Participation (per cent)		Coverage (per cent)	
	Old	New	Old	New
1	91.18	96.36	46.07	47.37
2	90.48	97.05	45.86	47.04
3	92.76	97.31	58.48	59.59
4	96.02	98.61	63.41	64.07
5	96.54	98.99	58.86	59.47

Notes: Participation denotes the survey-weighted share of households with strictly positive deposits. Coverage is the ratio (in per cent) of the HFCS-weighted deposit total to the BSR household deposit aggregate in the corresponding wave.

among positive depositors and assign that value to the false-zero cases in the same class. The resulting change in the percentage of deposit holders and in coverage (Table 4.1) is numerically small (at most equal to a 1.3% increase) but operationally relevant for the reliability of subsequent adjustments. Moreover, this imputation supports the augmentation of the GMM objective with a soft mean-of-holders moment, as defined in Section 4.5.

Step 1: Estimate the account-level target marginal from BSR. We begin by extracting the target account balance distribution from the BSR range-sum aggregates. The estimation adopts the specification validated in Chapter 3: a lognormal (LN) baseline and, as robustness, a lognormal-Pareto splice (PLN). This step translates the supervisory inputs – the amount shares by account-value bracket – into a precise, external marginal that subsequent micro reconstruction must honour. The rationale is straightforward: before touching the survey, we fix the "truth" at the account level as implied by BSR, so that all downstream adjustments are anchored to that object rather than to survey levels.

Step 2: Split household deposits into accounts using HFCS/SHIW side-information. Next, we create an account-level microstructure from the survey. Using the HFCS/SHIW variables on the number of deposit accounts and the main-account share, we partition each household's deposits into synthetic accounts. As my co-authors and I do in Neri et al. (2023), we apply operational rules for one/two/three or more accounts and impute missing counts by deposit size, so that the empirical main-account concentration is respected and each synthetic account retains the parent household identifier. This ensures that the ensuing account-level adjustments never break the household linkage that is needed later on.

Step 3: Enforce the BSR marginal via rank-preserving quantile mapping. We then align the split survey accounts to the BSR-consistent marginal by replacing each value with the corresponding quantile of the estimated account distribution, matching on (weighted) empirical ranks. Conceptually, we apply a

monotone probability-integral transform (Rosenblatt, 1952): ranks are kept, levels are mapped. This delivers exact agreement with the BSR-implied marginal while preserving the within-household ordering across multiple accounts and robustness to survey-level mismeasurement.

However, in heavy-tail settings (e.g., PLN with a large open top bin), pure rank mapping can unintentionally amplify the main-account share. To guard against this without altering the BSR marginal, we apply a household-local rank tilt before mapping and then re-uniformise ranks globally.

Specifically, let $X_{h\ell}$ denote the value of synthetic account ℓ in household h after splitting (Step 2), and let w_h be the household survey weight (assigned to each of its accounts). Let $U_{h\ell} \in (0,1)$ be the weighted empirical mid-rank of $X_{h\ell}$ among all synthetic accounts in the sample. These ranks are computed from the split-survey accounts and are not derived from BSR (Step 1 only delivers the target marginal F). To mitigate drift in the implied main-account share under heavy tails, we optionally tilt these ranks within household before mapping.

Define a within-household score $\eta_{h\ell}$ satisfying $\sum_{\ell=1}^{L_h} \eta_{h\ell} = 0$, where L_h is the number of deposit accounts in household h. We take $\eta_{h\ell} < 0$ for the main account and $\eta_{h\ell} > 0$ for minor accounts (e.g., $\eta_{h,\text{main}} = -(L_h - 1)/L_h$ and $\eta_{h,\text{minor}} = 1/L_h$), so that the tilt redistributes rank mass within household without changing its average level. Given a scalar tilt intensity β, define the tilted ranks

$$U^{\star}_{h\ell} = \text{logit}^{-1}\big(\,\text{logit}(U_{h\ell}) + \beta\,\eta_{h\ell}\big).$$

The tilted ranks $U^{\star}_{h\ell}$ are therefore computed, not observed. Because this within-household transformation generally destroys the global uniformity of ranks, we re-uniformise by replacing $\{U^{\star}_{h\ell}\}$ with their weighted empirical mid-ranks over all accounts, obtaining $\widetilde{U}_{h\ell}$. Mapping $\widetilde{X}_{h\ell} = F^{-1}(\widetilde{U}_{h\ell})$ then preserves the BSR marginal by construction.

The parameter β is common across households for parsimony and is chosen by a single calibration condition, for instance matching the implied global main-account deposit share after mapping to the SHIW/HFCS target (about 66%). Although β is global, its effect differs across households through $\eta_{h\ell}$ (which depends on L_h and on which account is the main one). This safeguard is activated only when diagnostics indicate top-bin concentration drift; otherwise we set $\beta = 0$ and revert to pure rank-preserving mapping.

Step 4: Resample at the household (cluster) level. To propagate realistic multi-account dependence, we bootstrap households, bringing along all their accounts in each draw to form replicate account micro-populations. Cluster-preserving resampling is standard when intra-cluster correlation matters (Rust and Rao, 1996; Davison and Hinkley, 1997). Here, it stabilises household-level summaries and maintains the empirical link between account multiplicity and household characteristics. In this step, we also assess that the account-level distribution of the bootstrap sample still matches the BSR marginal account distribution.

Step 5: Re-aggregate back to the household level. Within each bootstrap replicate, we sum accounts back to the (replicate, household) level. Because the account values already match both the BSR marginal and the SHIW estimated main account share, and households are resampled as intact clusters, the resulting deposit totals are coherent with supervisory constraints and retain survey-based household structure and information. Moreover, the bootstrap microdata is used to estimate the underlying household-level deposits distribution, deriving, as before, a LN or PLN specification.

Step 6: Construct household-level, BSR-consistent bracket shares. Finally, in order to compare newly derived household-level BSR data with the original ones, we tabulate the bootstrap household totals into the BSR brackets and compute amount shares; if needed, we scale by the BSR aggregate to enforce exact totals by bracket. The output is a household-level distribution that is, by construction, aligned with the administrative account-level aggregates and ready to be fed into the multivariate calibration developed in Section 6.1.3.

In short, the sequence first fixes the account-level target (Step 1), then builds a disciplined household-to-account partition from HFCS/SHIW (Step 2), and imposes the external marginal with minimal assumptions, optionally adding a rank-tilt safeguard that preserves the BSR marginal (Step 3), all while preserving intra-household dependence through clustered resampling (Step 4) and returning to the household as the unit of analysis (Steps 5). The final object is a household-level BSR estimated distribution that can be later used in the penalty definition of the calibration in Section 6.1.2–6.1.3. In order to compare the impact of this, we also reconstruct a BSR-like household-level data, tabulating the bootstrap household totals into the BSR brackets and computing amount shares, scaling to the totals observed in the BSR (Step 6).

It is necessary to note limitations due to data availability. The main-account deposit share is observed only for the sample in Wave 4 of the SHIW/HFCS. In the absence of analogous information for other waves, we impute the Wave 4 estimates to earlier and later years. This approach cannot capture potential changes over time in households' preferences regarding the intra-household split of deposits and may therefore introduce bias. Nevertheless, such behavioural changes are unlikely to be abrupt, and all survey waves report the number of deposit accounts per household, which allows us to accommodate time variation to a reasonable extent. Even so, the development of wave-specific evidence on main-account shares should be regarded as a priority for future improvement as and when data permit.

4.4 Theoretical framework

This subsection formalises the account-to-household reconstruction used to obtain household-level deposit totals that are coherent with the account-level BSR constraints. The procedure combines: (i) a rank-preserving, measure-transport mapping from survey-derived account values to the BSR-consistent marginal distribution; and

(*ii*) a cluster (household) bootstrap that preserves multi-account dependence before re-aggregation to the household level. In addition – prior to these steps – the survey microdata are aligned to the BSR scope by correcting obvious false zeros in deposits: households reporting at least one deposit account but zero deposits receive a small positive imputation based on survey-weighted class medians (classes by deciles of net wealth excluding deposits and by the number of deposit accounts $\{1, 2, 3+\}$).

Let households be indexed by $h = 1, \dots, H$, with survey weight $w_h > 0$ and (post-imputation) number of deposit accounts $m_h \geq 1$. Denote the account vector for household h by

$$\boldsymbol{c}_h \;=\; (c_{h1}, \dots, c_{hm_h}), \qquad \sum_{\ell=1}^{m_h} c_{h\ell} \;=\; C_h^{\mathrm{HH}},$$

where C_h^{HH} is the household's total deposits used for splitting. The splitting rules that determine the within-household allocation (main-account share and residual equal split across minors) are as in Neri et al. (2023). Define the account-level dataset

$$\mathcal{C} \;=\; \{(c_{h\ell},\, w_h) : h = 1, \dots, H;\; \ell = 1, \dots, m_h\},$$

i.e. each account inherits the household weight so that $\mathcal{C}$ represents the account population.

On the BSR side, let $0 = a_0 < a_1 < \cdots < a_{J-1} < a_J = \infty$ be the account-value bracket edges used to tabulate amount totals. Normalised by the grand total, the BSR provide amount shares $\widehat{s}_j$ which, in expectation, equal bin probabilities under the size-biased law, as defined in Equation (3.22).

Recall that $g_\theta(x) \propto x f(x; \boldsymbol{\theta})$ is the size-biased density induced by the account-level balance distribution $f(\cdot; \boldsymbol{\theta})$ (Patil and Rao, 1978; Vardi, 1982). Chapter 3 provides a consistent estimator $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ (LN/PLN family), hence a fully specified target CDF

$$F(x) \equiv F(x; \widehat{\boldsymbol{\theta}}) \;=\; \int_0^x f(t; \widehat{\boldsymbol{\theta}})\, dt, \qquad F^{-1}(u) \equiv \inf\{x :\, F(x) \geq u\}.$$

Construct the weighted empirical account-level CDF from $\mathcal{C}$,

$$\widehat{F}_{\mathcal{C}}(x) \;=\; \frac{\sum_{h=1}^{H} \sum_{\ell=1}^{m_h} w_h\, \mathbb{1}\{c_{h\ell} \leq x\}}{\sum_{h=1}^{H} \sum_{\ell=1}^{m_h} w_h}, \tag{4.1}$$

and define the rank (generalised) transform $U_{h\ell} = \widehat{F}_{\mathcal{C}}(c_{h\ell}) \in [0, 1]$. The baseline quantile map is

$$\widetilde{c}_{h\ell}(0) \;=\; F^{-1}(U_{h\ell})\,, \tag{4.2}$$

which is monotone and rank-preserving (Rosenblatt, 1952). Hence, under standard Glivenko–Cantelli conditions for weighted empirical CDFs, the re-mapped account

sample $\{\widetilde{c}_{h\ell}(0)\}$ has a weighted empirical CDF that converges uniformly to F; in finite samples it exactly matches F on the grid of empirical ranks (see also the monotone rearrangement rationale in Chernozhukov et al., 2013).

As noted, in heavy-tail settings (e.g., PLN), pure rank mapping can mechanically amplify the main-account share. To control this without altering the BSR marginal, we optionally apply a household-local rank tilt followed by global re-uniformisation. Let $\eta_{h\ell}$ satisfy $\sum_{\ell=1}^{m_h} \eta_{h\ell} = 0$ (e.g., $\eta_{h1} = -1$ for the main account, $\eta_{h\ell} = 1/(m_h - 1)$ for $\ell \geq 2$). For $\beta \geq 0$,

$$U^{\star}_{h\ell}(\beta) \;=\; \mathrm{logit}^{-1}\big(\mathrm{logit}(U_{h\ell}) + \beta\, \eta_{h\ell}\big),$$

and let $\widehat{F}_{U^{\star}(\beta)}$ be the weighted empirical CDF of $\{U^{\star}_{h\ell}(\beta)\}$ over all accounts. The re-uniformised ranks are

$$\widetilde{U}_{h\ell}(\beta) \;=\; \widehat{F}_{U^{\star}(\beta)}\big(U^{\star}_{h\ell}(\beta)\big).$$

Define the corresponding mapped account values as

$$\widetilde{c}_{h\ell}(\beta) \;=\; F^{-1}\big(\widetilde{U}_{h\ell}(\beta)\big). \tag{4.3}$$

We set $\widehat{\beta} = 0$ unless diagnostics indicate top-bin concentration drift, in which case $\widehat{\beta}$ solves the one-moment condition that matches the aggregate main-account share after mapping to its survey target (about 66%), i.e. $\Psi(\widehat{\beta}) = 0$, where $\Psi(\beta)$ is the difference between realised and target main-account shares (computed over multi-account households only). In what follows, we write $\widetilde{c}_{h\ell} \equiv \widetilde{c}_{h\ell}(\widehat{\beta})$ for the final mapped account values.

To propagate realistic multi-account dependence into household totals, resampling proceeds at the household (cluster) level (Rao and Wu, 1988; Rao et al., 1992; Rust and Rao, 1996; Davison and Hinkley, 1997). For bootstrap replication $b = 1, \ldots, B$:

1. Draw a multiset of household indices $\mathcal{H}^{(b)} = \{h_1^{(b)}, \ldots, h_{n_b}^{(b)}\}$ by sampling with replacement from $\{1, \ldots, H\}$ with probabilities proportional to w_h.
2. For each selected $h \in \mathcal{H}^{(b)}$, include all mapped accounts $\{\widetilde{c}_{h\ell} : \ell = 1, \ldots, m_h\}$. Define the bootstrap household total

$$Y_h^{(b)} \;=\; \sum_{\ell=1}^{m_h} \widetilde{c}_{h\ell}, \qquad \widetilde{w}_h^{(b)} \;=\; \text{replicate weight (e.g., count of times } h \text{ is selected).} \tag{4.4}$$

Because clustering is preserved, $(Y_h^{(b)})$ retains the empirical joint structure of multi-account holdings observed in HFCS/SHIW, while their marginal is anchored to F (or to the tilted-and-re-uniformised $\widetilde{U}_{h\ell}$) via the mapping step. This object can then be used to estimate the household-level distribution, consistent with both BSR account-level information and HFCS household-level account dependency information.

In order to compare how BSR household-level data differ from original account-level data, we then aggregate results following BSR reported brackets. Let $\mathcal{I}_j[a_{j-1}, a_j)$ be a set of household-total deposit brackets (by default, the same edges as BSR account brackets). Define the household-level amount share in replication b by

$$\widehat{s}_j^{(b)} \;=\; \frac{\sum_{h\in\mathcal{H}^{(b)}} \widetilde{w}_h^{(b)}\, Y_h^{(b)}\, \mathbb{1}\{Y_h^{(b)} \in \mathcal{I}_j\}}{\sum_{h\in\mathcal{H}^{(b)}} \widetilde{w}_h^{(b)}\, Y_h^{(b)}}, \qquad j = 1, \ldots, J. \tag{4.5}$$

By construction, $\sum_j \widehat{s}_j^{(b)} = 1$. A point estimate may be taken as $\widehat{s}_j = \frac{1}{B}\sum_{b=1}^{B} \widehat{s}_j^{(b)}$. Uncertainty can be summarised by the empirical dispersion of $\{\widehat{s}_j^{(b)}\}_{b=1}^{B}$, for example via a bootstrap standard error $\widehat{\mathrm{se}}(\widehat{s}_j) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{s}_j^{(b)} - \widehat{s}_j\right)^2}$ and/or percentile intervals from the bootstrap quantiles.

Let T denote the BSR grand total for household-sector deposits (conceptually aligned after coverage/timing adjustments). Household-level bracket totals that are distributionally consistent with the reconstructed microdata and level-consistent with the macro aggregate are then given by

$$T_j^{\mathrm{HH}} \;=\; \widehat{s}_j\, T, \qquad j = 1, \ldots, J. \tag{4.6}$$

Equation (4.6) re-allocates the macro total across household brackets using micro-founded shares $\widehat{s}_j$ derived from the mapped-and-bootstrapped household totals.

Properties and remarks.
In synthesis, this approach yields some important properties:

(i) **Measure-preservation at the account level.** The mapping (4.2) (or its tilted-and-re-uniformised variant (4.3)) is a monotone rearrangement that aligns the account-level marginal with F and hence with BSR range-sum constraints, while leaving ranks – and therefore dependence patterns tied to ranks – unchanged in the absence of the optional tilt.

(ii) **Dependence preservation.** Cluster resampling (4.4) preserves within-household multi-account dependence, a standard requirement for resampling with complex/clustered data (Rao and Wu, 1988; Rao et al., 1992; Rust and Rao, 1996; Davison and Hinkley, 1997).

(iii) **Macro coherence.** The re-scaling (4.6) ensures that, even if finite-sample mapping/bootstrapping induces small level deviations, reported household totals by bracket sum to the macro T while reflecting household structure.

(iv) **Identification stance.** The approach is semi-parametric: identification of the marginal at the account level is parametric (LN/PLN via Chapter 3); identification of household aggregation relies on observed HFCS/SHIW clustering (non-parametric in the joint distribution of account multiplicities and intra-household composition).

(v) **Activation rule for the tilt.** The rank-tilt is off by default and activated only when diagnostics flag tail-driven main-account inflation (e.g., top-bin overconcentration or a deviation of the global main-account share from its survey target beyond a tolerance band). The one-parameter choice of β restores that moment while exactly preserving the BSR marginal through re-uniformisation.

Finally, the methodology can be summarised as such:

1. Split household deposits into accounts $\boldsymbol{c}_h$ using HFCS/SHIW rules (after imputing the number of accounts for households with positive holding without this information); form $\mathcal{C}$.
2. Estimate $F(\cdot;\widehat{\boldsymbol{\theta}})$ from BSR (Chapter 3); compute ranks $U_{h\ell}$ and (optionally) apply the tilt $U^{\star}_{h\ell}$ and re-uniformisation $\widetilde{U}_{h\ell}$.
3. Map accounts: $\widetilde{c}_{h\ell} = F^{-1}(U_{h\ell})$ or $\widetilde{c}_{h\ell} = F^{-1}(\widetilde{U}_{h\ell})$ (tilted case).
4. Cluster bootstrap households; form $Y_h^{(b)} = \sum_{\ell} \widetilde{c}_{h\ell}$.
5. Compute $\widehat{s}_j^{(b)}$ via (4.5); average to $\widehat{s}_j$; set $T_j^{\mathrm{HH}} = \widehat{s}_j T$ for comparability.

4.5 Augmenting the GMM with a soft mean-of-holders moment

After reconciling BSR deposits to the household level, the identification of the distribution can be strengthened – especially in the upper tail – by augmenting the objective with a soft moment on the household mean among holders. This construction parallels the real-data augmentation introduced for Chapter 3 (see Section 3.5.2), where a client-based anchor from BSR is used; here, by contrast, the anchor is computed at the household level from HFCS participation. The additional information is the average deposits per holding household, constructed as

$$\widehat{\mu}_t^{H} \;=\; \frac{T_t^{\mathrm{BSR}}}{\widehat{N}_t^{H,+}}, \tag{4.7}$$

where T_t^{BSR} denotes year-t aggregate deposits from the sectoral accounts/BSR, and $\widehat{N}_t^{H,+}$ is the estimated number of households with strictly positive deposits. Since, for the purposes of this augmentation, only the count of deposit-holding households is required from HFCS, a basic imputation is applied to mitigate false-zero claims: in the denominator of (4.7), households are counted if they either report a strictly positive deposit amount or report at least one deposit account. This second criterion captures cases where a household reports owning a deposit account but records a zero amount; the resulting increase in $\widehat{N}_t^{H,+}$ is numerically modest – see Table 4.1 – yet improves the credibility of the denominator (and, by extension, the mean-of-holders anchor). For the client-based anchor and related coverage/duplication caveats, see also (3.29) and the discussion in Section 3.5.2.

Moreover, deposit participation in HFCS is already high (90% or above across waves). Even if all remaining zeroes were in fact false, the effect on the implied household average would be limited. Consequently, the mean anchor is imposed softly.

The anchor $\widehat{\mu}_t^H$ is obtained by dividing year-t deposit aggregates by the estimated number of holding households. Table 4.2 reports the resulting averages, which are plausible in level and trend, and contrasts them with two simple alternatives: (*i*) the average using non-imputed participation ("$\widehat{\mu}_t^H$ raw"), and (*ii*) the average assuming universal participation in HFCS ("$\widehat{\mu}_t^H$ all").

Table 4.2. Average deposits per holding household used as soft anchor.

Year	$\widehat{N}_t^{H,+}$	T_t^{BSR} (million €)	$\widehat{\mu}_t^H$ (€)	$\widehat{\mu}_t^H$ raw (€)	$\widehat{\mu}_t^H$ all (€)
2010	22,955,466	985,657	42,938	45,378	41,375
2014	23,989,367	1,092,104	45,525	48,829	44,180
2016	24,897,920	1,155,933	46,427	48,700	45,176
2020	24,896,012	1,327,853	53,336	54,776	52,595
2022	24,804,508	1,389,714	56,027	57,449	55,463

Let $g_t(\boldsymbol{\theta})$ denote the vector of bracket-share moment discrepancies used in Chapter 3, and let I_J be the $J \times J$ identity matrix (where J is the number of brackets used in the fit). Let $\mu_t^H(\boldsymbol{\theta})$ be the model-implied household mean among deposit-holding households under parameters $\boldsymbol{\theta}$.

In analogy to the client-based augmentation introduced in Chapter 3 (Equation (3.30)), we augment the baseline bracket-share criterion $g_t(\boldsymbol{\theta})^\top I_J\, g_t(\boldsymbol{\theta})$ with a soft penalty that anchors the mean among holders to the household-level target $\widehat{\mu}_t^H$. Specifically, the criterion for year t is

$$\mathcal{L}_t(\boldsymbol{\theta}) \;=\; g_t(\boldsymbol{\theta})^\top I_J\, g_t(\boldsymbol{\theta}) \;+\; \lambda_\mu \left(\frac{\mu_t^H(\boldsymbol{\theta}) \;-\; \widehat{\mu}_t^H}{\widehat{\mu}_t^H} \right)^2, \tag{4.8}$$

with a small tuning constant $\lambda_\mu > 0$ to reflect the softer reliability of $\widehat{\mu}_t^H$ relative to the bracket constraints. The relative-error scaling renders the penalty dimensionless and comparable across years. In practice, this additional moment stabilises the level of the household distribution (and, in particular, the multiplicity component in the account-to-household mapping), while the bracket moments continue to discipline the marginal shape and alignment to BSR totals. For comparison with the client-based specification and its objective, see (3.29)–(3.30).

Two remarks are in order. First, the augmented loss in (4.8) remains a standard GMM criterion: it simply appends a moment on the household mean among holders. Second, because total deposits T_t^{BSR} are already enforced through the bracket system, the mean moment is not redundant provided no hard penalty is imposed on the count of holding households; in this set-up, it acts as a stabiliser for the level, while

the bracket moments discipline the shape. For securities-accounts disentanglement (Chapter 5), a similar penalty is introduced, with some additional caveat: although the scope reconciliation is comparable, zero-reporting at the household level is potentially more salient for those instruments, which would place undue weight on a noisy anchor given the sparse nature of such holdings. Thus, we consider the estimated mean after reconciliation of zero claims as a soft upper bound of the true mean among holders. For the broader rationale and robustness protocol shared with the client-based augmentation, see Section 3.5.2.

4.6 Results and discussion

4.6.1 Simulated data

This subsection evaluates, by simulation, the account-to-household aggregation pipeline developed in this chapter. The design mirrors the empirical setting: a latent household-level distribution of deposits generates account multiplicity and sizes, while survey observations are a biased proxy that must be reconciled to account-level BSR targets and then re-aggregated to households. We consider twenty scenarios aligned with Chapter 3, Table B.5: α, τ, and the bracket scheme (pre-2022 vs post-2022) vary on the same grids as before; μ and σ are fixed at 9.2 and 1.5, tailored to this chapter to reflect the different role of multiplicity in deposits.[3] For each scenario we simulate $B = 100$ replications. In each replication, a large pseudo-population of 2,000,000 households is drawn from the household-level DGP to induce negligible Monte Carlo error on the "truth" used for validation.

Household total deposits are split into accounts via a size-dependent multiplicity mechanism consistent with the splitting step in Section 4.4. To avoid confusion with the account-level rank $U_{h\ell}$ used later in the quantile mapping, let $U_h \sim \mathcal{U}[0,1]$ denote the household-level deposits latent rank in the simulated household distribution. Conditional on U_h, we draw the number of accounts $m_h \in \{1,2,3,4\}$ and then allocate the household total across $\{c_{h\ell}\}_{\ell=1}^{m_h}$ using the same main-account and residual-split rules as in Section 4.4. The resulting simulated account sample $\{c_{h\ell}\}$ is then treated as $\mathcal{C}$ in Equation (4.1), from which we compute account ranks $U_{h\ell}$ and apply the mapping/bootstrapping pipeline exactly as in Section 4.4.

The number of accounts $m_h \in \{1,2,3,4\}$ is drawn with probabilities increasing

[3]Scenario grids reproduce the tails-and-breaks structure of Chapter 3 to keep comparability of identification patterns across chapters.

in U_h according to the operational rule

$$\begin{aligned}
\ell(x) &= \frac{1}{1+e^{-x}}, \\
\tilde{p}_2(U_h) &= \ell(-1.8 + 3.0\,U_h), \\
\tilde{p}_3(U_h) &= 0.6\,\ell(-3.0 + 5.0\,U_h), \\
\tilde{p}_4(U_h) &= 0.4\,\ell(-5.0 + 7.0\,U_h), \\
p_2(U_h) &= \min\{\tilde{p}_2(U_h),\, 0.8\}, \\
p_3(U_h) &= \min\{\tilde{p}_3(U_h),\, 0.5\}, \\
p_4(U_h) &= \min\{\tilde{p}_4(U_h),\, 0.4\}, \\
p_1(U_h) &= \max\left\{0,\; 1 - p_2(U_h) - p_3(U_h) - p_4(U_h)\right\}, \\
\pi_m(U_h) &= \frac{p_m(U_h)}{\sum_{j=1}^{4} p_j(U_h)}, \qquad m = 1,2,3,4, \qquad \Pr(m_h = m \mid U_h) = \pi_m(U_h).
\end{aligned}$$

to mimic the relationship between account multiplicity and holding size observed in HFCS, while the main-account share s_{main} for multi-account holders follows a truncated Beta$(12, 6)$ law restricted to $[0.45, 0.85]$, yielding a mean near 0.66 as in HFCS/SHIW while allowing realistic variability. The remaining mass $1 - s_{\text{main}}$ is allocated to the secondary accounts using the operational rule of Section 4.3 so as to preserve within-household dependence. From these account-level values, we form BSR-like account aggregates by bracket, thereby obtaining the pseudo-BSR account-level target used downstream in each replication.

A pseudo-HFCS sample of size 8,000 households is then drawn from the same latent household distribution. To mimic survey features, two distortions are introduced: the main-account share is fixed at 0.66 (removing its cross-household variability, in order to replicate the splitting rule applied in the real-data application), and a multiplicative attenuation is applied to each observed deposit with i.i.d. Beta$(6, 4)$ draws, producing the biased pseudo-HFCS amounts.[4]

The reconciliation pipeline of Section 4.3 is then applied: survey household deposits are split into accounts (with fixed 0.66 main share in the survey arm), survey account ranks are mapped to the quantiles of the pseudo-BSR account-level fit (LN or PLN, estimated with the bracket-share GMM and with DGP detection as developed in Chapter 3), and a cluster bootstrap resamples households (with all their accounts) to preserve within-household correlation. Finally, accounts are re-aggregated to household totals, and household wealth shares by BSR brackets are computed.

Performance is assessed on three layers. First, we evaluate goodness of fit on the estimation target – household bracket wealth shares – by comparing, for each replication, the true (unobserved) household shares $s^{\text{hh,true}}$, the biased pseudo-HFCS shares

[4] False-zero correction is applied as in Section 4.1: households declaring at least one deposit account but a zero amount are treated as participants with a small positive assignment confined to the bottom brackets. This affects participation counts but has negligible impact on the distributional shape above the lower bins.

$s^{\text{hh,biased}}$, and the reconstructed shares $s^{\text{hh,recon}}$ implied by the pipeline. Figures 4.1 and 4.2 show that reconstructed shares track the truth closely across scenarios, with discrepancies largely confined to the lowest brackets. To highlight local deviations, each panel encodes the absolute difference $|s^{\text{hh,recon}} - s^{\text{hh,true}}|$ in the shading of the true points (darker indicates larger deviations). A compact summary is provided in Table B.6, which reports, for each scenario, the median mean absolute error (MAE) across brackets for the reconstructed and biased shares and the associated inter-decile range over replications.

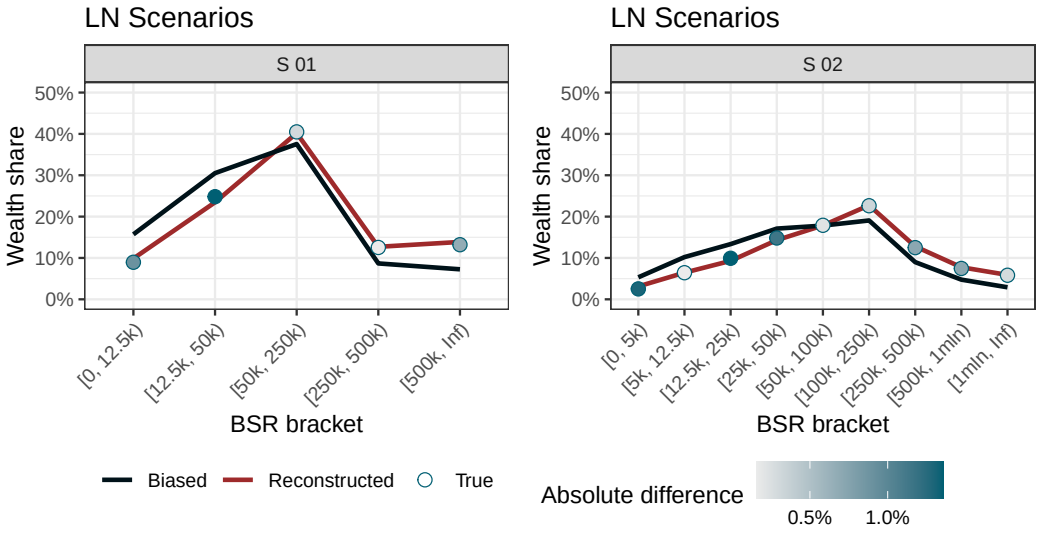

Figure 4.1. Simulated goodness of fit on household bracket wealth shares: LN scenarios (S 01–S 02).
Notes: Points denote $s^{\text{hh,true}}$; red lines denote $s^{\text{hh,recon}}$; dark grey lines denote $s^{\text{hh,biased}}$. Point shading encodes $|s^{\text{hh,recon}} - s^{\text{hh,true}}|$ by bracket (darker indicates larger local deviations). Scenarios vary the bracket scheme (pre- vs post-2022) as in Table B.5.

Figures 4.1–4.2 and indicate that reconstructed shares closely recover the true household-level composition. The median MAE typically shrinks from roughly 3–5% to below 1% in most scenarios, including heavy-tailed PLN cases (e.g., S 06 and S 12). Visual inspection suggests that the largest improvements arise in the tails, while central brackets are affected mainly through small mass reallocation; the biased sample's MAE is partly driven by under/over-assignment across the extreme bins. Overall, under a design that mirrors the empirical application, the aggregation–reconciliation procedure reproduces the latent bracket shares with high accuracy. Table B.6 in Appendix B provides the full table on these values.

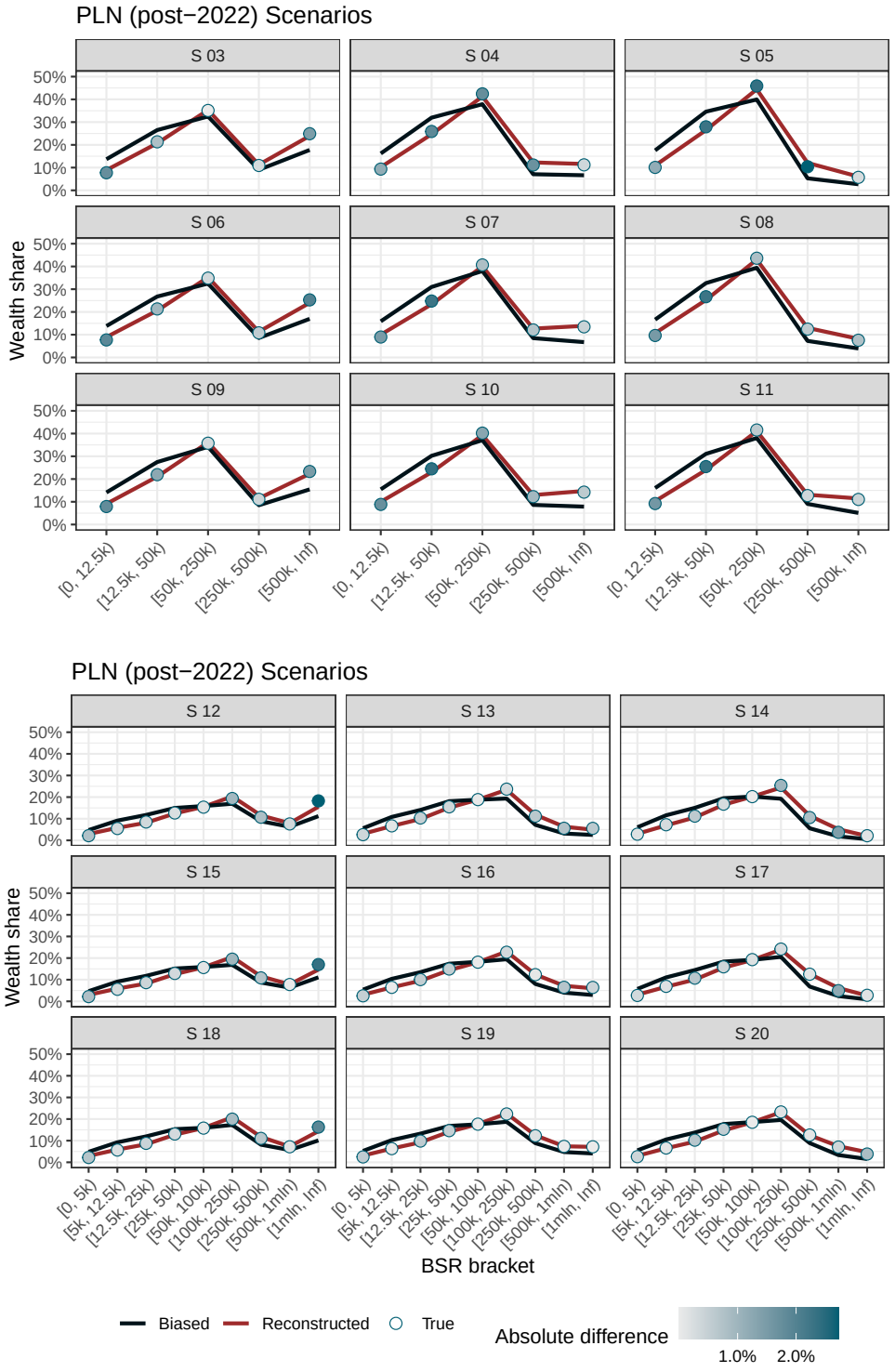

Figure 4.2. Simulated goodness of fit on household bracket wealth shares: PLN scenarios. *Notes:* Points denote $s^{\mathrm{hh,true}}$; red lines denote $s^{\mathrm{hh,recon}}$; dark grey lines denote $s^{\mathrm{hh,biased}}$. Point shading encodes $|s^{\mathrm{hh,recon}} - s^{\mathrm{hh,true}}|$ by bracket (darker indicates larger local deviations). Bracket structures differ in 2022, as reflected in the panels. Scenario codes correspond to Table B.5.

Second, we summarise distributional discrepancies using per-replication distances: total variation (TV) and the root-mean-squared error (RMSE) between reconstructed and true household share vectors, alongside the same metrics "before" reconciliation (biased vs truth). Figure 4.3 reports TV by scenario before and after reconciliation; in Appendix A, Figure A.3 shows the corresponding RMSE comparison. Across scenarios, both distances fall markedly after reconciliation, indicating that the pipeline substantially reduces distributional error relative to the biased pseudo-HFCS baseline.

Results indicate sizeable TV reductions across scenarios: the average reduction exceeds 70% in all scenarios and 80% in fourteen of them. RMSE exhibits similar declines: the average reduction exceeds 67% in every scenario and 78% in fourteen of them.

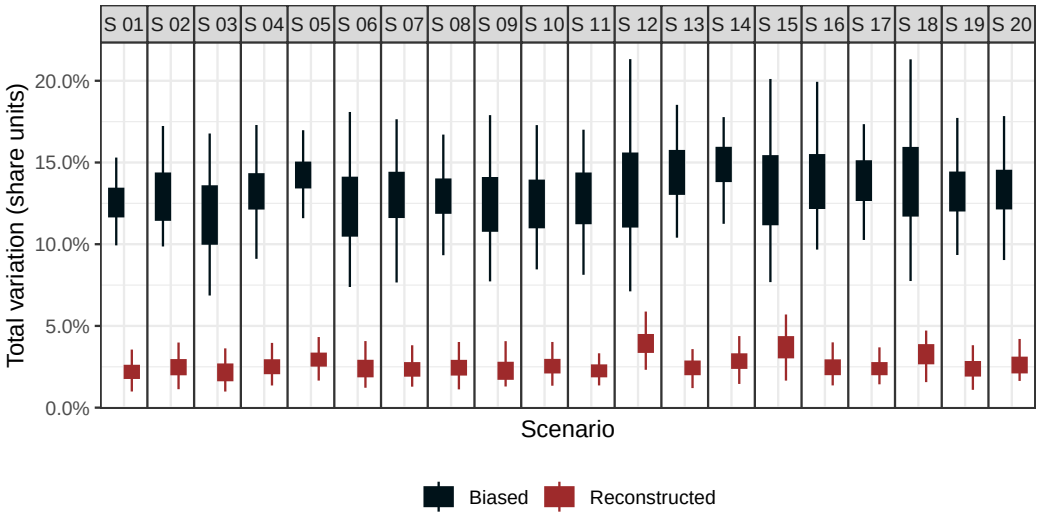

Figure 4.3. Total variation distance by scenario: Before (Biased) vs After (Reconstructed). *Notes:* Boxplots over $B = 100$ replications. "Before (Biased)" compares biased pseudo-HFCS household shares to truth; "After (Reconstructed)" compares reconstructed shares to truth.

Third, because identification is most fragile in the upper tail, we report tail-focused accuracy by scenario using the top-bin error and the cumulative error over the top two bins. This also helps disentangle the central part being seemingly unbiased in the starting shares, due to the combined effect of the top tail moving down to lower breaks and the central breaks moving down to lower breaks. As noted, Figures 4.1–4.2 already hint at this phenomenon. We now study it in more detail. Figure 4.4 maps the median top-bin error over the (α, τ) grid for pre-2022 and post-2022 brackets. Consistent with Chapter 3, errors are higher when τ encroaches on the open-top range, even tough in this case, they are higher when α implies heavier tail mass. In particular, the post-2022 scheme reports the highest errors, as the top tail bin for such breaks start at a very high value.

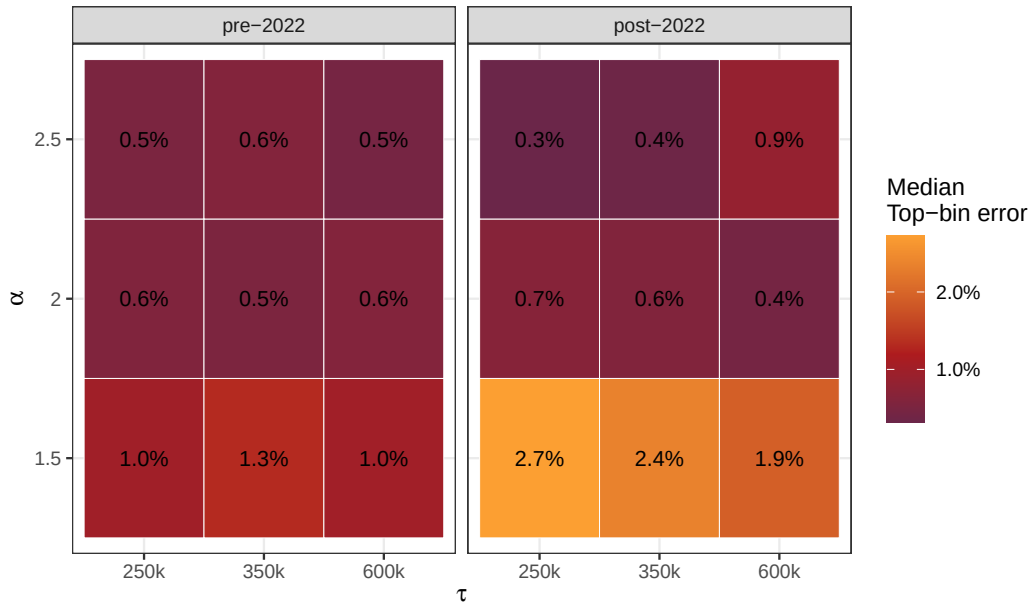

Figure 4.4. Tail-focused accuracy: median top-bin error by (α, τ) and bracket scheme.

To verify that the pipeline preserves recoverability of structural parameters, we re-estimate LN/PLN on the reconstructed household shares in each replication and compare the resulting parameters with the scenario truth. Figure 4.5 displays bias and Monte Carlo spread for $(\mu, \sigma, \alpha, \tau)$ across scenarios (points: mean error; whiskers: ± 2 standard deviation).

Mean errors for the body parameters are modest but not uniformly negligible, and they are larger than those achieved in Chapter 3. With whiskers set to ± 2 Monte Carlo standard deviations, the dispersion of estimation errors is substantial and the error distribution often includes zero, even when the mean error shows a systematic direction. An increased number of breaks helps to reduce the bias, as all post-2022 scenarios have lower bias on μ and σ with respect to their corresponding pre-2022 scenarios. Dispersion for α and τ increases as tail weight declines or τ approaches the open-top range, in line with the identification geometry documented

in Chapter 3. Nonetheless, the systematic under-estimation of μ and over-estimation of σ further support the anchor method developed in Section 4.5, which stabilises estimation around a credible target for the mean.

Table 4.3 reports the relative RMSE (rRMSE, in per cent), defined as

$$\mathrm{rRMSE}(\boldsymbol{\theta}) \;=\; 100 \times \frac{\mathrm{RMSE}(\boldsymbol{\theta})}{|\boldsymbol{\theta}_0|},$$

where $\boldsymbol{\theta}_0$ denotes the scenario truth for the relevant parameter. For the body parameters, rRMSE is generally lower under the post-2022 bracket design, consistent with the finer upper-bin geometry improving identification. By contrast, τ is the least precisely recovered parameter, reflecting both weak identification when the threshold lies near the open-top range and the fact that τ is selected by profiling over a discrete grid. In this simulation study – since threshold recovery is not the main focus of the chapter – we intentionally use a coarser τ-grid than in the empirical application to reduce computation across 20 scenarios and $B = 100$ replications; this discretisation mechanically inflates RMSE for τ.

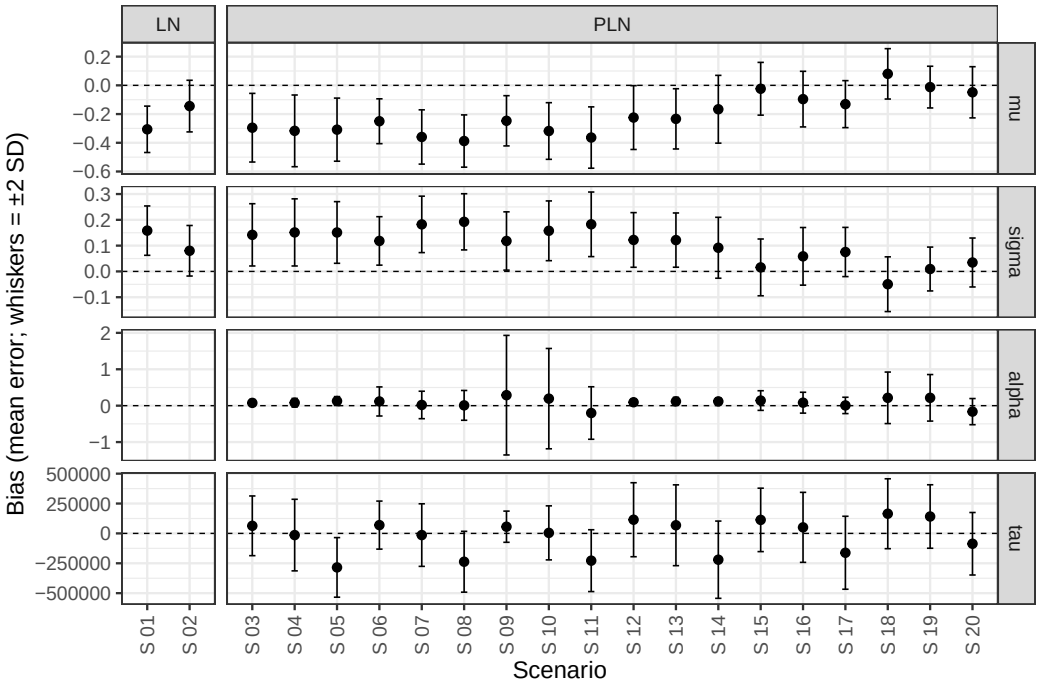

Figure 4.5. Parameter recovery from reconstructed household shares: bias and spread by scenario.

Notes: Points show mean estimation error over replications; whiskers are two standard deviations.

Finally, to motivate the soft mean-of-holders augmentation, we repeat the reconstruction but fit LN/PLN *without* the mean anchor and then compare, for each replication, the model-implied mean $\mu^H(\widehat{\boldsymbol{\theta}})$ with the scenario's true mean μ_0^H.

Table 4.3. Relative RMSE (per cent) of parameter estimates from reconstructed household shares by scenario.

Scenario	rRMSE(μ)	rRMSE(σ)	rRMSE(α)	rRMSE(τ)
S 01	3.4	11.0	-	-
S 02	1.8	6.3	-	-
S 03	3.5	10.3	5.9	55.8
S 04	3.7	11.0	6.7	42.8
S 05	3.6	10.8	9.4	51.7
S 06	2.8	8.5	11.5	48.7
S 07	4.0	12.7	9.4	37.4
S 08	4.3	13.3	10.2	44.8
S 09	2.8	8.7	34.7	34.3
S 10	3.6	11.2	28.5	32.2
S 11	4.1	12.9	16.5	43.6
S 12	2.7	8.9	6.9	76.8
S 13	2.8	8.8	8.7	51.9
S 14	2.2	7.3	8.5	45.4
S 15	1.0	3.8	9.7	69.5
S 16	1.5	5.4	8.2	44.0
S 17	1.7	5.9	5.6	37.0
S 18	1.3	4.8	16.5	87.9
S 19	0.8	2.9	15.4	55.3
S 20	1.1	3.9	9.7	26.0

Figure 4.6 plots bracket-share MAE against the relative mean error

$$\text{relative mean error} = \frac{|\mu^H(\widehat{\boldsymbol{\theta}}) - \mu_0^H|}{\mu_0^H}.$$

The cloud shows that even when the bracket fit is tight (most MAE, roughly 77%, below one percentage point), the level can drift materially – often several percentage points and occasionally well above 10 per cent (32% of runs have MAE below 1% and a relative mean error above 10%). This decoupling between shape and level is precisely what the soft anchor is designed to correct. In the empirical implementation that follows we include the mean moment; as the simulations indicate, it stabilises the level at negligible cost to bracket fit.

Taken together, the simulations show that the account-to-household aggregation and reconciliation pipeline reproduces the implied household-level bracket shares with high accuracy. MAE is below 1% on average in 18 out of the 20 scenarios, and the overall median value across all scenarios is 0.77%. Second, structural parameters

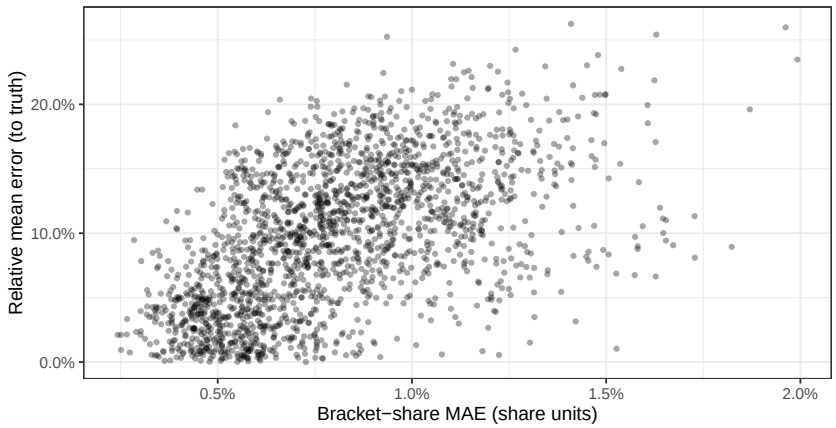

Figure 4.6. Bracket-share MAE vs relative mean error from non-augmented fits (per replication).
Notes: Each point is a replication. MAE is computed on household bracket shares; the relative mean error compares the model-implied mean to the scenario's true mean. The scatter documents that accurate bracket fit does not guarantee coherence in the level, motivating the mean-anchor augmentation.

display slightly higher biases than in Chapter 3 – as expected given the random assignment of main-account shares – but dispersion remains limited and recovery is reliable across scenarios (with the post-2022 break design mitigating extremes). Third, the model-implied mean is the most sensitive quantity: even when bracket MAE is below one percentage point, relative mean errors can be several percentage points and occasionally approach 10 per cent. This decoupling of level from shape motivates the soft mean-of-holders augmentation, which improves level coherence at negligible cost to bracket fit.

On this basis, we now apply the reconstruction procedure to the real data, replicating the same pipeline on BSR/HFCS and reporting the corresponding fit diagnostics, parameter estimates, and results under the augmented objective.

4.6.2 Real-data application

Following the simulation evidence, we now apply the same pipeline to observed BSR data, using the mean anchor as described in Section 4.5. From a computational perspective, the point-estimate pipeline (account-to-household reconstruction followed by LN/PLN selection and parameter estimation) takes roughly 20 minutes per HFCS wave under the current implementation; any additional uncertainty assessment based on repeating the cluster bootstrap scales approximately linearly in the number of bootstrap replications. The real-data procedure proceeds in two steps. First, we aggregate BSR deposits from the account unit to the household unit following Section 4.3, obtaining a BSR-coherent household-level microdataset. Second, we estimate the household distribution on this microdataset. For reporting and comparability across years, we bin household totals on the post-2022 break vector for all years to sharpen tail resolution.

After reconciliation, Table 4.4 reports year-specific PLN parameter estimates on the reconstructed BSR-coherent household microdata, contrasting the augmented objective (Section 4.5) with the baseline bracket-share GMM without augmentation. We apply the same model-selection and DGP testing protocol as in Chapter 3. All years prefer a PLN specification (we omit pseudo-p-values for brevity, as they are effectively zero in all cases). Because the BSR aggregates are treated as fixed administrative totals, conventional sampling uncertainty is not defined for the year-specific bracket-share inputs. However, the account-to-household reconstruction relies on survey microstructure (account multiplicity and within-household splits) and on clustered resampling. We therefore report bootstrap dispersion measures that quantify the sensitivity of $(\hat{\mu}, \hat{\sigma}, \hat{\alpha}, \hat{\tau})$ to the reconstruction pipeline conditional on the BSR totals. Specifically, for each year we re-run the full mapping and household-level estimation over B household-cluster bootstrap replicates, and summarise uncertainty by bootstrap standard errors and central (percentile) intervals computed from the empirical distribution of $\{\widehat{\boldsymbol{\theta}}^{(b)}\}_{b=1}^{B}$; the resulting dispersion measures are reported in Appendix B, Table B.7.

Table 4.4. PLN estimates on BSR-coherent household microdata: augmented vs non-augmented objective.

Year	Augmented					Non-augmented				
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\tau}$ (€)	μ^H_{model} (€)	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\tau}$ (€)	μ^H_{model} (€)
2010	9.67	1.28	1.67	249,000	40,396	8.93	1.57	1.90	811,000	26,492
2014	9.62	1.38	1.55	713,000	43,227	8.81	1.71	1.65	195,000	27,589
2016	9.69	1.30	1.57	225,000	43,950	9.19	1.48	1.62	341,000	32,469
2020	9.83	1.44	1.79	287,000	52,202	9.51	1.60	1.75	197,000	42,339
2022	10.05	1.27	1.91	507,000	54,957	9.89	1.33	1.99	675,000	49,678

Notes: Parameters are estimated by GMM on household-level bracket shares constructed from the BSR-coherent reconstructed microdata; the augmented specification adds the soft mean-of-holders anchor described in Section 4.5. "Implied mean" is computed from the fitted PLN. Bootstrap dispersion measures are reported in Appendix B, Table B.7.

The augmented estimates deliver a steadily rising model-implied mean from 2010 to 2022, matching the macro pattern in deposits. Dispersion, read jointly through $\hat{\sigma}$ and the mass located in upper brackets, increases into the mid-2010s (2014–2016) and then eases by 2020–2022: despite an uptick in $\hat{\sigma}$ in 2020, the upper-tail mass above €1 mln falls in 2020 and again in 2022 (see shares below), indicating a thinning open top relative to 2014–2016. Overall, the anchor behaves as a stabiliser of the level rather than a hard constraint, with smooth movements in $(\hat{\mu}, \hat{\sigma}, \hat{\alpha}, \hat{\tau})$ across years.

Compared with the anchors in Table 4.2, the augmented implied means show the same overall trends, but are slightly lower in absolute level, closer to the average value in the 100% participation scenario. The discrepancies are rather small and indicate that the added penalty acts as a stabiliser rather than as a hard constraint. For perspective, Table 4.4 also reports the non-augmented bracket-share GMM estimates; model choice remains PLN in every year, but the implied mean is systematically

lower and the fitted tail parameters are less stable across years, consistent with weaker level discipline absent the soft anchor.

Non-augmented fits achieve a similar bracket loss yet imply markedly lower means in early years, which is hard to reconcile with the implied number of deposit accounts we should have (for example, 2010 average implies over 37 million households, while in Italy we roughly had between 23 and 25 million). This is the level-shape decoupling evidenced in the simulations: shares can be matched while the level drifts. Moreover, non-augmented estimates show deeply different values across years, which are not too credible, as from 2010 to 2022 the average deposit holding should increase of approximately 87%.

Figure 4.7 compares household bracket shares (on post-2022 breaks) across the two specifications, year by year; the series are close in each bin, consistent with the simulation evidence. The figure shows how the non-augmented method yields almost equal shares, with slightly heavier shares in the lower breaks and a fainter top-tail. Nonetheless, both methods report a high share of households with deposit holdings above one million euros. Compared to the differences highlighted by the implied average, differences in coverage are negligible.

To illustrate the distributional implications, Figure 4.8 overlays the implied household-level densities from the two estimators, faceted by year. Augmented densities are slightly right-shifted at the mode/shoulder relative to non-augmented ones – the distributional counterpart of higher μ^{H}_{model}. Tail shapes remain similar, but the splice point $\hat{\tau}$ evolves more smoothly under the anchor. From 2014–2016 to 2020–2022 the extreme right tail attenuates (visible in the open-top share and the density's far-right mass), which aligns with the observed reduction in the proportion of households above €1 mln.

Finally, the BSR-coherent household shares allow a direct comparison with HFCS household shares without splitting HFCS into accounts. Table 4.5 reports HFCS and BSR household shares by deposit class (rows: post-2022 breaks) and by year (two columns per year: HFCS vs BSR).

Table 4.5. Household bracket shares by source and year (post-2022 breaks).

	2010		2014		2016		2020		2022	
Deposit class	**HFCS**	**BSR**	**HFCS**	**BSR**	**HFCS**	**BSR**	**HFCS**	**BSR**	**HFCS**	**BSR**
[0, 5k)	2.29	1.30	2.98	1.32	1.84	1.18	1.37	0.94	1.29	0.62
[5k, 12.5k)	13.44	5.04	12.55	4.54	8.38	4.56	6.41	3.41	6.00	3.10
[12.5k, 25k)	21.86	9.41	16.12	8.21	15.86	8.57	10.30	6.59	11.79	6.93
[25k, 50k)	22.91	15.33	15.15	13.50	18.09	14.11	14.64	11.79	16.49	13.22
[50k, 100k)	13.64	18.79	12.44	17.32	12.53	17.60	19.06	16.81	20.85	18.87
[100k, 250k)	8.49	21.88	10.47	22.22	8.82	21.04	18.40	25.06	21.09	26.00
[250k, 500k)	5.04	10.52	4.00	11.84	5.76	10.70	6.43	14.79	9.42	14.11
[500k, 1 mln)	1.53	6.60	2.44	6.89	7.57	7.22	3.43	8.65	4.61	8.00
[1 mln, ∞)	10.79	11.13	23.85	14.17	21.17	15.01	19.96	11.96	8.46	9.16

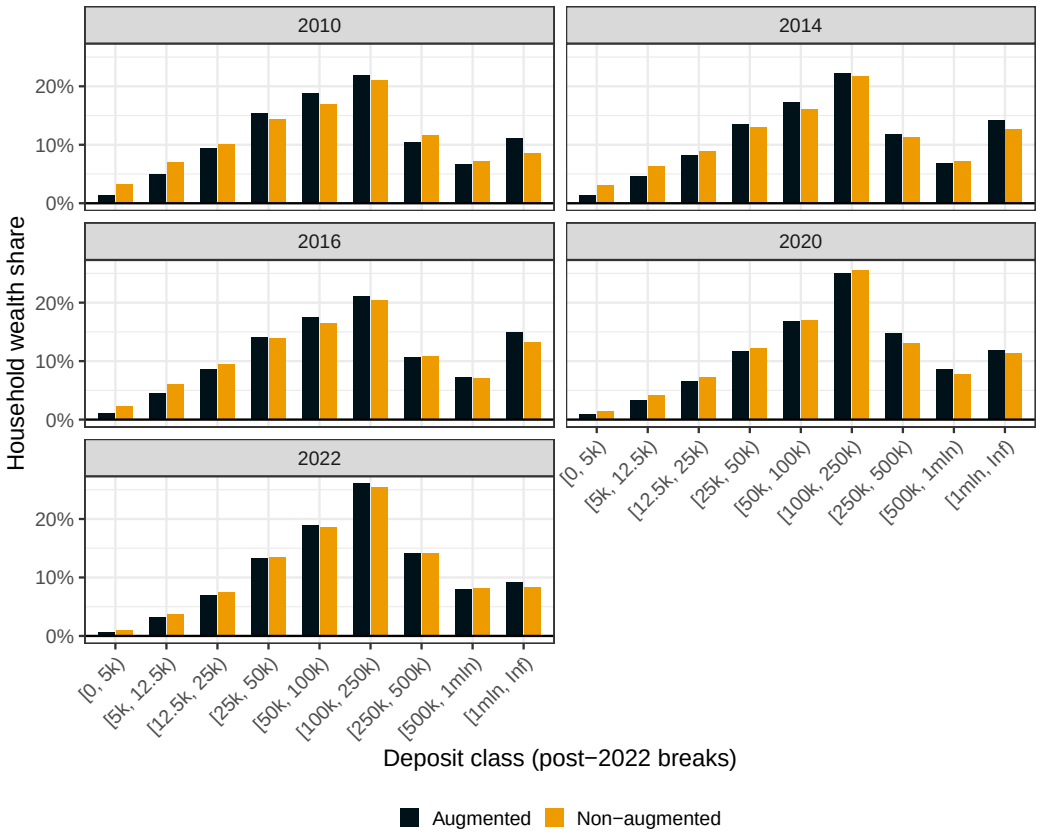

Figure 4.7. Household bracket shares (post-2022 breaks): augmented vs non-augmented, wrapped by year.

BSR allocates substantially more mass than HFCS to the middle and upper-middle brackets [50k, 500k), with compensating lower mass in [5k, 50k). The open-top share (€1 mln) is higher in BSR than HFCS in 2010, then the gap narrows, and by 2020–2022 BSR's open-top share declines to 11.96% and 9.16%, respectively, indicating a thinner extreme tail in the administrative-scope, household-coherent view. This pattern is consistent with survey under-measurement in middle/upper brackets and with the attenuation of the extreme tail post-2016 visible in the densities.

Bracket shares are virtually identical across objectives, but the augmented loss yields level-consistent densities and smoother tail-splice dynamics over time. Mean deposits rise steadily, while dispersion – measured by upper-bracket mass – peaks around 2014–2016 and eases by 2020–2022, when the share above €1 mln falls. This justifies adopting the mean anchor in the empirical application while preserving the bracket geometry required for downstream calibration.

Building on these household-level BSR targets for deposits, the next chapter undertakes the securities-account disentanglement – the final reconciliation of BSR information – recovering instrument-specific marginals (LN/PLN) for listed shares, bonds, and investment fund shares that are consistent with the bracketed account

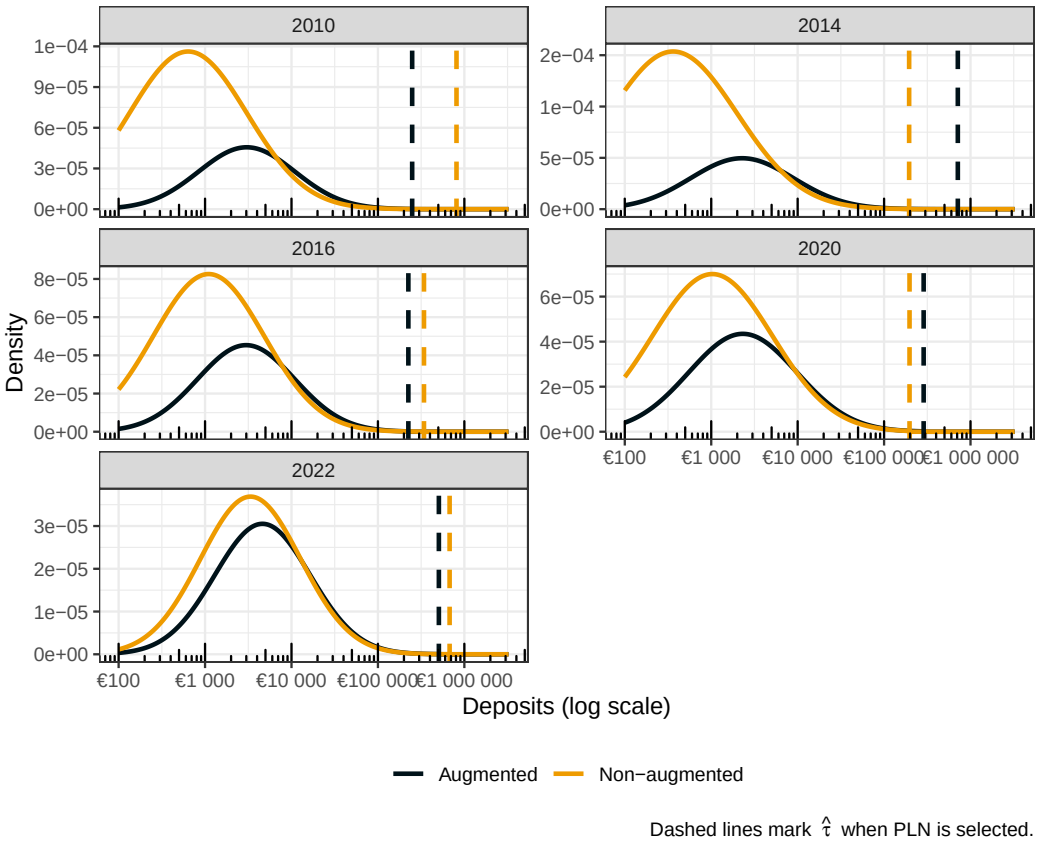

Figure 4.8. Implied household-level densities: augmented vs non-augmented, wrapped by year.

data and supply the BSR-coherent inputs required for the subsequent calibration.

Chapter 5

Securities Disentanglement and Dependence

5.1 Introduction and scope

This chapter addresses a specific limitation of the supervisory information on securities accounts discussed in Section 2.3. The Banking Supervisory Reports (BSR) publish outstanding amounts for listed shares, bonds, and investment fund shares by fixed brackets of the total securities account size. As documented in Section 2.4.3, these three components differ markedly by level and dispersion. Our objective is to recover, for each instrument separately, a credible and parsimonious marginal distribution that is consistent with the BSR bracket information. Concretely, for each of the listed shares, bonds, and investment fund shares, we estimate a semi-continuous marginal with an explicit mass at zero and a flexible positive part, specified as either lognormal (LN) or Pareto-lognormal (PLN) and selected by empirical diagnostics. The resulting instrument marginals are designed both for stand-alone analysis and as inputs to the multivariate calibration developed later in Chapter 6, Section 6.1.3. All reconciliation choices mapping BSR to Quarterly Sector Accounts aggregates are defined in Section 2.3 and are not repeated here.

The motivation is twofold. First, reliable instrument-level marginals permit coherent comparisons across instruments and enable concentration and tail studies analogous to the deposit case in Chapter 4, where account-to-household mapping was addressed explicitly. Second, within this thesis, accurately estimated marginals are a prerequisite for constructing calibration procedures that modify survey microdata while preserving realistic portfolio structure. In Section 6.1.2, the instrument marginals are used to calibrate household values, improving coverage and distributional similarity to BSR-implied kernels; Section 6.1.3 then extends the calibration to a multivariate setting that enforces cross-instrument coherence.

A practical issue is whether account-level and household-level quantities can be treated as equivalent for securities. For deposits, this equivalence does not hold: a non-negligible fraction of households hold multiple accounts, and Chapter 4 shows how the resulting account-to-household mapping is handled. By contrast, evidence

for securities indicates that only a small minority of households hold more than one securities account, and the median number of securities accounts equals one across the distribution of securities holdings (see Section 2.4.3 and Table 2.7). Moreover, the BSR preprocessing that reallocates the total amounts held in the below 50k bracket targets mostly secondary accounts, which further strengthens the account-household working equivalence for securities in this chapter. We therefore proceed under the operational assumption that, for securities, account-level and household-level distributions are close enough for the purposes of estimating instrument marginals. We estimate instrument marginals by matching BSR bracket information, using survey evidence only as side information to stabilise participation and within-bracket allocation.

The chapter yields for each instrument, a semi-continuous marginal by year with a zero mass and a positive part selected between LN and PLN according to data-driven diagnostics, exactly consistent with the BSR bracket information by construction of the loss, and which can be used in later chapters to reconcile HFCS microdata to BSR implied distributions. Moreover, this result also enables instrument-specific concentration analysis, as also explained in Chapter 3. Another additional result this chapter yields is ancillary indicators for false-zero reporting and instrument-specific concentration that can be deployed without additional assumptions.

The scope is deliberately focused. We treat each year separately and do not enforce household-level time consistency across waves. We do not attempt to reconstruct the full joint distribution of instruments, nor to extend the analysis beyond listed shares, bonds, and investment fund shares. Conceptual reconciliation between survey and national accounts is limited to what is implicit in the administrative coverage considered; broader definitional alignment is handled in other parts of the thesis. The aim in this chapter is to deliver transparent, auditable marginals that respect the administrative constraints and are suitable for downstream calibration.

The remainder of the chapter is organised as follows. Section 5.2 sets out the problem, defines observables and targets, and lists the working assumptions. Section 5.3 reviews the literature on recovering distributions from grouped data and on tools for semi-continuous variables. Section 5.4 presents the theoretical framework for the two-block construction. Section 5.5 details the estimation strategy and the simulation-matching algorithm. Section 5.6 reports simulation and empirical results, and Section 5.5.7 gives a concise assessment of the false-zero share in the survey with a preliminary model-free mitigation.

5.2 Problem identification and formalisation

This section formalises the estimation problem implied by the BSR securities tables described in Section 2.3. Let $Y_{ik} \geq 0$ denote holdings for account i of instrument $k \in \{B, L, F\}$ (bonds, listed shares, and investment fund shares). By the working equivalence stated in Section 5.1 and detailed below, account-level holdings are treated as a reasonable approximation to household-level holdings. Define the

zero indicator $Z_{ik} = \mathbb{1}\{Y_{ik} = 0\}$ and the total-securities amount $A_i = \sum_{k\in\{B,L,F\}} Y_{ik}$.

As described in Section 2.3, the BSR report outstanding amounts by value bracket of the total-securities account A. For a given reference year, let $0 = a_0 < a_1 < \cdots < a_J = \infty$ denote the bracket edges (with a coarser partition before 2022 and a finer one from 2022 onwards), and write $\mathcal{J} = \{[a_{j-1}, a_j)\}_{j=1}^{J}$ for the induced partition of the support of A. Let N denote the number of securities accounts recorded in the BSR tables for that year (equivalently, the number of accounts in the BSR universe to which the bracket classification applies). The BSR then provide instrument-specific bracket totals

$$T_{kj} \;=\; \sum_{i=1}^{N} Y_{ik}\, \mathbb{1}\{A_i \in [a_{j-1}, a_j)\}, \qquad k \in \{B, L, F\}, \quad j = 1, \ldots, J,$$

and the corresponding instrument totals $T_k = \sum_{j=1}^{J} T_{kj}$. The equality $T_k = \sum_{i=1}^{N} Y_{ik}$ holds by construction, since the brackets $\mathcal{J}$ partition the support of A and each account i is assigned to exactly one bracket. These objects are unnormalised truncated first moments rather than conditional means: neither the bracket probability $\Pr(A \in [a_{j-1}, a_j))$ nor the conditional mean $\mathbb{E}[Y_k \mid A \in [a_{j-1}, a_j)]$ is observed separately. Moreover, the BSR tables do not report instrument-specific zero counts or holder counts within brackets, and therefore do not identify the bracket-conditional zero probabilities for Y_{ik}. In the proposed strategy, this missing participation information is imported from HFCS for the active population, under the assumption that, conditional on $A_i > 0$, instrument-level zeros are credible.

The working equivalence between account-level and household-level holdings is maintained. HFCS evidence indicates that the vast majority of households with securities hold a single securities account (shown later in the chapter); residual multi-account inflation cannot be identified with current data and is therefore addressed through robustness analysis (Section 5.5.9).

A second caveat concerns the number of independent restrictions delivered by the BSR bracket totals (or their normalised shares). After reconciliation, the securities tables provide three brackets up to 2022 and five brackets from 2022 onwards. In the pre-2022 regime, this yields $3 \times 3 = 9$ instrument-by-bracket share cells, with one restriction implied by global normalisation, hence eight independent conditions. A fully flexible PLN specification for three instruments would require twelve parameters. The model is therefore under-identified in the early years. To obtain a feasible specification, we impose common Pareto-tail parameters across instruments, setting $\alpha_B = \alpha_L = \alpha_F$ and $\tau_B = \tau_L = \tau_F$. This reduces the number of free parameters to eight and allows estimation via the simulated share-matching criterion in Section 5.5. The restriction reduces tail flexibility, but preserves substantial shape variation through instrument-specific body parameters, and identification is primarily driven by the upper brackets, where multi-instrument holdings are more prevalent.

An alternative would be to exploit the finer pre-aggregation bracket structure, which includes an additional < 50k category and would increase the number of

usable conditions. We do not adopt this option, because introducing such a category would weaken the household-account equivalence underlying the reconciliation step, which is more consequential for the present objective than additional tail flexibility. We therefore prioritise a tighter parametric specification that preserves the household-level interpretation of the reconciled BSR aggregates.

In synthesis, the objective is to recover, for each instrument k, the positive-part marginal parameters in the BSR universe, which are required for the subsequent calibration chapters. The BSR shares alone do not identify these marginals because participation and dependence are latent in the aggregates; consequently, HFCS information is used to supply the missing participation and dependence structure, as set out in Section 5.5.

5.3 Literature review

This section reviews methods for recovering component marginals from information that is defined on the sum of components. In our application, the observables are, for each bracket $[a_{j-1}, a_j)$ of the total $A = Y_B + Y_L + Y_F$, the bracket totals T_{kj} for $k \in \{B, L, F\}$. After reconciliation, these BSR statistics are treated as targets after reconciliation (see Section 2.3). HFCS microdata are used only as side information on the zero pattern and on rank-based dependence among positive amounts. Throughout, we target globally normalised shares $\widehat{s}_{kj} = T_{kj}/T$ with $T = \sum_{k,j} T_{kj}$, derived as in Equation (3.8), which removes sensitivity to the overall coverage scaling and to the unknown prevalence of all-zero accounts. The review proceeds in three parts: a concise overview of approaches; implementable alternatives that are closest to our problem; and brief concluding remarks.

To the best of our knowledge, we are not aware of a prior application that recovers instrument-level marginals from totals indexed by brackets of the sum while fixing a two-block copula – governing joint participation and positive-part dependence – from external microdata and estimating the margins by simulate-aggregate-match. Accordingly, we develop a method tailored to these constraints and informed by adjacent literatures.

5.3.1 Concise overview: inversion from information on the sum

A convenient starting point is to recognise that the BSR objects are size-biased truncated first moments over brackets defined by the convolved distribution of A. These are functionals of the joint law of (Y_B, Y_L, Y_F): without additional structure, many different joint distributions can reproduce the same set of range-sum totals. Identification, therefore, hinges on side information about dependence or about at least one margin.

Characteristic-function arguments clarify the basic logic. If the components were independent and the full density of A were observed, then the characteristic function of A factors as the product of the characteristic functions of the components, and one could attempt deconvolution of the sum to recover the margins. In practice, only

range-sum functionals (truncated first moments) are observed, dependence across components is non-negligible, and deconvolution is known to be an ill-posed inverse problem that typically requires regularisation or auxiliary information (Carroll and Hall, 1988; Fan, 1991). These considerations motivate simulation-based strategies that impose a credible dependence structure, generate pseudo-data under candidate marginal parameters, and match model-implied range-sum targets to the BSR observables (in shares).

5.3.2 Implementable alternatives closely related to the problem

This subsection records alternatives that are practically implementable and conceptually close to our setting. Each method is described with a study-specific pointer and a brief account of how it would operate with range-sum information, together with the reasons why the copula-based approach remains preferable for the present problem.

Characteristic-function deconvolution for sums. When components are independent, the characteristic function of the sum factors, and one can invert the factorisation to retrieve the margins. Carroll and Hall, 1988 study optimal rates for deconvolution and show that the problem is ill-posed, with difficulty increasing in the smoothness of the contaminating distribution; Fan, 1991 derive optimal rates in non-parametric deconvolution in errors-in-variables settings. These results explain why regularisation or auxiliary information is required in practice. In our case, only range-sum functionals (truncated first moments) are observed, and the components are dependent, so a pure characteristic-function inversion is not feasible. The dependence supplied by HFCS plays the role of the missing ingredient: once the copula is fixed, the inversion reduces to matching the model-implied bracket shares $[a_{j-1}, a_j)$ rather than recovering densities from the characteristic function (Carroll and Hall, 1988; Fan, 1991).

Regularised inverse problems on a discretised grid. A classic route is to discretise the supports of B, L, F and of A and to view the forward map $(f_B, f_L, f_F) \mapsto \{T_{kj}\}$, where f_k denotes the instrument k density function, as an operator that can be inverted iteratively under positivity and smoothness constraints. The Richardson-Lucy scheme is the canonical iterative deconvolution in astronomy and imaging: Richardson, 1972 introduces a Bayesian iterative restoration algorithm for blurred images in the presence of noise; Lucy, 1974 proposes an iterative rectification technique for deconvolution problems; and later work links these iterations to likelihood-based EM reconstructions for Poisson data in emission tomography (Shepp and Vardi, 1982). Translated to our problem, one could set up an EM or RL-type iteration where the current guess of the component histograms is pushed through the joint operator that bins by A over $[a_{j-1}, a_j)$ and accumulates component totals, and then updated to reduce the discrepancy with BSR targets under non-negativity and smoothing penalties. This is attractive for transparency and for built-in regularisation. However, the forward operator depends on the joint distribution through the dependence between components. Assuming independence

would generally be contradicted by HFCS evidence and would misallocate totals across brackets. Embedding a copula inside the forward operator restores realism, but the method then converges to our simulate-aggregate-match workflow, with the additional complexity of tuning regularisation on coarse, heavy-tailed grids.

Bayesian hierarchical deconvolution with micro data as side information. A coherent alternative is to specify priors for the positive-part marginal parameters and for the copula parameters, write a likelihood for the BSR range-sum totals via simulation, and add a pseudo-likelihood or informative priors for the two copula blocks based on HFCS. In that set-up, posterior draws propagate uncertainty about dependence and margins to the implied bracket totals. This offers a full-probabilistic treatment at the cost of heavier computation and sensitivity to prior choices on tails and on degrees of freedom. The frequentist procedure used in this chapter can be read as the profiling step of such a hierarchical model, and the two approaches would be expected to align when priors are weakly informative.

Moment-based indirect inference for range-sum constraints. Another option is to define estimating equations by equating observed and model-implied bracket shares for each component, where the model-implied shares are computed under a specified copula. This delivers standard errors and over-identification tests in the spirit of grouped-data GMM, but with moments tailored to the BSR statistics rather than to counts or shares of a single variable. The attraction is the availability of a familiar inferential framework. The limitation is that the moments are highly non-linear in the tail parameters, and with few, wide bins $[a_{j-1}, a_j)$, they can be weak. Moreover, the mapping is sensitive to the dependence specification; in practice, once a copula is introduced to compute the model-implied shares, the resulting estimation problem is numerically close to the simulation-criterion approach described above.

Risk-aggregation bounds as diagnostics for identifiability. A complementary perspective comes from copula-based bounds for sums with given margins. Puccetti and Rüschendorf (2015) introduce the rearrangement algorithm to compute sharp bounds on expectations of supermodular functionals under unknown dependence; Embrechts et al. (2013) use these ideas to assess model uncertainty in value-at-risk aggregation. In our setting, such bounds provide a diagnostic for whether different dependence structures could, in principle, produce similar bracket totals even with fixed margins. This helps interpret the role of the HFCS-informed copula: by narrowing the admissible dependence, it makes the inversion from range-sum constraints informative enough to identify the marginal parameters.

5.3.3 Concluding remarks

The methods above can be adapted to the present problem, but they differ in how they handle the entanglement introduced by bracketing on the sum A. Characteristic-function deconvolution is elegant but relies on information that is not available here

and on independence that does not hold. Regularised unfolding is transparent but, without an explicit dependence model, is mis-specified; once a copula is embedded, it collapses to simulation-based matching. Bayesian hierarchical formulations are coherent but computationally more demanding and sensitive to prior choices, while moment-based indirect inference requires a copula to be workable and is numerically similar to the simulation criterion. Against this backdrop, the copula-based simulate-aggregate-match approach is preferable because it supplies exactly the missing joint structure required by the BSR range-sum constraints $[a_{j-1}, a_j)$, separates participation from positive dependence in a way that mirrors the data, and remains parsimonious and robust with coarse aggregation. The next sections implement this strategy and evaluate its performance on simulated and real data.

5.4 Copula theory and related tools

This section collects the copula and collateral theory needed to (*i*) model the dependence between the three semi-continuous security components; (*ii*) separate extensive (Z_k) from intensive margins ($Y_k \mid Y_k > 0$); and (*iii*) transport dependence learned from HFCS to the BSR setting for simulate-aggregate-match estimation. We use (Sklar, 1959; Nelsen, 2006; Joe, 2014) as general references throughout. To simplify notation, we suppress the account index i and write Y_k and $Z_k = \mathbb{1}\{Y_k = 0\}$ for the generic instrument holdings and zero indicator introduced in Section 5.2.

5.4.1 Preliminaries: Sklar's factorisation, copula density, and invariance

Let F be a d-variate CDF with continuous univariate margins F_k. By Sklar's theorem, there exists a unique copula $C : [0,1]^d \to [0,1]$ such that

$$F(y_1, \ldots, y_d) = C(F_1(y_1), \ldots, F_d(y_d)), \qquad y_k \in \mathbb{R}.$$

If F admits a joint density f and univariate densities f_k, then the copula density is

$$c(u_1, \ldots, u_d) = \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\prod_{k=1}^d f_k(F_k^{-1}(u_k))},$$

and the Jacobian factorisation $f(y) = c(F_1(y_1), \ldots, F_d(y_d)) \prod_k f_k(y_k)$ separates marginals from dependence. Copulas are invariant under strictly monotone marginal transforms; consequently, rank-based methods are natural for estimating dependence (Embrechts et al., 2003; Joe, 2014).

5.4.2 Semi-continuous outcomes: a two-block construction

The securities components are semi-continuous: a mass at zero and a continuous distribution on $\mathbb{R}_+$ for positives. With atoms at zero, copula representations are not unique; we therefore adopt an explicit two-block construction that separates participation and positive-part dependence (Joe, 2014).

Block 1 (zeros, bracket-conditional). Let $\mathcal{J}$ denote the fixed BSR partition of total securities A with brackets $[a_{j-1}, a_j)$, and let $j \in \{1, \ldots, J\}$ index a bracket. Write

$$\pi_k(j) = \Pr(Z_k = 1 \mid A \in [a_{j-1}, a_j),\, A > 0),$$

so that $\pi_k(j)$ is defined on the active universe aligned with the BSR tables. We posit a one-factor latent-Gaussian model, for each bracket j,

$$Z_k^*(j) = a_k(j) + \lambda_k(j)F + \varepsilon_k, \qquad F, \varepsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \qquad Z_k = \mathbb{1}\{Z_k^*(j) \le 0\}.$$

This implies

$$\Pr(Z_k = 1 \mid j) = \Phi\Big(\frac{-a_k(j)}{\sqrt{1 + \lambda_k(j)^2}}\Big),$$

and therefore we set

$$a_k(j) = -\sqrt{1 + \lambda_k(j)^2}\; \Phi^{-1}(\pi_k(j)),$$

so that the marginal zero-rate $\pi_k(j)$ is matched exactly in each j. Here $a_k(j)$ is a bracket-specific intercept chosen to match the marginal zero rate $\pi_k(j)$, while $\lambda_k(j)$ is a bracket-specific loading on the common participation factor F that induces dependence between the zero indicators within bracket j. The factor loadings $\lambda_k(j)$ govern co-movements of participation within j and are smoothed over j to reflect the monotone gradient in portfolio breadth. Simulation excludes the all-zero state $(Z_B, Z_L, Z_F) = (1, 1, 1)$ by truncation or rejection, aligning with the BSR universe $A > 0$.

In practice, $\{\pi_k(j), \lambda_k(j)\}$ are first estimated on HFCS quantile bins of A for stability, smoothed as functions of affluence, and then evaluated on the fixed BSR brackets $j \in \{1, \ldots, J\}$.

Block 2 (strictly positive amounts). Conditional on strictly positive amounts, dependence is modelled by an elliptical copula. We use the Student-t copula $C_t(R^+, \nu)$ with correlation matrix R^+ and degrees of freedom $\nu > 2$ to allow symmetric tail dependence (Demarta and McNeil, 2005; Joe, 2014). For finite ν, the t-copula exhibits equal upper and lower tail-dependence coefficients $\lambda_U = \lambda_L > 0$, which is empirically relevant for co-movements of positive positions.

The two blocks are simulated independently, an assumption that simplifies estimation and is consistent with HFCS evidence that participation co-movements can differ from dependence among strictly positive holdings. Robustness checks in Section 5.5.9 vary this assumption.

5.5 Estimation strategy (practical set-up)

This section describes the workflow that turns BSR range-sum aggregates and HFCS micro evidence into instrument-level positive-part marginals for bonds, listed shares, and investment fund shares. The exposition follows the order implemented. Copula

definitions and supporting results are collected in Section 5.4. Computationally, the Monte Carlo approximation of the forward map $\boldsymbol{\theta} \mapsto \widetilde{s}_{kj}(\boldsymbol{\theta})$ is the dominant cost of the procedure. Under the current parallel implementation, a full wave estimation runs in approximately 50 minutes (down from roughly 6 hours in the initial version). This cost also constrains the scale of the simulation study: we can now run $B = 100$ replications per scenario in about 8 hours when omitting the bootstrap tail-testing step, which is therefore excluded from the simulated loops for tractability.

The BSR objects we target are globally normalised instrument-by-bracket shares. For the fixed partition $\mathcal{J} = \{[a_{j-1}, a_j)\}_{j=1}^{J}$ of the total securities value $A = B+L+F$, let T_{kj} denote the instrument-k bracket totals defined in Section 5.2 and set

$$\widehat{s}_{kj} \;=\; \frac{T_{kj}}{T}, \qquad T = \sum_{k\in\{B,L,F\}} \sum_{j=1}^{J} T_{kj}, \qquad \sum_{k}\sum_{j} \widehat{s}_{kj} = 1.$$

Given a parameter vector $\boldsymbol{\theta}$ for the positive-part marginals, the model implies shares $s_{kj}(\boldsymbol{\theta})$. Since the forward map is not available in closed form, $s_{kj}(\boldsymbol{\theta})$ is approximated by Monte Carlo simulation, yielding $\widetilde{s}_{kj}(\boldsymbol{\theta})$. Estimation chooses $\widehat{\boldsymbol{\theta}}$ so that $\widetilde{s}_{kj}(\widehat{\boldsymbol{\theta}})$ matches $\widehat{s}_{kj}$ in a least-squares sense.

5.5.1 Copula-based reconstruction with survey-informed side information

The estimation problem is a deconvolution with side information. The BSR provide information on the sum A through bracketed first-moment totals but do not separately identify (*i*) joint participation across components and (*ii*) dependence among strictly positive amounts. The HFCS is used to supply both ingredients in a way that is consistent with what the BSR brackets can identify: a participation (zero) block and a positive (copula) block. Holding these HFCS-based objects fixed, the remaining unknowns are the positive-part marginals for each instrument, which are estimated to reproduce the BSR range-sum constraints in shares.

Operationally, the workflow is:

(i) Document sparse joint participation in HFCS and derive a dependence summary aligned with the information content of the BSR brackets (Section 5.5.3).

(ii) Estimate the participation (zero) block on HFCS, smooth it over affluence, and evaluate it on the fixed BSR brackets, excluding the all-zero state to align with the BSR universe $A > 0$ (Sections 5.5.4 and 5.5.6).

(iii) Estimate the positive-block dependence on HFCS positives using rank-based measures, and represent it with an elliptical copula (baseline Student-t) so that tail co-movement can be accommodated when present (Demarta and McNeil, 2005) (Section 5.5.4).

(iv) Specify positive-part marginals for each instrument using a LN or a PLN specification, optionally allowing for breadth-dependent level shifts motivated by HFCS evidence (Section 5.5.5).

(v) Combine the two blocks and the candidate marginals in a forward simulator that produces $\widetilde{s}_{kj}(\boldsymbol{\theta})$ on the fixed BSR partition (Section 5.5.6).

(vi) Estimate $\boldsymbol{\theta}$ by minimising a share-matching discrepancy between $\widetilde{s}_{kj}(\boldsymbol{\theta})$ and $\widehat{s}_{kj}$, using common random numbers and multi-start optimisation (Section 5.5.7).

(vii) Assess fit and sensitivity through block-specific diagnostics and robustness exercises; optionally, stabilise levels with a soft mean-of-holders anchor (Sections 5.5.9 and 5.5.10).

Two assumptions are central and are therefore monitored throughout. First, rank dependence among positive amounts in HFCS is a credible proxy for latent dependence in the reconciled universe, so that dependence can be transported without committing to the survey scale. Second, conditional on activity $\{A > 0\}$, instrument-level zeros behave as credible zeros rather than missing values. These assumptions are confronted with diagnostics and sensitivity analysis in Section 5.5.9

The construction is grounded in the standard separation of margins and dependence afforded by copulas (Sklar, 1959; Nelsen, 2006; Joe, 2014). What is distinctive here is the use of survey-informed dependence to unlock identification from range-sum aggregates defined on $A = B + L + F$, without requiring closed-form inversion. In practice, identifiability is assessed locally by verifying that materially different marginal parameter vectors do not yield indistinguishable bracket allocations under the fixed dependence blocks (Section 5.5.9). Finally, computation follows directly from the problem structure: the zero block is obtained from HFCS joint-zero frequencies under a low-dimensional latent-threshold representation; the positive block is estimated from HFCS pseudo-observations; and the simulated criterion is stabilised through common random numbers and multi-start optimisation. Uncertainty can be assessed by bootstrapping the HFCS dependence estimates and, separately, by perturbing the BSR targets within the reconciliation bounds from Section 2.3.

5.5.2 Data and working assumptions

BSR targets. The BSR provide, for each bracket $[a_{j-1}, a_j)$ of $A = B + L + F$, the instrument totals T_{kj}. These are size-biased truncated first moments contributed by active accounts: accounts with $(B, L, F) = (0, 0, 0)$ never enter and are therefore invisible. We target $\widehat{s}_{kj} = T_{kj}/T$, which removes sensitivity to residual coverage scaling after reconciliation and is unaffected by the unknown prevalence of triple zeros.

HFCS side information. HFCS microdata provide household-level B, L, F with sampling weights and many zeros. Two inputs are transported from HFCS to the BSR setting. First, bracket-conditional participation profiles are imported for the active universe $A > 0$: the zero rates $\pi_k(j)$ and a low-dimensional representation of joint participation (a one-factor latent-Gaussian block). Second, dependence among strictly positive amounts is imported through an elliptical copula estimated from weighted pseudo-observations, exploiting the invariance of rank dependence to

monotone marginal transforms (Section 5.4).

The working transport assumptions are those stated in Section 5.2. In particular, rank correlation among positives is assumed broadly preserved under monotone scale distortions, and instrument-level zeros are assumed credible within the active set. Relative to the DWA's proportional-allocation view – which implicitly assumes a common proportional bias and credible zeros for everyone – these assumptions are a relaxation: we allow heterogeneous distortions at the positive level, require only within-instrument rank preservation, and keep the credible-zero assumption within the active set. The differing participation profiles in HFCS – bonds more often held stand-alone and funds more often co-held – are handled explicitly by the zero block and carried over to simulation.

5.5.3 Dependence under sparse joint participation: HFCS evidence and implications

A distinctive feature of securities holdings in HFCS is the prevalence of single-instrument portfolios outside the top of the securities distribution. This matters for our estimator because the BSR objects $\widehat{s}_{kj}$ are size-biased, bracketed first moments: large-A accounts contribute disproportionately to T_{kj}, so the upper brackets carry most of the identifying power for cross-instrument dependence, whereas the lower brackets mainly discipline participation and the one-dimensional positive parts. When joint positives are rare in a bracket, the simulated range-sum totals in that bracket depend almost entirely on the marginals and the zero rates, with the positive-block dependence entering only weakly.

We therefore document portfolio breadth in HFCS and use it to guide the dependence summary fed into the simulator. We form weighted quantiles (quintiles) of total securities A, restrict to the active sample $\{h : Y_{hB} + Y_{hL} + Y_{hF} > 0\}$, and tabulate the weighted shares of households holding 1, 2, or 3 instruments. In the Italian waves considered here, the bottom 80% of the A distribution is dominated by single-instrument portfolios, while multi-instrument portfolios become frequent only in the top quintile. The pattern is stable across waves (Table 5.1).

The sparse-joint-participation pattern motivates two modelling choices. First, the participation block is estimated and smoothed as a function of affluence and then evaluated on the fixed BSR brackets. Second, the positive-block dependence is estimated on pooled upper-quantile joint positives (e.g., the top decile), because lower brackets contribute little information about cross-instrument co-movement once the BSR objects are size-biased. The breadth effect observed in HFCS – conditional on holding, single-instrument amounts are typically smaller than multi-holder amounts – motivates an optional breadth-dependent shift in the lognormal body (Section 5.5.5). Table B.8 in Appendix B summarises the distribution of portfolio types by wave and quintile of A.

Table 5.1. HFCS portfolio composition by securities-account quintile and wave for HFCS households with $A > 0$.

Wave	Quintile group	Number of non-zero holdings					
		1 instrument		2 instruments		3 instruments	
		HH	Share	HH	Share	HH	Share
1	bottom 80%	1,156	83.32%	210	14.99%	24	1.69%
	top 20%	206	48.24%	152	35.02%	55	16.74%
2	bottom 80%	1,037	82.76%	181	15.48%	27	1.77%
	top 20%	182	58.64%	111	28.36%	50	13%
3	bottom 80%	778	84.04%	138	14.3%	14	1.66%
	top 20%	143	53.61%	93	32.61%	36	13.78%
4	bottom 80%	797	79.64%	238	17.33%	41	3.03%
	top 20%	186	47.97%	195	36.82%	143	15.21%
5	bottom 80%	1,103	80.07%	269	17.48%	62	2.45%
	top 20%	267	42.43%	238	34.42%	199	23.16%

Notes: Weighted HFCS shares of active households, where active means at least one of $\{B, L, F\}$ is strictly positive. Columns report the percentage of households with exactly 1, 2, or 3 instruments within the quintile groups of total securities A. Totals report the number of households (column HH) and the weighted percentage (column Share).

5.5.4 Rank-based estimation of dependence from HFCS

Zeros (bracket-conditional; estimated on HFCS, evaluated on BSR brackets). Estimate $\pi_k(j)$ as weighted zero rates in HFCS bins of A, and infer a low-dimensional loading vector $\lambda(j) = (\lambda_B(j), \lambda_L(j), \lambda_F(j))$ by matching bin-wise pairwise tetrachorics to a one-factor structure, with weakly increasing smoothing across j. Evaluating the smoothed profiles on the fixed BSR brackets yields $\{\pi_k(j), \lambda_k(j)\}_{j=1}^{J}$. In simulation, the all-zero state is excluded by truncation or rejection to align with the BSR universe $A > 0$.

Positives. Let $\mathcal{H}_{k,+} = \{h : Y_{hk} > 0\}$ denote the HFCS strictly positive subsample for instrument k, with survey weights $\{w_h\}$ and total weight $W_{k,+} = \sum_{h \in \mathcal{H}_{k,+}} w_h$. We define the weighted empirical CDF on positives by the weighted midrank transform

$$\widehat{F}^{w}_{k,+}(y) \;:=\; \frac{1}{W_{k,+}} \sum_{h \in \mathcal{H}_{k,+}} w_h \left(\mathbb{1}\{Y_{hk} < y\} + \frac{1}{2} \mathbb{1}\{Y_{hk} = y\} \right),$$

and set the pseudo-observations as $U_{hk} = \widehat{F}^{w}_{k,+}(Y_{hk}) \in (0, 1)$, defined for $h \in \mathcal{H}_{k,+}$. The midrank convention ensures that U_{hk} remains in the open unit interval even in the presence of ties. For elliptical copulas, Kendall's τ maps to the linear correlation

via

$$\rho_{k\ell} \;=\; \sin\!\left(\frac{\pi}{2}\,\tau_{k\ell}\right),$$

for both Gaussian and t copulas (Lindskog et al., 2003; Demarta and McNeil, 2005). We therefore estimate R^+ by computing weighted $\widehat{\tau}_{k\ell}$ on pairs with joint positivity (pooled in the upper tail) and then transforming to $\widehat{\rho}_{k\ell}$, with a final nearest-correlation projection if needed. As a robustness alternative, one can maximise the weighted copula pseudo-likelihood $\sum_h w_h \log C_t(U_{h\cdot}; R^+, \nu)$ (Tsukahara, 2005; Genest and Rémillard, 2008).

These HFCS estimates are then treated as side information. In the simulate-aggregate-match step we fix the zero block at the smoothed bracket-conditional profiles $\{\pi_k(j), \lambda_k(j)\}$ and fix the positive-block dependence at the pooled upper-tail R^+, while allowing the positive-part marginals $F_k^+(\cdot\,;\vartheta_k)$ to vary.

5.5.5 Marginal model for positives: Pareto-lognormal splice

For each component k, the strictly positive margin $F_{k,+}$ is specified as a PLN. Denote the body parameters by (μ_k, σ_k), the tail index by $\alpha_k > 1$ (finite first moment), and the splice threshold by τ_k. In the empirical implementation, τ_k is profiled on a grid aligned with BSR brackets, as in Chapter 3.

To reflect the breadth effect observed in HFCS, we allow an optional breadth-dependent location shift γ_k, calibrated from HFCS and treated as fixed in the simulation step; γ_k captures the average log-difference in Y_k between multi-instrument holders and single-instrument holders within the active population.

$$\log Y_k \;=\; \mu_k + \gamma_k\,\mathbb{1}\{K \geq 2\} + \sigma_k\,\varepsilon, \qquad K = \sum_r (1 - Z_r),$$

applied to the lognormal body (PLN) or to the LN mean.

5.5.6 From dependence and margins to simulation

Given $\{\pi_k(j), \lambda_k(j)\}_{j=1}^J$, $\widehat{R}^+$ (and $\widehat{\nu}$ if applicable), fixed BSR bracket weights $\{\omega_j\}_{j=1}^J$, and a vector of candidate marginal parameters $\boldsymbol{\theta} = \{\vartheta_k\}_{k=B,L,F}$, proceed as follows:

(i) Draw $j \sim \text{Categorical}(\omega_1, \ldots, \omega_J)$.

(ii) Draw the participation factor $F \sim \mathcal{N}(0,1)$ and idiosyncratic shocks $\varepsilon_k \sim \mathcal{N}(0,1)$; set $Z_k = \mathbb{1}\{a_k(j) + \lambda_k(j)F + \varepsilon_k \leq 0\}$, with $a_k(j)$ chosen so that $\Pr(Z_k = 1 \mid A \in [a_{j-1}, a_j), A > 0) = \pi_k(j)$ (Section 5.4). If $Z_B = Z_L = Z_F = 1$, reject and redraw to enforce activity ($A > 0$).

(iii) Draw U^+ from the Student-t copula $C_t(\widehat{R}^+, \widehat{\nu})$ independently of Step *(ii)*.

(iv) For each k, set $Y_k = 0$ if $Z_k = 1$; else set the positive amount by the LN/PLN quantile and the breadth shift:

$$\log Y_k \leftarrow \log F_{k,+}^{-1}(U_k^+; \vartheta_k) \;+\; \gamma_k\,\mathbb{1}\{K \geq 2\}, \qquad K = \sum_r (1 - Z_r).$$

Here $F_{k,+}^{-1}(u;\vartheta_k)$ denotes the generalised inverse (quantile function) of the strictly positive marginal $F_{k,+}(\cdot;\vartheta_k)$, for $u \in (0,1)$.

Aggregating by $A = B + L + F$ and binning by $[a_{j-1}, a_j)$, $j = 1, \ldots, J$, produces simulated bracket totals $\widetilde{T}_{kj}(\boldsymbol{\theta})$ and simulated shares

$$\widetilde{s}_{kj}(\boldsymbol{\theta}) \;=\; \frac{\widetilde{T}_{kj}(\boldsymbol{\theta})}{\sum_{k',j'} \widetilde{T}_{k'j'}(\boldsymbol{\theta})}.$$

Common random numbers are fixed across evaluations of $\boldsymbol{\theta}$ to stabilise the objective.

5.5.7 Targets, side information, and methodology

From BSR we work with the observed shares $\widehat{s}_{kj}$. From HFCS we estimate, prepare, and then fix for calibration: (i) the bracket-conditional participation profiles $\{\pi_k(j), \lambda_k(j)\}_{j=1}^{J}$ (estimated on HFCS quantiles, evaluated on fixed BSR brackets); and (ii) a positive-block correlation R^+ (from weighted Kendall's τ on joint positives in the upper tail), with a Student-t degrees-of-freedom parameter ν when tail co-movement is empirically present. The unknowns are the positive-part marginal parameters $\boldsymbol{\theta} = \{\vartheta_k\}_{k=B,L,F}$, with $\vartheta_k = (\mu_k, \sigma_k)$ for LN and $\vartheta_k = (\mu_k, \sigma_k, \alpha_k, \tau_k)$ for PLN. Given side information and candidate $\boldsymbol{\theta}$, simulation produces $\widetilde{s}_{kj}(\boldsymbol{\theta})$; estimation chooses $\boldsymbol{\theta}$ to align $\widetilde{s}_{kj}(\boldsymbol{\theta})$ with $\widehat{s}_{kj}$.

The baseline criterion is defined on shares:

$$Q(\boldsymbol{\theta}) = \sum_{k,j} w_{kj} \left(\widetilde{s}_{kj}(\boldsymbol{\theta}) - \widehat{s}_{kj}\right)^2, \qquad w_{kj} = 1 \text{ at baseline.}$$

Defining the loss on shares prevents the top bin from dominating mechanically and keeps the target comparable across years with different bracket partitions.

Optimisation is staged to stabilise the simulated criterion in a heavy-tailed setting.

Step 1: PLN-for-all coarse fit. We begin by fitting all three margins as PLN. The splice thresholds τ_k are profiled on a grid aligned with $\{a_j\}$ (plus a few internal knots), as in Chapter 3. Global optimisers are unnecessarily slow and brittle on simulated criteria; instead, we run many non-overlapping random starts within wide but sensible bounds and pass each to a box-constrained local optimiser (L-BFGS-B), keeping the best solution in terms of $Q(\boldsymbol{\theta})$.

Step 2: Tail-form check (LN vs PLN) by margin. With the Step-1 dependence blocks fixed, we test, instrument by instrument, whether PLN is warranted relative to LN. Holding the other two margins at their Step-1 fits, we re-fit the best LN and best PLN for margin k on the same BSR shares and with identical simulation draws, and compute $\Delta Q_k = Q_{\text{LN}} - Q_{\text{PLN}}$. A positive ΔQ_k favours PLN. When needed, we implement a parametric bootstrap under the LN null: we

simulate pseudo-tables with k constrained to LN at its fitted parameters, re-estimate LN and PLN on each pseudo-table, and form an empirical p-value for ΔQ_k. This comparison mirrors the LN-versus-PLN tail check in Chapter 3 (Section 3.4): we measure improvement by the drop in the same share-based loss used for estimation and, when needed, calibrate it by a parametric bootstrap under the LN null; the only difference is that here the loss is computed from simulated multivariate shares conditional on the HFCS-informed dependence blocks.

Step 3: Refit with selected forms. Given the selected form for each margin, we re-estimate the parameters using the same τ_k grid (for those margins that remain PLN) and the same multi-start scheme. If all three margins were already correctly specified as PLN in Step 1, this step is skipped.

Step 4: Fine-tune body parameters. Finally, we fix τ_k and α_k at their selected values and re-optimise only (μ_k, σ_k) inside narrow boxes centred at the current estimates (e.g., $\mu_k \pm 0.6$, $\sigma_k \pm 0.3$), using a denser set of starts than in Step 1. Once tail mass is pinned down, curvature in (μ, σ) sharpens, and this pass stabilises the fit in the body. The output of this step is the final parameter vector $\widehat{\boldsymbol{\theta}}$.

False zeros implied by HFCS totals

Another secondary, but still noticeable, result this method yields is a preliminary assessment of the false-zero share in the HFCS. Let the HFCS positive set for instrument k be $\mathcal{S}_{v,k} = \{h \in \{1, \dots, H\} : Y_{hk}^{\mathrm{HFCS}} > 0\}$, with survey weights $\{w_h\}$ and weighted size $H_{vk} = \sum_{h \in \mathcal{S}_{v,k}} w_h$, and denote its complement by $\mathcal{S}_{nv,k}$. Map the weighted empirical ranks of Y_{hk}^{HFCS} on $\mathcal{S}_{v,k}$ into the BSR-estimated positive marginal by the inverse F_k^+ transform, analogously to the deposits mapping in Chapter 4; denote the mapped values by Y_{kh}^q. By construction of this rank-preserving quantile mapping, the positive-part expectation is matched:

$$\mathbb{E}[Y_{hk}^q \mid Y_{hk}^q > 0] = \mathbb{E}[Y_k^{\mathrm{BSR}} \mid Y_k^{\mathrm{BSR}} > 0].$$

The HFCS-implied total after mapping is

$$\widehat{T}_k \;=\; \sum_{h \in \mathcal{S}_{v,k}} w_h \, Y_{hk}^q,$$

and satisfies

$$\mathbb{E}[\widehat{T}_k] \;=\; H_{vk} \, \mathbb{E}[Y_k^{\mathrm{BSR}} \mid Y_k^{\mathrm{BSR}} > 0]. \tag{5.1}$$

On the BSR side,

$$\mathbb{E}[T_k] \;=\; \mathbb{E}\Big[\sum_{i=1}^{N} Y_{ik}\Big] \;=\; N \, \mathbb{E}[Y_k^{\mathrm{BSR}} \mid Y_k^{\mathrm{BSR}} > 0], \tag{5.2}$$

with $T_k = \sum_j T_{kj}$ and N counting BSR positives for instrument k. Combining (5.1)–(5.2) yields

$$\frac{H_{vk}}{N} \;=\; \frac{\mathbb{E}[\widehat{T}_k]}{\mathbb{E}[T_k]} \qquad \Rightarrow \qquad r_k^{fz} \;\equiv\; 1 - \frac{H_{vk}}{N} \;\approx\; 1 - \frac{\widehat{T}_k}{T_k},$$

which provides a preliminary estimate of the HFCS false-zero share for instrument k (r_k^{fz}). Here N denotes the (latent) BSR count of positives for instrument k and is not observed; it cancels in the expectation ratio, which motivates the empirical plug-in $\widehat{T}_k/T_k$ under matched reference years. An alternative, but equivalent, estimation procedure relies on estimating $\widehat{T}_k$ as the expected value given by the estimated marginal, multiplied by the HFCS weighted number of households with a positive holding of instrument k, namely

$$\widehat{T}_k \;=\; \mathbb{E}[Y_k^{\mathrm{BSR}} \mid Y_k^{\mathrm{BSR}} > 0] \sum_{h=1}^{H} w_h \, \mathbb{1}\{Y_{hk} > 0\}\,.$$

We apply both estimation procedures, as they may be susceptible to different biases. The comparison is carried out after the reconciliations in Section 2.3.

The equality in expectations underlying the zero-mass inference relies on a correctly specified positive-part marginal used in the quantile mapping; severe misspecification would bias r_k^{fz}. Any dependence between A and the instrument components is handled within the simulation-matching step rather than in the zero-share algebra. The remainder of the chapter develops the copula construction for semi-continuous outcomes and the estimation set-up that implements this scheme.

5.5.8 Specialisation to this thesis

In the empirical implementation, we adopt the following choices. First, the bracket-conditional zero block $\{\pi_k(j), \lambda_k(j)\}$ is estimated with survey weights and monotone smoothing on HFCS and evaluated on the fixed BSR brackets $j = 1, \ldots, J$; the simulated process is truncated to exclude the all-zero state, aligning the extensive margin with the BSR universe. Second, for the positive block we use a Student-t copula with R^+ estimated by weighted Kendall's τ and ν fixed at a value supported by HFCS diagnostics; robustness varies ν. Third, the PLN splice is profiled over a threshold grid and then refined by continuous optimisation of body and tail parameters, consistent with Chapter 3; breadth-dependent shifts γ_k are fixed from HFCS. Finally, estimation targets the observed shares $\widehat{s}_{kj}$ on the fixed partition, which eliminates sensitivity to the unknown prevalence of triple zeros and to residual coverage scaling in BSR after reconciliation.

5.5.9 Diagnostics and robustness

Diagnostics target both blocks and the stability of the share-matching objective. For the zero block, compare observed and model-implied joint-zero frequencies by bracket j and inspect the spectrum of $\widehat{\lambda}_k(j)$ for monotonicity. For the positive block, report weighted sample versus model Kendall's τ and, when ν is finite, the implied

tail-dependence coefficients against empirical tail co-exceedance rates. Goodness-of-fit for copulas can be assessed via parametric bootstrap of Cramér–von Mises-type functionals of the empirical copula (Genest and Rémillard, 2008); in practice, we prioritise stability of the simulate-aggregate-match objective over formal copula goodness-of-fit, given the coarse BSR aggregation. An additional specification check uses the Rosenblatt transform.

Robustness exercises vary the most consequential dependence choices: pooling of the upper tail for R^+, shrinkage toward a bracket-constant R^+, and setting $R^+ = 0$ below the top quantile. These exercises are interpreted in light of the size bias of $\widehat{s}_{kj}$, which concentrates identification of cross-instrument dependence in the upper brackets.

5.5.10 Augmenting the GMM with a soft mean-of-holders moment

We optionally enrich the share-matching criterion with a soft moment on the mean among holders. For each instrument $k \in \{B, L, F\}$ and wave t, let T_{kt} denote the BSR/QSA aggregate. We construct a conservative upper bound for the mean among holders by first estimating the number of positive holders from HFCS and then augmenting it with a basic imputation based on securities account reporting. Specifically, define

$$\begin{aligned} N_{kt}^{H,+,\text{obs}} &= \sum_h w_h \, \mathbb{1}\{Y_{hk} > 0\}, \\ N_t^{\text{acc},0} &= \sum_h w_h \, \mathbb{1}\{A_h = 0 \text{ and reports at least one securities account}\}, \end{aligned}$$

and estimate the conditional positive probability within the active set,

$$\pi_{kt}^{+} := \Pr(Y_k > 0 \mid A > 0),$$

computed on HFCS and smoothed over t. The basic imputation allocates the $A = 0$ but account-reporting mass across instruments using π_{kt}^{+}, yielding the augmented count

$$N_{kt}^{H,+,\text{lb}} = N_{kt}^{H,+,\text{obs}} + \pi_{kt}^{+} N_t^{\text{acc},0}.$$

Because securities accounts are prone to under-reporting of positive positions, $N_{kt}^{H,+,\text{lb}}$ is a lower bound for the true number of positive holders. We therefore treat

$$\overline{\mu}_{kt}^{H,\max} = \frac{T_{kt}}{N_{kt}^{H,+,\text{lb}}} \tag{5.3}$$

as a conservative upper bound for the mean among holders.

We then define the augmented loss used in this chapter. Let $\mu_{kt}^{H}(\boldsymbol{\theta})$ be the model-implied mean among holders under $\boldsymbol{\theta}$ (either the closed-form LN/PLN mean or a simulation estimate; we use the parametric form at baseline). Define

$$\begin{aligned} r_{kt}(\boldsymbol{\theta}) &= \log\left(\frac{\mu_{kt}^{H}(\boldsymbol{\theta})}{\overline{\mu}_{kt}^{H,\max}}\right), \\ \delta_t &= \log(1 - \text{tol}_{\downarrow,t}) < 0, \end{aligned}$$

with $\mathrm{tol}_{\downarrow,t} \in [0,1)$ a tolerance for being below the upper bound. The asymmetric two-sided penalty per margin is

$$\begin{aligned} \mathcal{P}_{kt}(\boldsymbol{\theta}) \;&= \lambda_{\uparrow,t}\,\big[r_{kt}(\boldsymbol{\theta})_+\big]^2 \;+\; \lambda_{\downarrow,t}\,\big[\,(-r_{kt}(\boldsymbol{\theta})-\delta_t)_+\,\big]^2, \\ x_+ &:= \max\{0,x\}, \end{aligned} \tag{5.4}$$

where $\lambda_{\uparrow,t}$ (overshoot) and $\lambda_{\downarrow,t}$ (large undershoot) are non-negative weights. The wave-t augmented criterion is

$$\mathcal{L}_t(\boldsymbol{\theta}) \;=\; \underbrace{\sum_{k,j} w_{kj}\,\big(\widetilde{s}_{kj}(\boldsymbol{\theta})-\widehat{s}_{kj}\big)^2}_{Q_t(\boldsymbol{\theta})} \;+\; \sum_{k\in\{B,L,F\}} \mathcal{P}_{kt}(\boldsymbol{\theta}), \qquad w_{kj}=1 \text{ at baseline.} \tag{5.5}$$

The penalty is scale-free and does not disturb the bracket scaling; it stabilises the level of each marginal while brackets discipline shape.

For LN, $\mu^H_{kt}(\boldsymbol{\theta}) = \exp(\mu_k + \frac{1}{2}\sigma_k^2)$. For PLN (smooth-fit splice at τ_k with tail index $\alpha_k > 1$), write c_k for the fitted lognormal mass below τ_k. The parametric first moment among holders is

$$\mu^H_{kt}(\boldsymbol{\theta}) \;=\; \underbrace{\exp(\mu_k+\tfrac{1}{2}\sigma_k^2)\,\Phi\Big(\tfrac{\log\tau_k-\mu_k-\sigma_k^2}{\sigma_k}\Big)}_{\text{body contribution}} \;+\; \underbrace{(1-c_k)\,\frac{\alpha_k\tau_k}{\alpha_k-1}}_{\text{Pareto tail}}.$$

Concerning $\mathrm{tol}_{\downarrow,t}$ and $(\lambda_{\downarrow,t}, \lambda_{\uparrow,t})$, we follow two principles. Tolerances reflect survey credibility by wave:

$$\mathrm{tol}_{\downarrow,t} = \begin{cases} 0.50, & t \in \{1,2,3\}, \\ 0.15, & t = 4, \\ 0.10, & t = 5, \end{cases} \qquad \delta_t = \log(1-\mathrm{tol}_{\downarrow,t}).$$

Weights are set to be commensurate with the scale of the share loss. Let $L_{\max,t}$ be the per-wave upper bound for the share loss and fix a reference level $L_t^{\mathrm{ref}} = \eta_t\, L_{\max,t}$ with $\eta_t \in (0,1)$. Choose design deviations at which the penalty should contribute L_t^{ref}: a below-bound factor $b_{\downarrow,t} < 1$ and an above-bound factor $b_{\uparrow,t} > 1$. With

$$u_t^\star = \max\{0, -\log b_{\downarrow,t} - \delta_t\}, \qquad v_t^\star = \log b_{\uparrow,t},$$

set

$$\lambda_{\downarrow,t} \;=\; \eta_{\downarrow,t}\,\frac{L_t^{\mathrm{ref}}}{(u_t^\star)^2}, \qquad \lambda_{\uparrow,t} \;=\; \eta_{\uparrow,t}\,\frac{L_t^{\mathrm{ref}}}{(v_t^\star)^2}, \tag{5.6}$$

with $\eta_{\downarrow,t}, \eta_{\uparrow,t} \in (0,1)$. In practice, $b_{\downarrow,t}$ is looser in early waves and tighter in later waves (for example $0.35 \to 0.75$), while $b_{\uparrow,t}$ is kept around 1.5.

Table 5.2 contrasts the upper-bound mean $\overline{\mu}^{H,\max}_{kt}$ with the model-implied mean from the non-augmented fit, by wave and instrument; the last column reports the model to anchor ratio (per cent).

Table 5.2. Upper-bound mean versus non-augmented model mean among holders (securities components).

Wave	Instrument	QSA upper bound (€)	Model mean (€)	Model / QSA (per cent)
1	Bonds	192,026	50,877	26.49
1	Shares	73,414	18,773	25.57
1	Funds	199,238	52,478	26.34
2	Bonds	161,156	18,283	11.35
2	Shares	95,521	6,491	6.80
2	Funds	311,649	34,326	11.01
3	Bonds	123,850	17,703	14.29
3	Shares	87,112	7,117	8.17
3	Funds	320,799	50,755	15.82
4	Bonds	95,686	20,688	21.62
4	Shares	71,799	14,526	20.23
4	Funds	324,134	81,199	25.05
5	Bonds	91,732	43,909	47.87
5	Shares	100,137	32,656	32.61
5	Funds	251,124	107,781	42.92

Notes: $N_{kt}^{H,+,\mathrm{lb}} = N_{kt}^{H,+,\mathrm{obs}} + \pi_{kt}^{+} N_t^{\mathrm{acc},0}$. "QSA upper bound" is $\overline{\mu}_{kt}^{H,\max} = T_{kt}/N_{kt}^{H,+,\mathrm{lb}}$ from (5.3); "Model mean" is the expected value from the non-augmented disentanglement. The last column reports 100 × Model/QSA.

The mean-of-holders anchor respects the direction of the survey bias, is activated only when informative via $\mathrm{tol}_{\downarrow,t}$, and is scaled to the share loss through (5.6). Empirically, it disciplines implausible movements in the level of the margins across waves without compromising the fit on bracket shares.

5.6 Results

5.6.1 Simulated data

The method is first evaluated under controlled scenarios that mimic the main challenges of the real application. Starting from known marginal laws and a known copula, we generate pseudo-BSR shares by binning simulated A over $[a_{j-1}, a_j)$ and accumulating instrument-specific truncated first moments. We then generate a pseudo-HFCS micro sample from the same joint distribution and apply instrument-specific multiplicative biases to positive amounts while leaving zeros unchanged. From this "biased micro" we re-estimate the bracket-conditional participation profiles $\{\pi_k(j), \lambda_k(j)\}$ and the positive-block dependence R^+ as with real data, and then run the four-step pipeline treating these dependence blocks as fixed. Five scenarios vary body location and dispersion, tail heaviness and thresholds, marginal forms

(LN vs PLN by instrument), dependence strength and tail dependence, and the size and variability of the applied biases.

Two design choices clarify what the simulations are intended to validate. First, the Monte Carlo evidence focuses on the behaviour of the estimator conditional on the correct marginal family class. LN–PLN discrimination is computationally intensive because each decision requires repeated re-estimation under resampled pseudo-tables; we therefore run the tail-test block only as a diagnostic on a subset of Base-scenario replications. This separation mirrors the empirical setting, where the three security components are expected to be strongly heavy-tailed and PLN is the relevant class for all margins.

Second, the simulation adopts an informative ("best-case") bracket design and allows margin-specific tail parameters (α_k, τ_k). This isolates the behaviour of the estimator under adequate tail identification, avoiding confounding from coarse-bracket non-identification, and it prevents scenarios with genuinely lognormal margins from being distorted by an imposed shared-tail restriction. In the empirical application, a shared-tail specification is primarily a stabiliser under limited bracket information; reproducing it systematically in a large-scale Monte Carlo would add substantial computational time and reporting burden without improving validation of the core disentanglement mapping. Moreover, identification issues are materially alleviated under the post-2022 BSR break design, so best-case evidence is directly informative for the high-resolution perimeter that will matter operationally. In short, BSR provide truncated first-moment targets by brackets of the sum; HFCS provides participation and rank-dependence blocks; the copula transports that structure; and a share-based simulation criterion pins down the positive-part marginals with enough discipline to be informative under bracketed aggregates.

This subsection examines the non-augmented estimator introduced in Section 5.5.7. The simulation design mirrors the empirical setting: BSR-style bracket totals for the sum $A = Y_B + Y_L + Y_F$, semi-continuous margins with participation zeros, and copula-driven dependence separating the zero block from the positive block. Throughout, the bracket partition and sample sizes are kept fixed to match the calibration stage. Specifically, we set $n_{\text{macro}} = 150{,}000$, $n_{\text{micro}} = 8{,}000$, and a population target of $n_{\text{accounts}} = 50{,}000$ under the account-household working equivalence for securities, with bracket cut-offs (0, 50,000, 100,000, 200,000, 500,000, 1,000,000, ∞).

To balance numerical tightness and replication scale, we report two Monte Carlo regimes. Parameter recovery for $(\mu, \sigma, \alpha, \tau)$ and LN–PLN model choice are evaluated on $R = 15$ replications of the Base scenario. These runs use tighter optimisation hyperparameters and implement the full pipeline, including the LN–PLN tail-test block and τ-profiling. By contrast, body parameters, bracket-fit and dependence diagnostics are evaluated on $R = 100$ replications per scenario, focusing on objects that are both operationally central and numerically stable at scale: body parameters (which deal with the greatest share of the probability density and allow for more parsimonious runs without loss of precision), share recovery by bracket, dependence among positives R^+, CAF overlays, and top-bracket share errors by instrument.

The evidence is organised in two parts. First, we provide a deep-dive on the Base scenario with $R = 15$ replications, featuring PLN-type tails where present, moderate cross-instrument dependence, and moderate participation zeros. Second, we present a compact sensitivity analysis over five scenarios, each evaluated with $R = 100$ replications:

- **S 01:** Base scenario;
- **S 02:** Heavy zeros and stronger tail on Y;
- **S 03:** Near-LN (no tails), moderate zeros;
- **S 04:** Strong positive dependence (t-copula) and tail on Y;
- **S 05:** Strong zero-block dependence and HFCS stress.

The estimator is evaluated along three dimensions. First, we report tail-test outcomes (LN vs PLN) as diagnostic evidence in the Base subset, as well as parameters estimation. Second, conditional on the correct marginal family class, we measure body parameter recovery with

$$\Delta\mu = \hat{\mu} - \mu, \qquad \Delta\sigma = \hat{\sigma} - \sigma.$$

Third, we quantify fit to BSR bracket totals using the sum of squared deviations and a normalised RMSE across brackets, complemented by the maximum absolute residual to highlight worst-case deviations. Because dependence is estimated upstream of calibration, we compare the positive-block correlation R^+ to the truth using the Frobenius norm and the maximum off-diagonal error; for designs employing a t-copula we fix the degrees of freedom at $\nu = 4$. For the zero block, we report the bracket-conditional participation rates $\{\pi_k(j)\}$ and summarise the zero-block correlation structure via the implied correlation matrix R^0, defined analogously to R^+. In the Base scenario with $R = 15$ runs, we begin by establishing whether the tail test selects the correct family at each margin, as this sets the entry point for calibration. Table 5.3 reports margin-wise accuracy and confusion counts across runs.

Table 5.3. Base scenario (15 runs): tail-test model choice – accuracy and confusion by margin.

Margin	Accuracy (per cent)	Correct	Total	PLN-to-LN	LN-to-PLN
Margin 1	73.3	11	15	0	4
Margin 2	66.7	10	15	5	0
Margin 3	46.7	7	15	0	8

Notes: Accuracy is Correct/Total. "PLN-to-LN" and "LN-to-PLN" count misclassifications by direction. Tail tests are run margin-wise under the Base scenario with $R = 15$ replications.

The selection procedure identifies the true family with decent probability, except for the third margin. This is consistent with the design choice to impose stronger

co-movement between two components in the Base scenario, which reduces the effective marginal information available for discrimination. In the real-data application, however, model choice is less central because all three security components are, by construction, strongly concentrated in the upper tail and are therefore expected to be consistent with PLN behaviour. We then summarise marginal parameter recovery. Figure 5.1 displays run-wise errors for $\Delta\mu$ and $\Delta\sigma$, and, when PLN is selected, for $\Delta\alpha$ only. Body parameters are recovered precisely, with medians near zero and tight

Figure 5.1. Base scenario (15 runs): parameter errors $(\Delta\mu, \Delta\sigma, \Delta\alpha)$ by margin.

interquartile ranges. For $\Delta\mu$ (median value) we obtain 0.037 for M1, 0.046 for M2, and 0.019 for M3. For $\Delta\sigma$, the median equals -0.014 in M1, -0.049 in M2, and 0.005 in M3. Where PLN is selected, tail parameters remain centred but are noisier: $\Delta\alpha$ median in M2 is -0.006, with $\mathrm{RMSE}(\alpha) = 0.129$.

Focusing on the Base design (S 01, $R = 15$), only Margin 2 (M2) carries a PLN tail with a finite threshold $\tau_0 = 200{,}000$. Since τ is profiled on a discrete grid, we summarise identification via hit-rate metrics rather than $\Delta\tau$ itself, following Section 3.5.1. Table 5.4 reports the MAE to the snapped truth and the hit rates within 0 per cent (snapped), 5 per cent, and 10 per cent of τ_0 for M2. The snapped hit rate is 53.3%; both the 5% and 10% bands are also 53.3%, showing tau is still the most difficult value to estimate; the MAE (snapped) is 60.000.

Table 5.4. PLN threshold identification: MAE (snapped), hit rate (snapped), and hit rates within 5 per cent and 10 per cent of τ_0.

Scenario	Margin	MAE (snapped)	Hit rate (per cent)		
			Snapped	*Within 5%*	*Within 10%*
S 01	Margin 2	60.000	53.3	53.3	53.3

Notes: The snapped truth is the grid point closest to τ_0 within the candidate set used in estimation. "Hit rate within 5%/10%" is the share of runs with $|\hat{\tau} - \tau_0| \leq 0.05\,|\tau_0|$ or $\leq 0.10\,|\tau_0|$. Rows include only margins with finite τ_0 (PLN truth).

Calibration quality against BSR brackets is evaluated next. Let S^{obs} denote the pseudo-BSR bracket shares implied by the data-generating process (the target moments), and let S^{fit} denote the corresponding bracket shares implied by the estimated model after calibration. Figure 5.2 reports residuals $S^{\text{fit}} - S^{\text{obs}}$ by instrument and bracket, complemented in the text by scalar metrics (SSD and normalised RMSE).

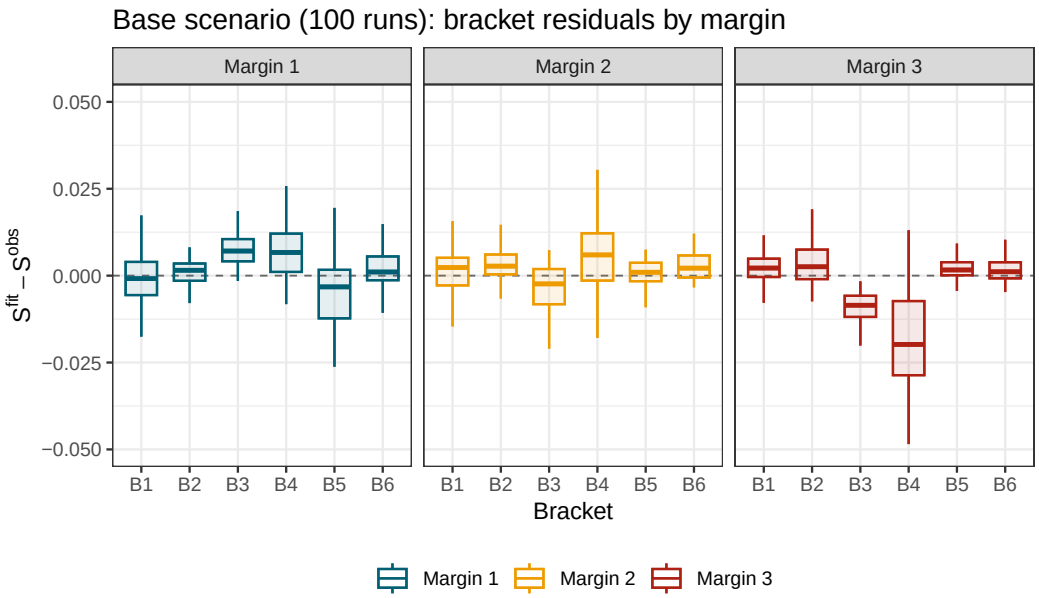

Figure 5.2. Base scenario (100 runs): bracket residuals $S^{\text{fit}} - S^{\text{obs}}$ by instrument and bracket.

Residuals are generally moderate, but they are not uniformly centred at zero across brackets and their dispersion varies systematically across the partition. In particular, the bracket carrying the largest share mass (B4 in this partition) exhibits the widest absolute variability. This is mechanically consistent with a levels-based loss: for a given relative discrepancy, a larger target share implies a larger absolute residual. In addition, B4 lies around the upper-body/lower-tail transition, so small run-to-run shifts in tail profiling and in positive-block dependence tend to reallocate mass across adjacent cut-offs, amplifying residual variability in that bracket even

when overall fit remains controlled. The corresponding normalised RMSE has median 12.7% (IQR 9.1%).

Because dependence is estimated upstream, we compare the positive-block correlation R^+ to the truth using the Frobenius norm and the maximum off-diagonal error. Figure 5.3 reports the Frobenius norm across runs. Figure A.4 in Appendix A complements it with the maximum off-diagonal error across runs.

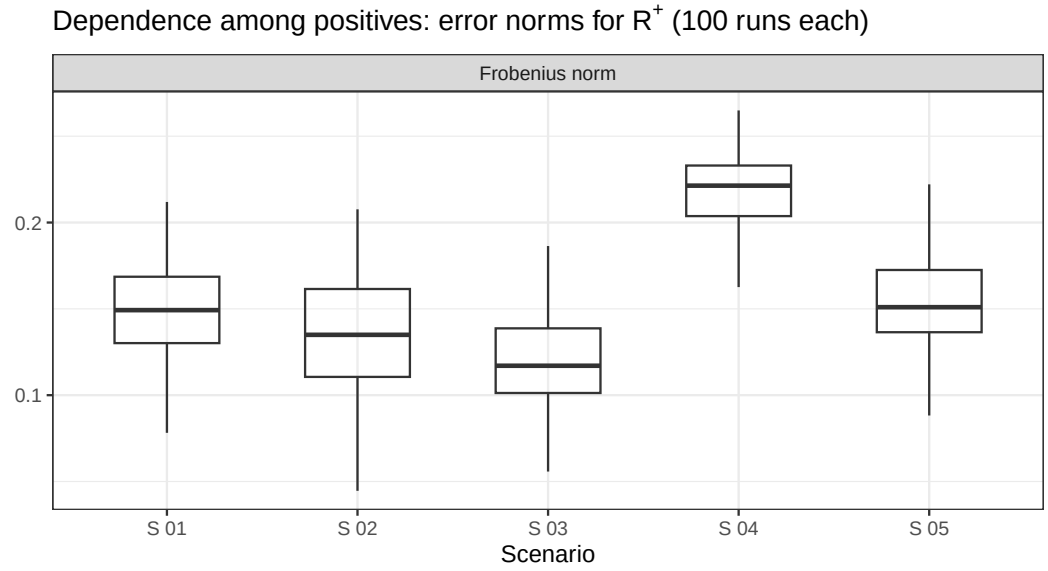

Figure 5.3. Five-scenario panel (100 runs each): dependence among positives – $\|R^+_{\text{hat}} - R^+_{\text{true}}\|_F$ across runs.

Positive-block dependence is recovered closely: concerning scenario S 01, the median Frobenius gap is 0.15 (IQR 0.04), while the median maximum off-diagonal error is 0.07 (IQR 0.02).

To link parameter and dependence recovery to distributional fit, Figure 5.4 overlays the model-implied CAF on the observed CAF by instrument, displaying the run-wise median with 95% coverage ribbons. Agreement is close through the body, while deviations concentrate in the very top range where identification is weakest.

Turning to sensitivity, we contrast five scenarios (S 01–S 05). Marginal parameter errors by scenario and margin are shown in Figure 5.5; this isolates the role of zeros, tail thickness, and dependence for recovery.

Patterns are consistent with identification geometry: heavier zeros and thicker tails enlarge the dispersion of PLN tail parameters and increase top-bracket sensitivity, while μ and σ remain comparatively stable. Across scenarios and margins, the median absolute error has a 95% central interval $|\Delta\mu| \in [-2.75, 1.2]$ and $|\Delta\sigma| \in [-0.39, 2.6]$; where PLN applies, the median $|\Delta\alpha|$ has a 95% central interval

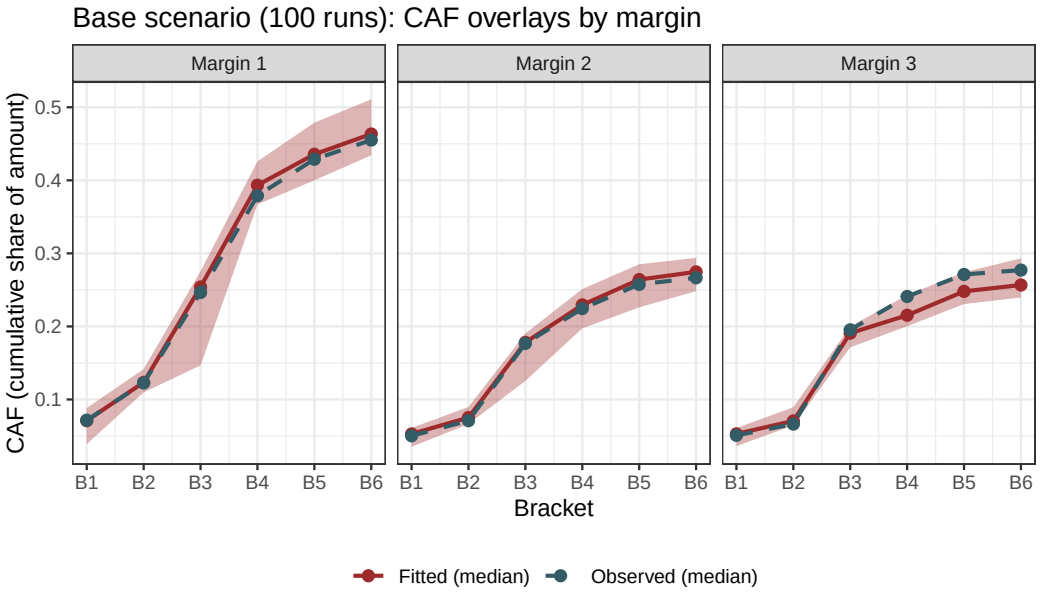

Figure 5.4. Base scenario (100 runs): CAF overlays (truth vs fitted) by instrument; median across runs with 95% coverage ribbons.

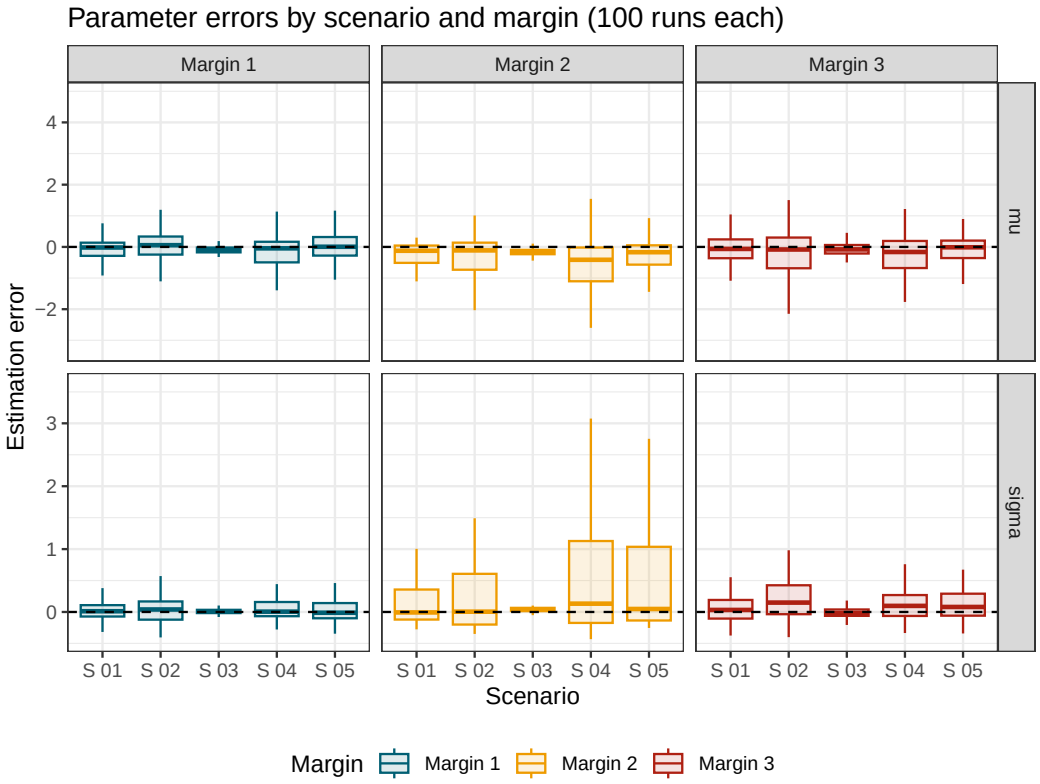

Figure 5.5. Five-scenario panel (100 runs each): parameter errors by scenario and margin.

of [−0.26, 4.09], with the wider end reflecting tails in heavy-zero designs.

Turning to the zero-block correlation, we recover the bracket-conditional participation rates $\{\pi_k(j)\}$ and the zero-block correlation R^0, and compare each object to its scenario truth. Figure 5.6 reports the maximum off-diagonal error for R^0, while Figure A.5 in Appendix A reports the L_2 norm and the Frobenius norm for R^0 and for $\{\pi_k(j)\}$. Degrees of freedom in the t-copula are fixed at $\nu = 4$ in S 04.

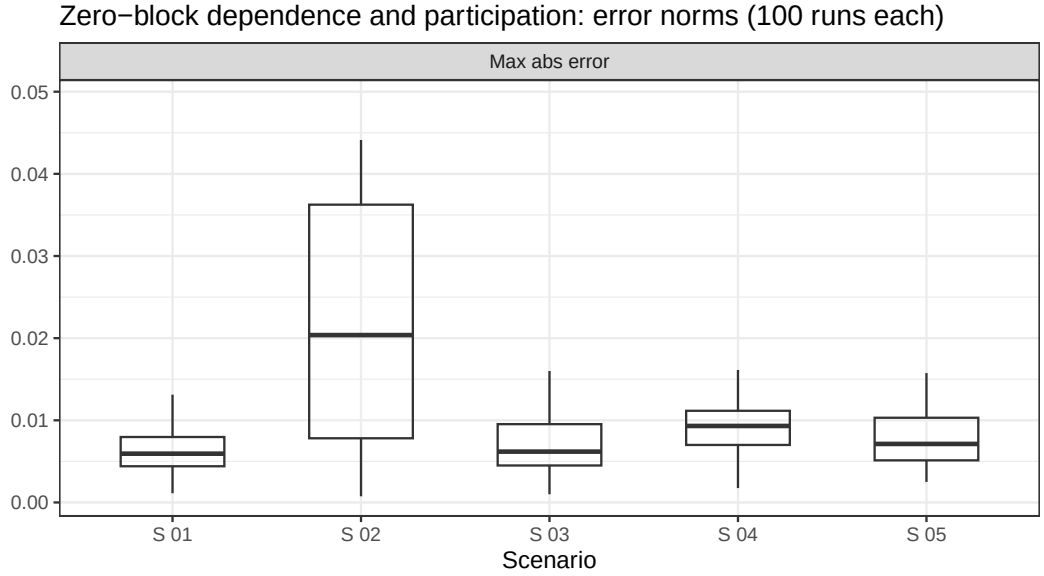

Figure 5.6. Five-scenario panel (100 runs each): zero-block dependence and participation – maximum off-diagonal error for R^0.

Allowing stronger positive dependence with a t-copula improves R^+ recovery relative to Gaussian, while under strong zero-block dependence and HFCS stress (S 05) the main errors concentrate in $\{\pi_k(j)\}$ and R^0; bracket fit for totals remains comparatively robust, consistent with the limited information on extensive-margin co-movements conveyed by range sums.

Finally, we assess sensitivity at the top via the absolute error in the highest bracket share by instrument (Figure 5.7). The top-bracket diagnostic concentrates power where identification is most fragile and links directly to policy-relevant tail mass. Top-bracket median absolute errors by scenario and Margin are reported in Appendix B, Table B.9. Top-bracket errors increase with tail thickness and participation sparsity, but remain contained in Near-LN designs; dependence-sensitive scenarios confirm that R^+ specification matters for tail alignment, whereas zero-block features primarily affect extensive-margin fit. In particular, S 04 exhibits the widest range (up to about 0.73%), consistent with heavier joint tails shifting mass in the top bin.

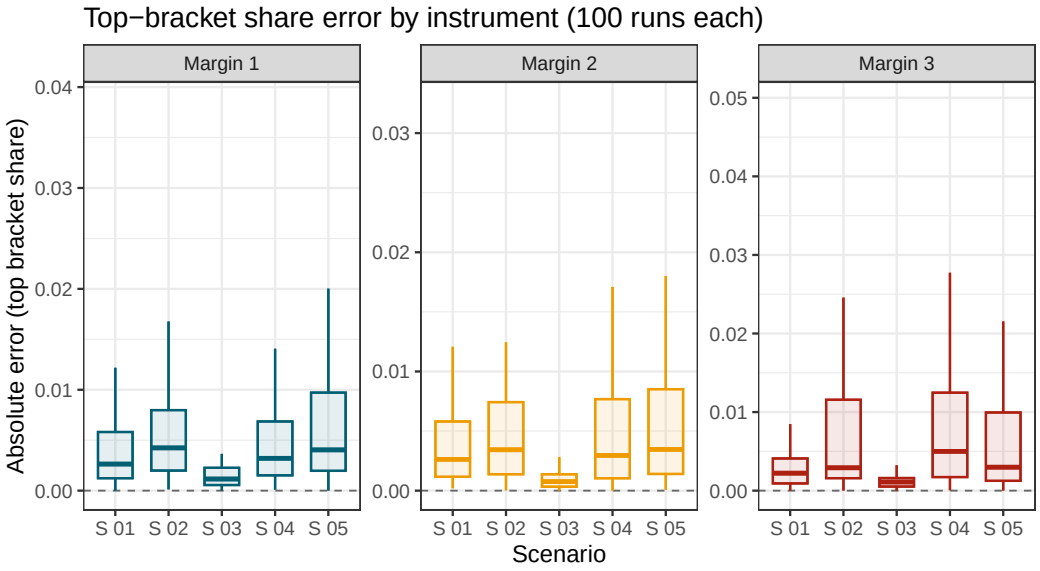

Figure 5.7. Five-scenario panel (3 runs each): top-bracket share error by instrument.

5.6.2 Real-data application

We now apply the estimator to BSR/HFCS data under the same reporting perimeter as in the simulations. All years are binned on the post-2022 break vector to maintain comparability across time. The three margins correspond to the actual instruments observed in BSR securities: bonds, listed shares, and investment fund shares. Estimation proceeds as before: margins are fitted with LN/PLN, participation is handled via the bracket-conditional zero block, and dependence is separated into a zero-block structure and a positive-block structure. The mean among holders is included as a soft upper bound via the asymmetric penalty calibrated in Section 5.5.10; we report whether, and how strongly, the anchor binds year by year.

We begin with a compact per-year snapshot that summarises marginal families and parameters, dependence, and participation. Table 5.5 reports, for each year and margin, $\widehat{\mu}, \widehat{\sigma}$, $\widehat{\alpha}$, and $\widehat{\tau}$ (since PLN is selected on each margin), along with the implied model mean among holders $\widehat{\mu}^H$ and the log anchor gap $\Delta_\mu = \log(\widehat{\mu}^H / \mu^H_{\text{anchor}})$.

Table 5.5. Real-data overview by year and margin: marginal parameters, model-implied mean vs anchor with augmented estimation.

Year	Instrument	Model	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\alpha}$	$\widehat{\tau}$	$\widehat{\mu}^H$	gap Δ_μ
2010	Bonds	PLN	11.13	0.98	1.05	1,584,893	184,851	-0.04
2010	Listed shares	PLN	7.26	2.35	1.05	1,584,893	78,403	0.07
2010	Investment funds	PLN	9.18	1.88	1.05	1,584,893	216,057	0.08
2014	Bonds	PLN	11.62	0.90	3.62	1,584,893	166,617	0.03
2014	Listed shares	PLN	9.91	1.99	3.62	1,584,893	95,587	0.00
2014	Investment funds	PLN	12.13	0.90	3.62	1,584,893	274,893	-0.13
2016	Bonds	PLN	10.09	1.32	1.05	1,459,673	124,829	0.01
2016	Listed shares	PLN	5.40	3.18	1.05	1,459,673	92,414	0.06
2016	Investment funds	PLN	11.40	0.96	1.05	1,459,673	319,506	0.00
2020	Bonds	PLN	9.22	1.63	1.05	1,583,683	98,165	0.03
2020	Listed shares	PLN	6.40	2.67	1.05	1,583,683	74,171	0.03
2020	Investment funds	PLN	11.13	1.12	1.05	1,583,683	345,920	0.07
2022	Bonds	PLN	10.37	1.36	1.50	316,228	94,238	0.03
2022	Listed shares	PLN	10.15	2.01	1.50	316,228	102,739	0.03
2022	Investment funds	PLN	11.70	0.90	1.50	316,228	256,154	0.02

For comparison, Table B.10 in Appendix B reports the estimates obtained with the standard method without the anchor. Results clearly show why the anchor is important in this estimation. With comparatively similar fits, the standard model yields very erratic mean values, which in particular years are also hardly credible. Providing an anchor as a soft penalisation – given that it is a sensible upper bound and becomes more credible as survey quality improves over time – allows estimates to depict more precisely the latent truth.

To convey correlation matrices without overwhelming the reader, Figure 5.8 shows year-by-year upper-triangle heatmaps for R^+ and R^0. This format preserves exact values while making temporal shifts legible. Both matrices display consistently positive entries across years; with few exceptions, patterns are quite stable over time.

We then assess calibration against the BSR targets. Figure 5.9 plots the residuals $S^{\text{fit}} - S^{\text{obs}}$ by instrument and bracket for each year; SSD and normalised RMSE are reported in the caption. Residuals remain small throughout, with discrepancies concentrated in the open-top bracket, consistent with the information content of range sums.

To describe distributional shape among holders, Figures A.6 overlay model-implied densities by year for each instrument on a log scale with peaks rescaled to equal one in order to improve readability. In Appendix A, Figure 5.10 reports the same unnormalised figure for completeness.

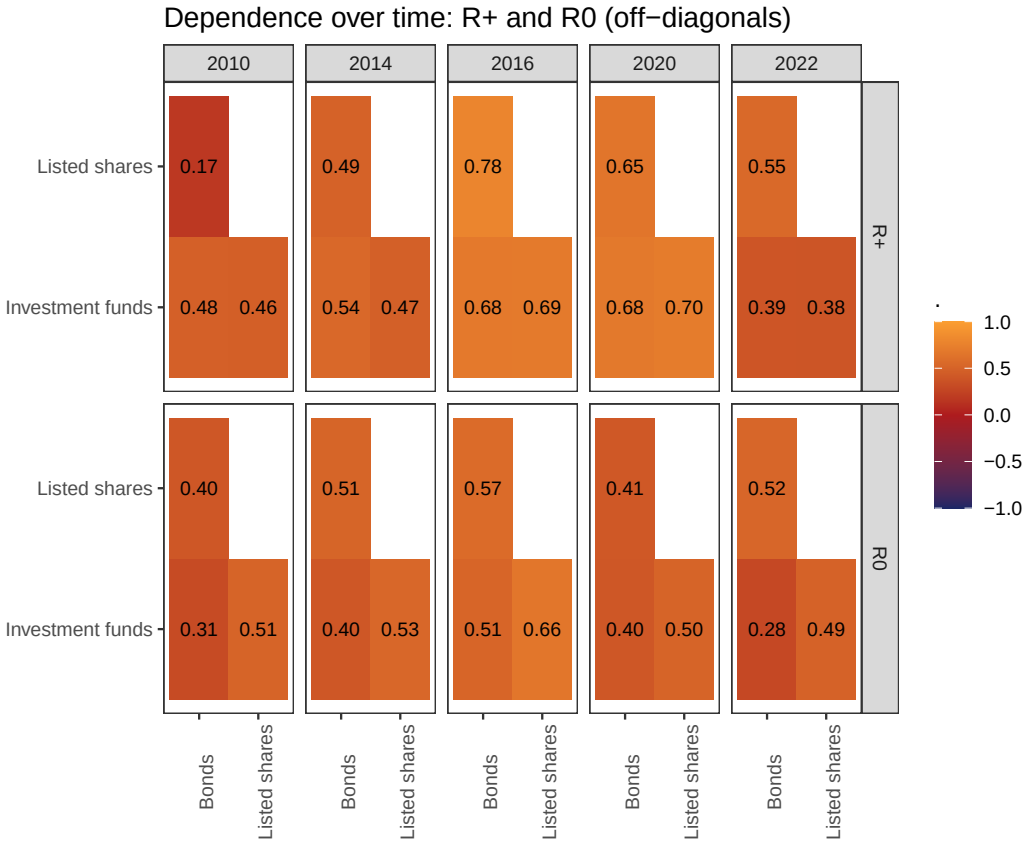

Figure 5.8. Instrument cross-dependence over time: R^+ and R^0 (upper-triangle heatmaps by year) with Frobenius drift $\|R_t - R_{t-1}\|_F$.

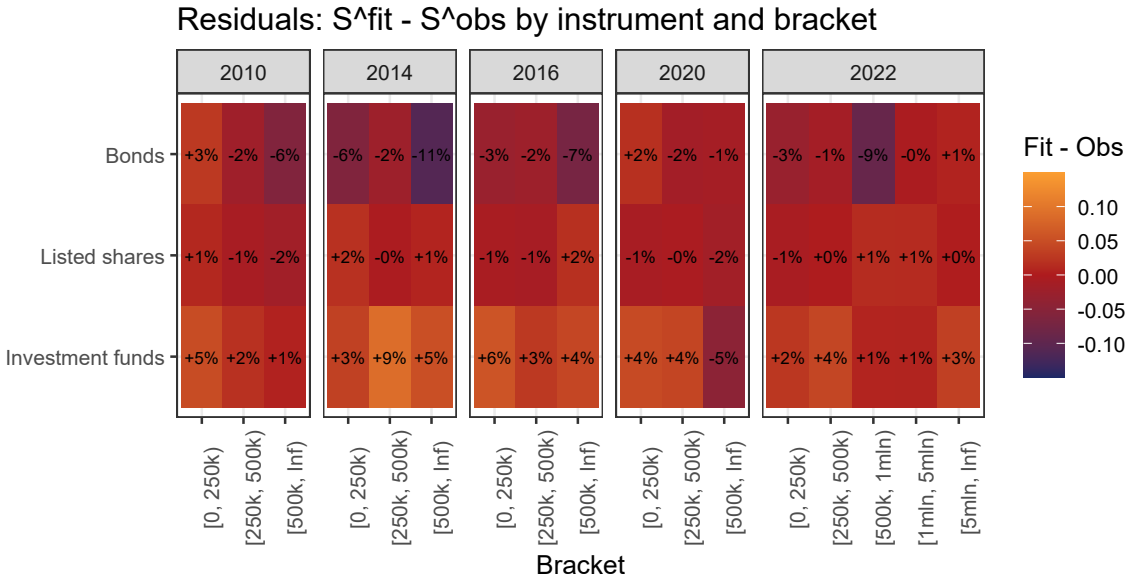

Figure 5.9. Bracket residuals $S^{\text{fit}} - S^{\text{obs}}$ by instrument and year (post-2022 breaks).

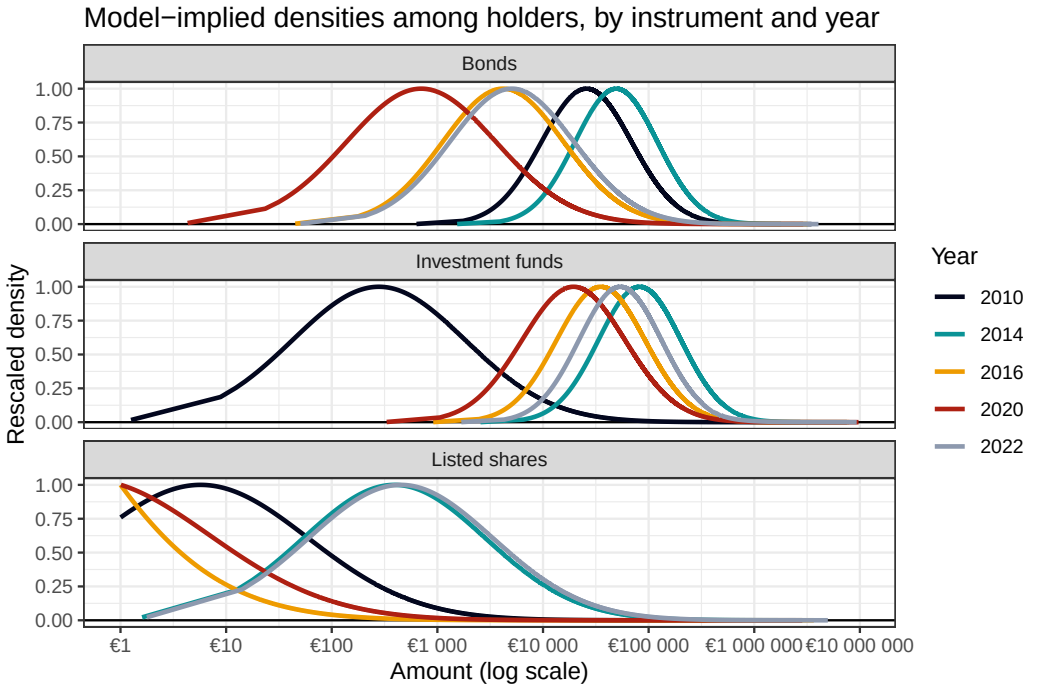

Figure 5.10. Yearly densities (holders only) by instrument; log scale on the x-axis.

Anchor coherence is reported next. Figure 5.11 shows the ratio $\widehat{\mu}^H/\mu^H_{\text{anchor}}$ by instrument and year. Values near or below 1.00 indicate the anchor is either non-binding or binding weakly, whereas sustained ratios above the tolerance would suggest upward pressure from the likelihood. The anchor appears to bind in several years, with ratios gravitating near 1.00. This highlights the need for more informative bracket data: bracket shares are recovered steadily even when the mean binds, but finer breaks – especially at the top – would sharpen identification and downstream analysis.

To link parameters and dependence to the calibration target, Figure 5.12 overlays the model-implied CAF on the observed CAF by instrument and year (median values). Agreement is close through the body; deviations, when present, concentrate at the very top where identification is weakest and dependence choices matter most.

Observed CAF curves exceed fitted CAFs for listed shares in almost all years. A pattern common to all instruments is the improvement in the latest survey wave. This reflects the combined effect of a revised sampling scheme – which improved information for these instruments – and the addition of BSR breaks. Further refinement of both lower and upper breaks would likely yield additional gains in household-level fit. In general, these administrative data offer a practical, accessible route to correct several biases that affect household surveys.

To conclude, this method delivers promising results. Simulations show that, although model selection can improve, the underlying truth is recovered reliably.

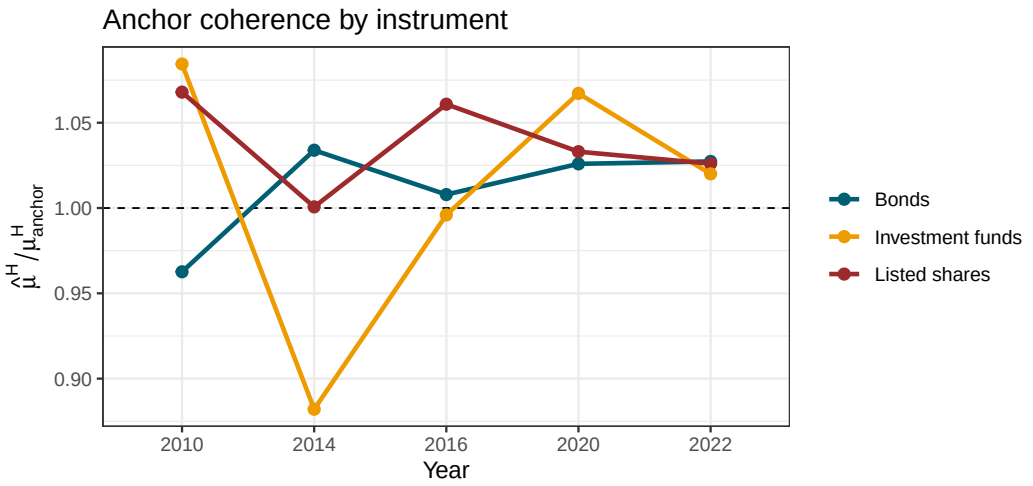

Figure 5.11. Anchor coherence: model-implied mean among holders vs anchor ($\widehat{\mu}^H/\mu^H_{\text{anchor}}$) by year.

Real-data results highlight the crucial role of the number of brackets and illustrate how a mean anchor can improve estimates; in low-information settings, however, the anchor may bind strongly. Nevertheless, bracket fit remains stable and comparable to the non-anchored method. These results complement the prior BSR data work and yield fully comparable BSR-based estimates on the HFCS perimeter. The following chapters will exploit these data to adjust HFCS microdata, providing calibration methodologies that correct household values while accounting for BSR administrative information.

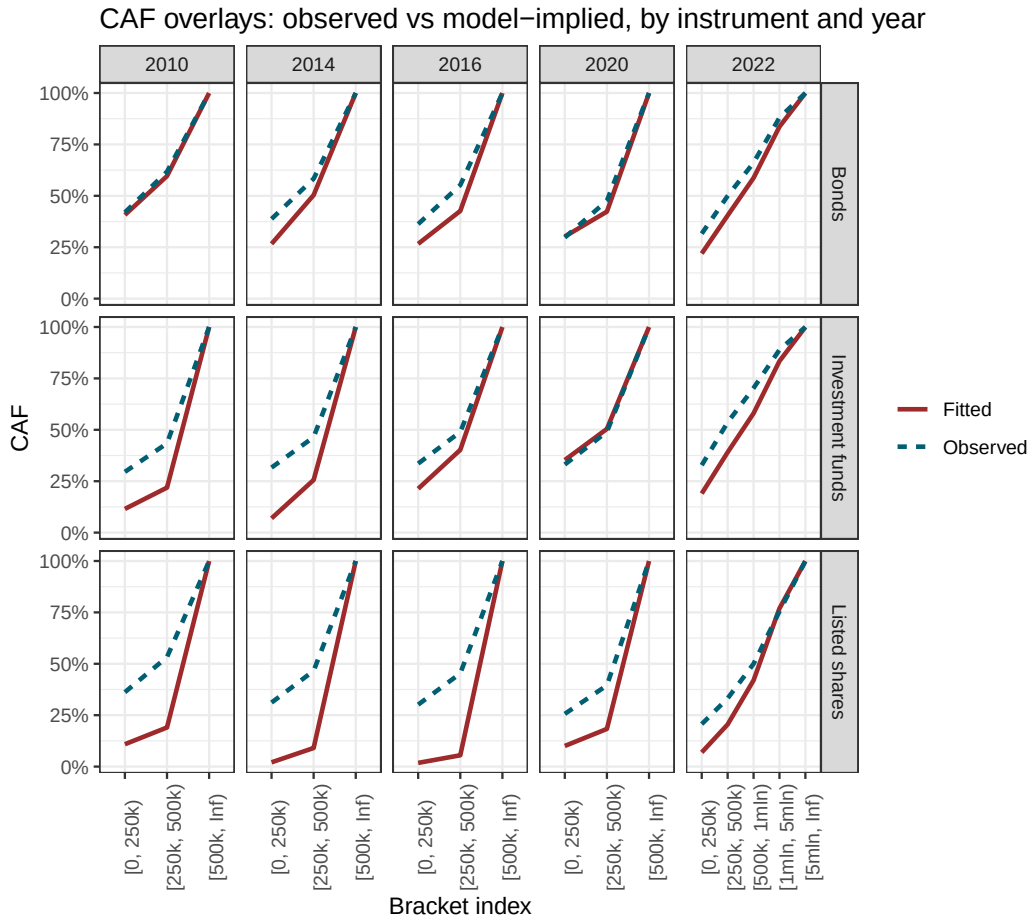

Figure 5.12. CAF overlays (observed vs model-implied) by instrument and year; median values.

Chapter 6

Value-based calibration of HFCS survey-data on wealth components

6.1 Introduction

This chapter proposes a calibration strategy to incorporate Banking Supervisory Reports (BSR) information into the Household Finance and Consumption Survey (HFCS) microdata. The objective is to align instrument values at the household level with external bracket totals, while preserving the survey's weighting structure and the distributional information reconstructed in earlier chapters. In previous work (Neri et al., 2023), we implemented the calibration for deposits only and after splitting HFCS households into accounts, owing to the absence of a household-level bridge to BSR and the lack of a method to disentangle marginal distributions for securities accounts. Nevertheless, the impact of that procedure was sufficiently material that it was adopted in the standard Italian Distributional Wealth Accounts workflow and will therefore serve here as the benchmark.[1] By contrast with the earlier account-based implementation, the present chapter operates directly on household deposits and extends the calibration to securities.

The path to this chapter was incremental. The thesis first examined what bank supervisory range-sum aggregates do and do not reveal about household portfolios. For deposits, BSR data are account-based; given the marked difference between household and account levels, we reconstructed household-level BSR distributions by exploiting HFCS account-household links (Chapter 4). For securities, BSR deliver, by brackets of total securities, instrument totals that are size-biased first moments contributed by accounts with at least one positive holding. This structure excludes triple zeros and distorts co-positivity in ways that matter for inference. Section 5.2 used this fact to diagnose false zeros in HFCS and to characterise identification limits when co-positives are sparse in high-value brackets; Chapter 5 then disentangled

[1] For scope, compilation principles and production constraints, see European Central Bank (2024), *Distributional Wealth Accounts: Methodological Note*, available at https://data.ecb.europa.eu/sites/default/files/2024-01/DWA%20Methodological%20note_0.pdf.

instrument contributions from range sums, clarifying how bracket totals relate to instrument marginals under truncation and what that implies for calibration targets. These steps established the administrative-data backbone for reconciliation with the survey.

The chapter develops three calibration models that build on each other. The first targets bracket totals directly through bounded value adjustment and explicit control of class changes. The second retains value adjustment and fixed weights, but replaces hard bracket constraints with a composite distributional loss that aligns both the marginal shape and the cumulative contribution of amounts, and is implemented univariately by instrument. The third extends the second to a multivariate setting by adding a dependence layer that controls cross-instrument co-movement while preserving what has been achieved in the marginals.

6.1.1 Calibration with class-change penalisation

The scope of the first calibration expands along two dimensions. First, deposits are calibrated directly at the household level by leveraging the account-to-household reconstruction in Chapter 4 and the associated main-account shares. Second, the same logic is extended to securities-account instruments – listed shares, bonds, and investment fund shares – using the instrument-specific marginal breaks recovered in Chapter 5. These breaks are estimated from BSR range-sum data via the parametric distributions validated in Chapter 5 and yield instrument-level bracket totals for each year once combined with the corresponding aggregates. In practice, for each instrument, theoretical bracket shares are computed as size-biased truncated first moments divided by the total; multiplying these shares by the aggregate restores the BSR object used as the target in calibration.

A key design choice is to leave household sampling weights unchanged. Weight calibration would propagate to all survey domains; instead, we adopt a deliberately univariate-by-instrument approach that alters only the values of the instrument under calibration. This choice is consistent with recent applications that compute household-specific calibration factors and apply them to the variables of interest rather than to the weights, thereby limiting spillovers to other outputs (Cantarella et al., 2021). Concretely, each observed value is multiplied by a bounded adjustment factor. This design introduces a specific challenge: because BSR constraints are expressed as totals by value ranges ("classes"), adjusting values can push households across class boundaries. Consequently, bracket totals may be satisfied under the pre-adjustment classes yet missed once classes are recomputed on adjusted values.

Two safeguards are therefore embedded. First, adjustment factors are constrained within instrument-specific bounds to prevent excessive departures from the initial distribution; baseline bounds are fixed at $[\underline{\varphi}_k, \overline{\varphi}_k] = [0.5, 10]$ for each instrument k. Second, calibration is implemented with an iterative, class-aware stabilisation step that limits reclassification near bracket thresholds while preserving the ability to improve global fit; the full procedure is described in Section 6.3.1.

6.1.2 Calibration with composite distributional loss

The second calibration builds on three premises established earlier in the thesis: first, that BSR tabulations, once harmonised with HFCS definitions and coverage, contain informative signals about the cross-sectional shape of household portfolios at the instrument level; second, that suitable parametric forms can parsimoniously represent these shapes for calibration; and third, that participation must be handled explicitly so that the relevant subpopulation for each instrument is well identified before any value adjustment. In contrast to the class-aware calibration in Section 6.3.2, which prioritises aggregate alignment through hard bracket constraints, this approach prioritises distributional coherence and recovers aggregates indirectly through a cumulative-amount alignment.

The central device is a composite loss that balances a divergence on the marginal shape with a divergence on the cumulative contribution of amounts. The former aligns the entire empirical CDF of adjusted values with the BSR-derived target; the latter aligns the relative cumulative amount function, so that totals and bracket-level aggregates emerge as a consequence of distributional fit rather than as hard constraints. This design is motivated by two limitations of bracket-based calibration for skewed, heavy-tailed instruments: adjusted observations may accumulate at bracket thresholds, and the procedure becomes difficult to scale to a genuinely multivariate setting without proliferating endogenous class reassignments across instruments.

This calibration contributes to a broader line of work that uses external information to improve the distributional properties of survey microdata. Classical survey calibration adjusts weights to satisfy auxiliary constraints under convex distance measures; Deville and Särndal (1992) show how chi-square and related distances deliver consistent estimators while preserving design-based properties. Here, calibration cannot act on weights and must act on values directly, which breaks convexity and requires smooth surrogates for the indicator functions used to compare distributions. The marginal shape divergence is based on the Cramér–von Mises distance, evaluated on a quantile grid; this choice is motivated by its compatibility with sampling weights and its sensitivity across the support, including the tails. For alternative metrics that emphasise geometric transport, Peyré et al. (2019) survey Wasserstein distances and discuss their computational trade-offs.

In practical terms, calibration acts on values rather than on weights. Let x_h denote the reported amount for household h for a given instrument and w_h the fixed survey weight; we introduce multiplicative adjustments $\varphi_h > 0$ and define $y_h = \varphi_h x_h$. From BSR we recover a target marginal distribution F (with density f) for the instrument and the relative cumulative amount function

$$s(y) \;=\; \frac{\int_0^y x f(x)\,\mathrm{d}x}{\int_0^\infty x f(x)\,\mathrm{d}x}\,,$$

which is the share of the total amount contributed by units with value at most y.

The corresponding instrument total is $T = N\,\mathbb{E}(X)$, with N the population size for the instrument; Chapters 3, 4 and 5 provide the parametric backbone for F and address account-to-household aggregation issues and the disentanglement of deposit and securities accounts. Participation is addressed upstream so that calibration targets the appropriate subpopulation; operationally, units with $x_h = 0$ are kept at $\varphi_h = 1$ and excluded from the margin-specific divergences, and where false-zeros are identified, their recovery precedes calibration.

The divergence on cumulative amounts uses a smooth approximation to the cumulative amount function (CAF) of the adjusted sample and matches it to the BSR-derived target in relative terms. The latter is a size-biased CDF that arises from weighted distributions; Patil and Rao (1978) show that size-biasing by x reweights the population so that cumulative amounts can be studied through a proper distributional object. This relative formulation stabilises the scale of the problem and embeds the total at the upper tail by construction; when the grid includes bracket thresholds, it also promotes coherence of bracket totals without imposing them as hard constraints.

To fix ideas, the working loss is

$$L(\boldsymbol{\varphi}) = L_{\text{shape}}(\boldsymbol{\varphi}) + \lambda_{\text{caf}}\, L_{\text{caf}}(\boldsymbol{\varphi}) + \lambda_{\text{stab}}\, L_{\text{stab}}(\boldsymbol{\varphi}),$$

where L_{shape} implements the smoothed Cramér–von Mises discrepancy on the log-scale, L_{caf} compares the adjusted relative CAF to $s(y)$ on a fixed grid that includes a far-tail anchor and L_{stab} stabilises the adjustment weights penalising for the overall changes as in the first model. Because value calibration induces a non-convex optimisation problem, the implementation follows a two-stage scheme (fit, then stabilise) under box constraints; Section 6.4 reports the algorithm, tuning choices, and stopping rules.

6.1.3 Multivariate composite calibration

Finally, the chapter derives a single multivariate calibration that adjusts HFCS microdata to be simultaneously consistent with instrument-level totals and with an interpretable, regime-specific dependence structure across instruments. The objective is practical: to provide a coherent, survey-based micro dataset that aligns with Distributional Wealth Accounts (DWA) targets constructed from administrative and national accounts sources, while preserving portfolio features that matter for downstream analysis and dissemination.

What remains unresolved after a purely marginal calibration is the joint behaviour of instruments. Different configurations of unit-level adjustments can map to the same totals and similar univariate shapes yet imply very different co-movements. In practice, dependence between instruments varies across the distribution of holdings: small portfolios often reflect transactions and precautionary balances, whereas large positions reflect deliberate risk choices. If dependence is left unconstrained, adjustments intended to repair totals and tails can inadvertently distort cross-instrument structure. The multivariate layer addresses this by aligning the survey's implied

dependence with regime-specific targets while preserving what has been achieved in the marginals.

The dependence layer operates on a scale that removes marginal effects by construction. For each instrument, the adjusted value is mapped through the parametric probability integral transform (PIT) associated with its calibrated PLN margin and then converted to a Gaussian score. Dependence is targeted through regime-specific correlations of these Gaussian scores over common value segments. In the baseline, segments are $[0, 250\text{k})$ (low) and $[250\text{k}, \infty)$ (high); regime membership is implemented with smooth log-scale windows to preserve differentiability, and PITs are clipped at the extremes for numerical stability.

Targets are defined once per year on the Gaussian-score scale. For each segment $b \in \{\text{low, high}\}$, the target correlation matrix $R_b^\star$ is estimated on co-positives using survey weights and projected to the nearest positive definite form with minimal adjustment; when segment support is thin, targets may be shrunk towards a global correlation in a diagnostic-driven and parsimonious way. During calibration, the same mapping is applied to the adjusted microdata and the implied segment correlations are compared to their targets through a Frobenius gap appended to the Stage 1 objective:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\varphi}) \;=\; \sum_{k=1}^{K} \Big\{ L_{\text{shape},k}(\boldsymbol{\varphi}) \;+\; \lambda_{\text{caf}}\, L_{\text{caf},k}(\boldsymbol{\varphi}) + \lambda_{\text{stab}}\, L_{\text{stab}}(\boldsymbol{\varphi}) \Big\} + \\ + \lambda_{\text{dep}} \sum_{b \in \{\text{low, high}\}} \big\| R_b(z(\boldsymbol{\varphi})) \;-\; R_b^\star \big\|_F^2 \,.\end{aligned} \tag{6.1}$$

The tuning λ_{dep} is introduced conservatively and increased until dependence improves without harming the marginal fit or the near-alignment of totals.

In practice, this chapter refrains from simulated exercises. Because the calibration targets are tied to the supervisory range design and the main trade-offs concern exact total matching versus distributional coherence, simulated data add limited value beyond basic solver validation. All calibrations are therefore performed on the real HFCS microdata after the household-level reconstruction of Chapter 5, with instrument-specific targets derived from the estimated BSR-consistent distributions. By contrast with the earlier account-based deposit calibration, re-estimating that specification would mainly reflect the inclusion of an additional wave and minor weight revisions, without producing new substantive insights; it is thus used solely as a benchmark for diagnostics.

Specifically, Section 6.2 reviews the relevant literature. Section 6.3 distils the theoretical implications of this literature concerning all three models developed. Section 6.4 details the implementation, including the algorithm, tuning choices, and stopping rules. Section 6.5 presents the empirical results and compares them across the three models.

6.2 Literature review

This section reviews the strands of literature that inform the three methodologies developed in this chapter. It proceeds in blocks. It first recalls calibration in official statistics and the production safeguards that motivate bounded updates. It then focuses on range-sum constraints and value-class totals, clarifying why multiplicative adjustment of instrument values – rather than weight calibration – is natural when constraints are truncated first-moment totals by bracket, as in BSR. It next situates the distribution-aware calibration of Section 6.1.2 within minimum-distance alignment of weighted distributions, emphasising smooth distributional distances, survey-weighted empirical processes, and cumulative-amount targets. Finally, it positions the multivariate extension of Section 6.1.3 within the literature on modular dependence modelling and feasible, survey-consistent correlation targets.

6.2.1 Calibration in official statistics: principles and practice

Calibration estimators adjust survey quantities so that selected auxiliary totals are matched exactly while perturbations remain controlled by a chosen distance. The formulation of Deville and Särndal (1992) unifies these ideas by minimising a convex distance from base weights subject to linear constraints, yielding design-consistent estimators with improved efficiency when auxiliary information is predictive. Deville et al. (1993) extend the approach to generalised raking, showing how the classical iterative proportional fitting of Deming and Stephan (1940) emerges as a special case. Complementary reviews such as Särndal (2007) and subsequent syntheses document the widespread adoption of calibration in production, its variance properties, and the practical safeguards needed to avoid undue distortions.

Two operational lessons are central here. First, calibration requires safeguards against extreme adjustments, particularly when micro and benchmark distributions differ materially. Early contributions (Potter, 1990) advocated trimming and bounding strategies, and modern production guidance confirms the routine use of truncated or logit updates and explicit bounds on adjustment factors or g-weights (Haziza and Beaumont, 2017). Second, when calibration serves not only variance reduction but also error correction (nonresponse, coverage), quasi-randomisation arguments justify a broad use of auxiliary information and instrumental-variable extensions (Kott, 2006; Chang and Kott, 2008).

Software infrastructures encode these practices. INSEE's `CALMAR` macros operationalise generalised calibration with options for raking and bounds; ISTAT's `ReGenesees` provides a design-based and model-assisted platform for calibration and post-calibration analysis; and the `sampling` package in R (`gencalib`) implements linear, raking, truncated and logit updates with user-specified bounds and iteration controls. These tools reflect a production consensus: bounds and diagnostics are intrinsic to calibration workflows rather than ad hoc additions.

The modern calibration theory in complex surveys links design-based estimators

to auxiliary information by minimising a discrepancy subject to benchmarking equations. As Deville and Särndal (1992) show, adjusting sampling weights to satisfy auxiliary totals under convex distances yields generalised regression estimators, encompasses raking and post-stratification, and preserves desirable design properties. Subsequent developments broaden this paradigm to incorporate nonresponse and coverage error under plausible models while retaining convexity and algorithmic stability (e.g. Wu and Sitter, 2001; Haziza and Beaumont, 2017). In practice, convex calibration is attractive because it guarantees a unique optimum and scales well to large frames.

In the present chapter, weights remain fixed; consequently, the reconciliation lever shifts to reported values. The implications for value adjustment and the resulting non-convexity are discussed in Section 6.2.6.

The methodological implication for the present chapter is that non-convexity and possible multi-modality are to be expected, so regularisation and well-chosen starting values play a central role. The use of smooth indicators, continuation in bandwidths, and the option of weak shape constraints are standard tools to control these features without reverting to hard class constraints. Related reconciliation frameworks similarly adjust series to respect external benchmarks while controlling distortion, reinforcing the value-level perspective adopted here (Dagum and Cholette, 2006).

6.2.2 Penalised and 'soft' calibration

Penalised calibration modifies pure distance minimisation under exact constraints by introducing penalties that either relax some margins or regularise updates. This literature arises along three, partly overlapping, lines.

Penalised margins and mixed-model assistance. Guggemos and Tillé (2010) propose replacing some hard calibration constraints with penalties in the criterion, obtaining a design-based estimator assisted by mixed models; the penalty weights directly control how tightly a given margin is enforced. Related partial penalisation schemes allow a subset of margins to hold exactly, while others are satisfied approximately, with tuning trading off distance to base weights against lack of fit on softer margins (Goga and Shehzad, 2014).

Ridge and model-calibration regularisation. Ridge-type stabilisation is employed in calibration to mitigate multicollinearity or ill-posedness when the auxiliary set is high-dimensional. Contributions in the model-calibration tradition (Wu and Sitter, 2001) provide a framework for embedding predictive structure into calibration without sacrificing design-based justification; principal-component and ridge variants can be used when the auxiliary space is large relative to sample size (e.g. (Cardot et al., 2017)).

Production implementations. Penalised variants have progressively entered production-oriented software (e.g. INSEE's `icarus`, which includes penalised calibration and tight-bound options).

Our class-change control is conceptually distinct from penalising margins or adding ridge regularisation. In the class-aware model of Section 6.1.1, margins remain hard: class-level range-sum totals T_{kj} must be attained. Stabilisation acts on a behavioural feature of the solution path – reclassification of units across brackets induced by value adjustments – via an ex-post, class-aware projection of multipliers for bracket-changers into a near-unity trust region; the baseline choice and its practical motivation are stated with the sequential scheme in Section 6.3.1. This preserves exact benchmarks while regularising iteration dynamics.

6.2.3 Range-sum constraints and value-based benchmarking analogues

BSR provide totals by value range ("brackets") of each instrument. Let $[a_{j-1}, a_j)$ be the j-th bracket and T_{kj} the corresponding aggregate total for instrument k. Such constraints are size-biased truncated first moments rather than simple frequencies or grouped means. Normalising by the instrument total produces bracket shares under the size-biased (length-biased) distribution of the instrument, not under the original density. This distinction is classical in the theory of weighted distributions (Patil and Rao, 1978; Vardi, 1982).

Two consequences follow. First, representing the constraints via class indicators and calibrating weights is, for the purpose of matching class totals, equivalent to multiplying values by adjustment factors and matching truncated sums; the latter representation is more direct for range-sum constraints because the target is linear in values. Second, because values are adjusted, units may cross class boundaries. Unconstrained L_2 updates can therefore generate large reclassification flows, particularly when micro and benchmark distributions disagree in the upper tail. In our setting, this is handled by an iterative class-aware stabilisation step in which reclassifiers are temporarily projected to a near-unity band before the next solve, while keeping the class totals T_{kj} as hard constraints.

Analogies with benchmarking and reconciliation are useful. In temporal benchmarking, a quadratic criterion is minimised subject to aggregation constraints, with solutions known to behave well in production (Denton, 1971; Dagum and Cholette, 2006; Buono et al., 2018). Our problem is cross-sectional rather than temporal, but it shares two properties: (*i*) linear constraints and (*ii*) a quadratic loss penalising deviation from a baseline. The key difference is that range-sum constraints depend on endogenous class assignment, which motivates explicit control of reclassification.

6.2.4 Positioning and novelty for the class-aware model

Section 6.1.1 implements a bounded, value-based L_2 calibration to BSR class totals with two safeguards: (*i*) global bounds on multiplicative adjustments $\varphi_{hk} \in [0.5, 10]$; (*ii*) a class-change control implemented as an ex-post projection – after each bounded fit, any unit that has reclassified has its multiplier projected to $[0.95, 1.05]$, and adjusted values are recomputed before the next iteration. This production-friendly device stabilises convergence and preserves local rank information, while allowing necessary reclassifications to satisfy feasibility. The device follows (Neri et al., 2023) and is here stated explicitly, formalised as a sequential bounded calibration with reclassification-triggered projection, and extended to additional instruments.

Relative to the literature, the novelty lies not in a new optimisation paradigm but in a targeted, class-change trust-region heuristic tailored to range-sum constraints and implemented at the level of instrument values. It differs from (*i*) penalised-margin approaches (which relax margins) and (*ii*) ridge/model-calibration regularisation (which operates in auxiliary space): here, margins remain exact and the control regulates who can move across classes by projecting reclassifiers to a near-unity band after each iteration, then restoring baseline bounds in the subsequent solve.

A caveat is the potential accumulation of adjusted observations near bracket cut-offs. Because range-sum targets are piecewise-constant in class membership, value-based adjustments under tight class-change control may produce visible mass near boundaries. This trades some micro-level smoothness for a quick, transparent reconciliation to external class totals. In a production context – where alignment with BSR objects and macro aggregates is the priority – this trade-off is acceptable and documented; the distribution-aware and multivariate models of Sections 6.1.2 and 6.1.3 explicitly target micro-smoothness and cross-instrument coherence and report diagnostics on boundary accumulation.

6.2.5 Summary of implications for class-aware calibration

Three implications guide the empirical implementation of the class-aware model:

(i) **Criterion and representation.** With range-sum constraints, an L_2 distance on multipliers/values provides a direct linear mapping from adjustments to class totals, avoids a proliferation of class dummies in a weight-calibration setup, and facilitates explicit bounds.

(ii) **Stability.** Global bounds on multipliers are essential to prevent unrealistic adjustments; an ex-post projection band for reclassifiers curbs oscillations and preserves extensive-margin patterns.

(iii) **Production stance.** The approach prioritises transparency and robustness over algorithmic sophistication. It is compatible with DWA requirements and mirrors established benchmarking philosophies, while acknowledging the boundary-accumulation drawback.

6.2.6 From weight-based calibration to distribution-aware adjustment

The modern calibration theory in complex surveys links design-based estimators to auxiliary information by minimising a discrepancy subject to benchmarking equations. As Deville and Särndal (1992) show, adjusting sampling weights to satisfy auxiliary totals under convex distances yields generalised regression estimators, encompasses raking and post-stratification, and preserves desirable design properties. Subsequent developments broaden this paradigm to incorporate nonresponse and coverage error under plausible models while retaining convexity and algorithmic stability (e.g. Wu and Sitter, 2001; Haziza and Beaumont, 2017). In practice, convex calibration is attractive because it guarantees a unique optimum and scales well to large frames.

In the present chapter, weights must remain fixed, so the adjustment lever shifts from weights to reported values. The conceptual link with the calibration literature is nevertheless direct: the objective remains a distance-minimisation problem, but the object of adjustment becomes the vector of amounts. The advantage is the ability to inject parametric information about the entire marginal distribution of an instrument into microdata without modifying the sampling design – the cost is non-convexity once distributional penalties depend nonlinearly on adjusted values. This motivates the use of smooth surrogates, analytic gradients, and continuation strategies detailed later in the chapter.

6.2.7 Distributional distances for alignment: CvM and alternatives

A natural way to steer adjusted values towards a target marginal distribution is to penalise the discrepancy between a design-weighted empirical cumulative distribution function and a reference cumulative distribution function. Cramér–von Mises statistics provide a classical, global measure of discrepancy that balances centre and tails, and their properties are well documented (Stephens, 1974; Anderson and Darling, 1954). In complex surveys, replacing the empirical distribution with its design-weighted counterpart delivers valid procedures and large-sample approximations (Wang, 2012), which makes CvM a pragmatic choice when weights must be honoured in the criterion itself. Beyond this conceptual fit, CvM is convenient computationally: a smoothed indicator on the log-scale transforms the step process into a differentiable map, and the resulting gradients are inexpensive to compute on quantile grids.

Alternatives illuminate the trade-offs. Optimal transport distances endow the sample space with a geometry and measure discrepancies through mass relocation. The computational state of the art relies on entropic regularisation and Sinkhorn scaling (Cuturi, 2013), with a broad survey provided by Peyré et al. (2019). While attractive for problems where transport geometry is central, Wasserstein criteria are typically more expensive at high resolution and, in this setting, offer limited practical gains once totals are handled through a cumulative amount alignment. Kernel-based

integral probability metrics, notably the maximum mean discrepancy, supply powerful two-sample tests (Gretton et al., 2012), and energy distances provide a metric based on pairwise norms (Székely and Rizzo, 2013; Rizzo and Székely, 2016). Both families are flexible and can be weighted, but they shift the computational burden to kernel or distance choices. By contrast, CvM aligns naturally with the target quantile grid used to evaluate parametric splices and lends itself to straightforward smoothing and differentiation.

6.2.8 Weighted empirical processes and smoothing devices

Using design-weighted empirical processes is standard in complex surveys for inference and for the construction of test statistics (Binder, 1983; Lumley, 2004). A practical difficulty for optimisation is the non-differentiability of indicator kernels. A well-established remedy is to replace the indicator with a smooth, monotone transition and to treat the bandwidth as a tuning parameter. As Horowitz (1992) shows for the smoothed maximum score, smoothing discontinuous criteria yields tractable optimisation and preserves consistency under appropriate rates; subsequent work develops bootstrap methods and critical values for smoothed objectives (Horowitz, 2002). In the present application, adopting a log-scale Gaussian CDF as a smooth surrogate stabilises tail behaviour and produces closed-form gradients for both the shape and the cumulative-amount components, enabling scalable projected-gradient or quasi-Newton routines under simple box constraints.

6.2.9 From bracket totals to cumulative amount functions

BSR information arrives as bracket totals that encode cumulative amounts by threshold rather than only cumulative counts. The continuous analogue is the cumulative amount function $A(y) = \int_0^y t \, \mathrm{d}F(t)$, which measures the contribution to the total from units with value at most y. When normalised by the instrument total $T = N \, \mathbb{E}(X)$, the relative curve $s(y) = A(y)/T$ is a size-biased distribution function that weights the marginal by the value itself (Patil and Rao, 1978). This object links naturally to the Lorenz curve and to the generalised Lorenz curve – the former measures concentration at a fixed mean, whereas the latter scales by the mean to allow welfare comparisons under different means (Gastwirth, 1972; Shorrocks, 1983). For calibration, the key point is that aligning $s(y)$ along a grid that includes reported thresholds reconciles bracket totals continuously and recovers the overall total at high quantiles by construction. In implementation, the grid also includes a far-tail anchor and the endpoint is explicitly up-weighted within the loss, which sharply discourages total drift without imposing a hard total constraint.

Relative CAF alignment also mitigates two artefacts of class-based calibration. First, it avoids discontinuities and boundary pile-ups by replacing hard threshold indicators with smooth transitions on the log-scale. Second, it promotes coherent scaling by constraining the cumulative path of amounts rather than only the end points. In practice, a design-weighted, smoothed CvM criterion for $s(y)$ can be

combined with a dominant shape term so that distributional fit remains primary while aggregates are recovered through the cumulative channel.

6.2.10 Value adjustment in survey practice: precedents and implications

Most calibration procedures act on weights; direct value adjustment is more common in imputation, small-area estimation, and statistical reconciliation when external information dictates the scale or distribution of a variable. In such contexts, values are perturbed under external constraints and identification hinges on modelling assumptions rather than on design alone. The methodological implication for the presentt chapter is that non-convexity and possible multi-modality are to be expected, so regularisation and well-chosen starting values play a central role. The use of smooth indicators, continuation in bandwidths, and the option of weak shape constraints are standard tools to control these features without reverting to hard class constraints. Related reconciliation frameworks similarly adjust series to respect external benchmarks while controlling distortion, reinforcing the value-level perspective adopted here (Dagum and Cholette, 2006).

6.2.11 Targets for wealth margins and links to earlier chapters

Parametric targets translate BSR information into smooth marginal benchmarks for calibration. In line with the estimation results established earlier in the thesis, the lognormal–Pareto splice provides a parsimonious representation of heavy-tailed instruments and can be estimated reliably on bracketed data (Scollnik, 2007). Evidence on the thickness of wealth tails underscores the importance of a flexible upper tail (Vermeulen, 2018). To avoid repetition, alternatives to the splice are not reviewed here; the focus is on how parametric targets interface with distributional penalties and with the cumulative-amount alignment used in this chapter.

6.2.12 Dependence and multivariate extension

Extending calibration across instruments requires a representation that separates marginals from dependence. Sklar's theorem guarantees that any multivariate distribution can be decomposed into its marginal distribution functions and a copula that encodes dependence (Nelsen, 2006; Joe, 2014). This factorisation supports a modular strategy for the multivariate model of Section 6.1.3: marginal alignment via CvM terms is combined across instruments, and a dependence penalty restores joint structure without reintroducing class reassignments. Semiparametric and rank-based methods make copula estimation practical when marginals are treated separately (Genest and Rémillard, 2008).
A fuller discussion of dependence targets and feasibility is provided in Section 6.2.15.

6.2.13 Two-stage optimisation and stabilisation

Because adjustments act on values, the composite criterion is non-convex, and its terms pull in different directions early in the search. To ensure priority for distributional fit while retaining parsimony, the optimisation proceeds in two stages. Stage 1 minimises a pure fit objective composed of the log-scale CvM shape term and the relative CAF term, under box bounds and initialisation at the total-matching scale. Bandwidths may follow a short continuation schedule (coarser to finer) to stabilise tails. Stage 2 then activates a normalised L_2 stabiliser on $(\varphi - 1)$ with heteroscedastic weights (down-weighting large holdings), but only under tight tolerances that prevent any material degradation of the Stage 1 fit. Tolerances are enforced by smooth hinge penalties and an activation gate that suppresses the stabiliser whenever either fit component exceeds its allowance. This staging avoids a tug-of-war between terms, yields transparent identification (fit first, then shrink among near-ties), and improves numerical stability.

6.2.14 Computation, regularisation, and diagnostics

Because the criterion acts on values and includes distributional penalties, the optimisation problem is non-convex. Three devices make computation reliable in practice. First, smoothing of indicator kernels ensures differentiability of the loss and stabilises tail contributions (Horowitz, 2002). Second, continuation schedules on bandwidths and on the relative weight of the cumulative component guide the search from coarse to fine alignment while keeping the shape dominant early on. Third, weak shape controls can be used to suppress artefacts without overriding the distributional target – isotonic projections and related order-restriction methods provide efficient projections onto monotone cones (Robertson et al., 1988). Diagnostics play a central role: log-scale density overlays against the parametric target, relative CAF plots on grids that include reporting thresholds (with far-tail anchoring and endpoint behaviour), and bracket-by-bracket residuals provide direct checks of fit and stability under alternative grids and smoothing levels.

In sum, the literature supports a distribution-aware, minimum-distance calibration programme in which a log-scale, design-weighted Cramér–von Mises term anchors the marginal shape and a relative cumulative-amount term reconciles aggregates. Weighted empirical processes legitimise the use of design weights in both components; smoothing delivers differentiability and computational stability; Lorenz–generalised Lorenz theory clarifies the role of relative CAF; and copulas offer a coherent route to multivariate extension. These ingredients combine to yield a defensible alternative to bracket-based calibration when weights cannot be altered and the objective is to recover microdata whose distribution matches BSR-derived targets.

6.2.15 From calibrated margins to calibrated dependence: copulas, Gaussian scores, survey weighting, and feasible targets

Sklar's theorem establishes that any multivariate distribution can be factorised into its marginal cumulative distribution functions and a copula that captures dependence (Nelsen, 2006; Joe, 2014). This factorisation justifies a modular strategy for survey calibration: first align the margins to external information, then align the dependence conditional on those margins. Two broad estimation principles support such a two-step perspective. Inference functions for margins (IFM) estimate marginal parameters and then copula parameters; under mild regularity and correct copula specification, the two-step estimator is consistent and asymptotically efficient for many families (Joe, 2005). Semi-parametric copula methods treat the copula parametrically but the margins non-parametrically, using empirical or smoothed ranks; they offer robustness to marginal misspecification and provide tractable asymptotics (Tsukahara, 2005). Both strands motivate the present design: margins are calibrated first, and dependence is targeted on a transformed scale that removes marginal effects by construction.

A practical and widely used device for dependence alignment is the Gaussian-score (or normal-score) transform: for each instrument k, map an adjusted holding Y_k to $U_k = F_k(Y_k)$ through its calibrated marginal, and then to $z_k = \Phi^{-1}(U_k)$. This transformation standardises scale and tail behaviour and places all instruments on a common reference where second-order structure is summarised by correlations of the z_k. Compared with directly penalising copula parameters, targeting correlations of Gaussian scores yields a low-variance and interpretable signal that can be estimated with survey weights and supports regime-specific targets by value segments (Patton, 2006).

Allowing dependence to vary with the level of holdings is consistent with evidence that co-movement strengthens in the upper tail and is weaker or more idiosyncratic at low balances. Tail-dependence concepts from extreme-value theory formalise different joint-tail behaviours (Ledford and Tawn, 1996; Coles et al., 1999), though their direct estimation is often fragile in household surveys, where support at very high values is thin. Segmenting the support using common value thresholds across instruments – here, $[0, 250\text{k})$, $[250\text{k}, \infty)$ – offers a pragmatic compromise. To maintain differentiability of the calibration loss, hard membership is replaced by smooth windows on the log-scale that approximate the common breaks.

Survey consistency matters for dependence estimation. Design-based treatments of empirical and rank-based functionals motivate the use of survey weights when computing moments and correlations (Binder, 1983; Lumley, 2004). On the Gaussian-score scale, design-weighted Pearson correlations of (z_k, z_ℓ) provide natural targets by regime. In practice, clipping the PITs away from $\{0, 1\}$ stabilises score extremes and prevents undue leverage of a few units in thin segments, while preserving smooth derivatives for optimisation.

Feasibility and variance control for target correlations are essential. Empirical

regime-specific correlation matrices can fail to be positive definite because of sampling variability, sparse co-positives, or heavy weights. Enforcing feasibility via nearest-PD projections is well established; Higham's characterisation and algorithms are widely used to produce the closest correlation matrix under the Frobenius norm (Higham, 2002). Beyond feasibility, shrinkage improves conditioning and reduces variance when regime support is limited. Ledoit–Wolf shrinkage towards a structured target delivers well-conditioned estimators in moderately high dimensions (Ledoit and Wolf, 2004). In this chapter, shrinkage is diagnostic-driven and minimal, preserving empirical signals while ensuring stable targets for calibration.

The final methodological block is the composite loss. Let $z(\varphi)$ denote the Gaussian scores implied by the current adjustments φ. For each regime b, let $R_b^\star$ be the feasible target correlation matrix and $R_b(z(\varphi))$ its survey-weighted empirical counterpart under smooth membership. A dependence penalty of the form

$$L_{\mathrm{dep}}(\varphi) \;=\; \sum_b \big\| R_b\big(z(\varphi)\big) \;-\; R_b^\star \big\|_F^2$$

aggregates discrepancies in a convex, rotation-invariant way and admits simple gradients via the score Jacobians.

The composite objective is non-convex because adjustments act on values and because the component losses pull in different directions early in the search. A staged optimisation mitigates this. Stage 1 focuses on fit: for each marginal, the log-scale CvM shape term and the relative CAF term (with tail anchoring and endpoint up-weighting), augmented by a dependence term that is turned on gradually once marginals are essentially aligned. Stage 2 performs stabilisation: a normalised, heteroscedastic L_2 penalty on $(\varphi - 1)$, gated by smooth hinges that freeze the stabiliser whenever any of the fit terms (shape, CAF, dependence) threatens to exceed preset tolerances. This separation protects tail alignment, prevents over-shrinkage of large holdings, and yields transparent diagnostics: improvements in dependence are not purchased at the cost of marginal misfit or total drift. The design mirrors two-step practices in IFM and semiparametric copula estimation (Joe, 2005; Tsukahara, 2005), but it is implemented as a continuous optimisation with differentiable surrogates tailored to survey-weighted functionals.

6.2.16 Alternatives for multivariate reconciliation and dependence preservation

A first alternative is bracket-based multivariate benchmarking. Extending the class-constrained calibration of Section 6.1.1 to multiple instruments would enforce cross-instrument totals by class. However, endogenous class reassignment proliferates combinatorially across instruments and thresholds. The result is boundary pile-ups, sensitivity to the order of updates, and heavy iteration to maintain coherence across marginals; the approach also struggles to represent smooth changes in dependence across value regimes.

A second alternative is transport-based multivariate alignment. Optimal transport can, in principle, match a joint target distribution by solving a global mass-relocation problem. Entropic regularisation and Sinkhorn scaling have made high-dimensional transport more tractable (Cuturi, 2013; Peyré et al., 2019). Yet, in survey settings with thin joint tails and fixed weights, the computational burden, the sensitivity to sparsity in the upper regimes, and the challenge of preserving survey-weighted structure limit applicability. Moreover, specifying a credible joint target beyond marginals requires additional modelling assumptions that are not supported by administrative sources.

A third route is full copula calibration. One could estimate a single parametric copula on calibrated margins and penalise deviations in copula parameters. This is elegant when a common copula fits well across the support. In practice, wealth portfolios exhibit regime changes: dependence is weaker and more heterogeneous at low balances and stronger, sometimes more structured, at high balances. Conditional or time-varying copulas can capture such patterns (Patton, 2006), but they require conditioning rules that are not naturally available in range-sum aggregates. Mixtures of copulas or vine copulas increase flexibility at the cost of high dimensionality and difficult identification in the tails. By targeting regime-specific correlations of Gaussian scores, the present method captures salient dependence changes with minimal structure and diagnostic transparency.

A fourth family is based on partial correlations and graphical models. Penalising discrepancies in partial correlations directly aligns conditional dependence; sparse targets connect to graphical lasso and related precision-matrix estimators (Banerjee et al., 2008; Friedman et al., 2008). These tools are powerful in high-dimensional contexts with strong sparsity. In the present application, where the number of instruments is modest and the primary uncertainty lies in tails and segment support, the additional nonlinearity and positive definiteness constraints in the loss offer limited gains over correlation alignment, while complicating survey-weighted estimation and diagnostics. Partial-correlation targets remain a viable extension when diagnostics reveal specific conditional-dependence misspecification.

Finally, tail-focused approaches from extreme-value theory offer refined measures of tail association (Ledford and Tawn, 1996; Coles et al., 1999; Heffernan and Tawn, 2004). They are appealing when the primary inferential target is joint extremes. In household surveys, however, the limited number of co-positives in the upper regimes makes tail-only calibration unstable. Incorporating tail-aware checks as diagnostics (e.g., co-exceedance rates within the high segment) is feasible within the present framework, and small penalties can be added as needed without altering the core objective.

6.2.17 Positioning and contribution

The method proposed in Section 6.1.3 contributes a dependence-preserving calibration tailored to survey microdata and to administrative information that credibly

identifies marginals but only partially constrains joint structure. Its novelty lies in four design choices. First, it targets dependence on the Gaussian-score scale induced by the calibrated marginals, removing scale and tail effects by construction. Second, it defines and estimates regime-specific correlation targets using common value segments across instruments. This choice reflects economic behaviour (transactions and buffers at the bottom; deliberate risk at the top), sidesteps heterogeneity in administrative brackets, and keeps targets comparable across years. Third, it ensures feasibility and stability of the targets themselves. Regime-specific correlation matrices are made positive definite through nearest-PD projection and regularised only when diagnostics indicate instability, using minimal shrinkage towards structured targets. Fourth, it integrates the dependence term into a composite loss that respects the primacy of marginal alignment: the dependence penalty is normalised and introduced progressively in Stage 1; Stage 2 activates a normalised, heteroscedastic stabiliser under tolerances that encompass shape, CAF and dependence.

Relative to bracket-based multivariate benchmarking, the method avoids endogenous class reassignment and boundary pile-ups. Relative to transport-based alignment, it reduces modelling requirements and computational burden while retaining survey-weighted structure. Relative to full copula calibration, it provides regime flexibility without committing to a single family across the support. Relative to graphical-model approaches, it prioritises interpretability and survey compatibility over sparsity-driven conditional structure, while leaving room for partial-correlation extensions when diagnostics warrant them.

In sum, the chapter advances a multivariate, survey-compatible calibration that preserves the gains of the distribution-aware model of Section 6.1.2 on the marginals and adds a minimal yet effective dependence layer. It leverages copula modularity without forcing a global parametric fit; it estimates targets in a way that is feasible under complex designs; and it embeds these targets in a smooth, staged optimisation that keeps distributional alignment primary while making joint structure empirically credible and operationally tractable.

6.3 Theoretical framework

This section collects the three calibration models into a single theoretical framework. All models keep survey weights fixed and act on values through multiplicative adjustments. Model 1 targets BSR class totals through bounded L_2 calibration with endogenous class membership and a class-aware projection scheme. Model 2 replaces hard class constraints with a composite distributional objective that aligns both the marginal shape and the relative cumulative amount function (CAF). Model 3 extends Model 2 to a multivariate setting by adding a dependence penalty on Gaussian-score correlations computed after a parametric probability integral transform (PIT).

Throughout, households are indexed by h and instruments by $k \in \{1, \ldots, K\}$. We write $w_h > 0$ for the fixed survey weight. For each instrument k, let $x_{hk} \geq 0$ be

the reported amount, $\varphi_{hk} > 0$ the adjustment factor, and

$$y_{hk} \;=\; \varphi_{hk}\, x_{hk}$$

the adjusted value. Non-holders are kept fixed: $\varphi_{hk} = 1$ whenever $x_{hk} = 0$.

Taken together, the three models form a staged calibration pipeline. Model 1 enforces exact agreement with BSR range-sum (class-total) constraints under explicit household-level bounds, but it leaves within-class distributional shape only weakly identified and it treats instruments separately. Model 2 therefore replaces hard class constraints with a distribution-aware criterion that targets the full marginal CDF and the cumulative-amount path implied by the BSR-estimated parametric margin. Finally, Model 3 extends the same marginal machinery to the multivariate setting by adding a dependence penalty on a marginal-free (PIT / Gaussian-score) scale, so that portfolio co-movements are restored without reintroducing weight calibration.

6.3.1 Model 1: class-penalised bounded calibration to BSR class totals

This subsection formalises bounded, value-based calibration to BSR class totals. It first treats a fixed class partition, deriving feasibility ranges and the Karush–Kuhn–Tucker (KKT) characterisation (Kuhn and Tucker, 1951). It then states the global problem with endogenous class membership and presents the sequential scheme with class-aware projection. Brackets for instrument k are indexed by $j \in \{1, \ldots, J_k\}$.

Objects, notation, and class totals

For each instrument k, let J_k be the number of BSR value classes and let

$$0 < a_{k1} < a_{k2} < \cdots < a_{k\,J_k-1}$$

denote the corresponding finite cut-offs. Classes follow the left-closed, right-open convention:

$$[0, a_{k1}), \quad [a_{k\,j-1}, a_{kj}) \;(j = 2, \ldots, J_k - 1), \quad [a_{k\,J_k-1}, \infty).$$

Denote by $T_{kj} \geq 0$ the BSR target total for instrument k and class j. The HFCS-implied class total induced by $\mathbf{y} = (y_{hk})$ is

$$\widehat{T}_{kj}(\mathbf{y}) \;=\; \sum_h w_h\, y_{hk}\, \mathbb{1}\{y_{hk} \in \mathcal{C}_{kj}\}, \tag{6.2}$$

where $\mathcal{C}_{kj}$ denotes the j-th class interval listed above. Model 1 constrains multipliers to lie in instrument-specific bounds $\varphi_{hk} \in [\underline{\varphi}_k, \overline{\varphi}_k]$. Throughout, we use baseline bounds $[\underline{\varphi}_k, \overline{\varphi}_k] = [0.5, 10]$, treated as a conservative default that prevents implausibly large household-level adjustments and stabilises the calibration in the presence of sharp bracket thresholds.

Fixed-class problem: equality-constrained quadratic programme

Fix a reference vector $\mathbf{y}_k$ (e.g. the current iterate for instrument k) and the associated class sets

$$\mathcal{B}_{kj} = \{h : \ a_{k,j-1} \le y_{hk} < a_{k,j}\} \qquad (j = 1, \dots, J_k).$$

With classes fixed, (6.2) is linear in $\mathbf{y}$ and hence in $\boldsymbol{\varphi}_{\cdot k} = (\varphi_{hk})_h$ over households with $x_{hk} > 0$. Define the weighted incidence matrix $\mathbf{B}_k \in \mathbb{R}^{J_k \times H_k}$ and target vector $\mathbf{t}_k \in \mathbb{R}^{J_k}$ by

$$(\mathbf{B}_k)_{j,h} = w_h \, x_{hk} \, \mathbb{1}\{h \in \mathcal{B}_{kj}\}, \qquad \mathbf{t}_k = (T_{k1}, \dots, T_{kJ_k})^{\mathsf{T}}.$$

Let $\mathbf{W}_k = \operatorname{diag}(w_1, \dots, w_{H_k}) \in \mathbb{R}^{H_k \times H_k}$ denote the diagonal matrix of survey weights restricted to households with $x_{hk} > 0$ (so that H_k is the number of such households for instrument k). The fixed-class calibration for instrument k solves

$$\begin{aligned} \min_{\boldsymbol{\varphi}_{\cdot k}} \quad & (\boldsymbol{\varphi}_{\cdot k} - \mathbf{1})^{\mathsf{T}} \mathbf{W}_k \, (\boldsymbol{\varphi}_{\cdot k} - \mathbf{1}) \\ \text{s.t.} \quad & \mathbf{B}_k \, \boldsymbol{\varphi}_{\cdot k} = \mathbf{t}_k, \qquad \underline{\varphi}_k \le \varphi_{hk} \le \overline{\varphi}_k \ \ \forall h. \end{aligned} \tag{6.3}$$

Feasibility ranges (fixed classes). For each j, the attainable class total lies in

$$\mathcal{I}_{kj} \ = \ \Big[\sum_{h \in \mathcal{B}_{kj}} w_h \, (\underline{\varphi}_k x_{hk}) \, , \ \sum_{h \in \mathcal{B}_{kj}} w_h \, (\overline{\varphi}_k x_{hk}) \Big] \ = \ \Big[\sum_{h \in \mathcal{B}_{kj}} w_h \, y^L_{hk} \, , \ \sum_{h \in \mathcal{B}_{kj}} w_h \, y^U_{hk} \Big], \tag{6.4}$$

where $y^L_{hk} = \underline{\varphi}_k x_{hk}$ and $y^U_{hk} = \overline{\varphi}_k x_{hk}$ denote the lower- and upper-bound adjusted values. A necessary and sufficient condition for joint feasibility under fixed classes is $T_{kj} \in \mathcal{I}_{kj}$ for all j (classes are disjoint).

Karush–Kuhn–Tucker characterisation. Let $\boldsymbol{\lambda}_k \in \mathbb{R}^{J_k}$ be the multipliers for the equalities and $\boldsymbol{\mu}_k, \boldsymbol{\nu}_k \ge \mathbf{0}$ those for the lower and upper bounds. The KKT (Kuhn and Tucker, 1951) system is

$$\begin{aligned} & 2 \, \mathbf{W}_k (\boldsymbol{\varphi}_{\cdot k} - \mathbf{1}) + \mathbf{B}_k^{\mathsf{T}} \boldsymbol{\lambda}_k - \boldsymbol{\mu}_k + \boldsymbol{\nu}_k = \mathbf{0}, \\ & \mathbf{B}_k \, \boldsymbol{\varphi}_{\cdot k} = \mathbf{t}_k, \qquad \underline{\varphi}_k \mathbf{1} \le \boldsymbol{\varphi}_{\cdot k} \le \overline{\varphi}_k \mathbf{1}, \\ & \boldsymbol{\mu}_k \odot (\boldsymbol{\varphi}_{\cdot k} - \underline{\varphi}_k \mathbf{1}) = \mathbf{0}, \quad \boldsymbol{\nu}_k \odot (\overline{\varphi}_k \mathbf{1} - \boldsymbol{\varphi}_{\cdot k}) = \mathbf{0}, \quad \boldsymbol{\mu}_k, \boldsymbol{\nu}_k \ge \mathbf{0}, \end{aligned} \tag{6.5}$$

with $\odot$ denoting the Hadamard product.

Global problem with endogenous class membership

Brackets $\{a_{k,j}\}_{j=0}^{J_k}$ are exogenous. What is endogenous is the class membership induced by the adjusted values:

$$c_{hk}(\mathbf{y}) = j \quad \Longleftrightarrow \quad a_{k,j-1} \le y_{hk} < a_{k,j}.$$

Equivalently, for $x_{hk} > 0$,

$$\mathbb{1}\{c_{hk}(\mathbf{y}) = j\} \ = \ \mathbb{1}\Big\{ \tfrac{a_{k,j-1}}{x_{hk}} \le \varphi_{hk} < \tfrac{a_{k,j}}{x_{hk}} \Big\},$$

so that the equality constraints $\widehat{T}_{kj}(\mathbf{y}) = T_{kj}$ depend on $\boldsymbol{\varphi}$ through state-dependent indicators. The objective

$$\sum_{k,h} w_h \, (\varphi_{hk} - 1)^2$$

is strictly convex, but the feasible set is a union of convex regions separated by threshold hyperplanes $\varphi_{hk} = a_{k,j}/x_{hk}$. Hence, the global programme is piecewise convex and globally non-convex; gradients are well defined within any region with fixed membership and discontinuous on thresholds. Units with $x_{hk} = 0$ are unaffected by class rules and can be treated by a zero bin or by fixing $\varphi_{hk} \equiv 1$.

Within any region where memberships are frozen, the problem reduces to the convex QP in (6.3). The practical algorithm therefore alternates between: (*i*) freezing classes at the previous iterate $\mathbf{y}^{(t-1)}$ and solving the fixed-class QP; (*ii*) updating $\mathbf{y}$ and applying an ex-post, class-aware projection to curb reclassification near thresholds; (*iii*) recomputing classes and repeating. No global optimality claim is made; selection by the diagnostic $G^{(t)}$ in (6.6) retains the best iterate observed.

Operationally, the equalities are enforced by freezing class membership at the previous iterate, so that the implied class totals are linear in the current multipliers; the full iterative scheme and the projection step are stated in Section 6.3.1.

No global optimality claim is made; selection by the diagnostic $G^{(t)}$ in (6.6) retains the best iterate observed. This sequential approach is sufficient for production, where the class-level targets T_{kj} are externally mandated benchmarks (taken as fixed inputs from the BSR aggregates) and the bounds $[\underline{\varphi}_k, \overline{\varphi}_k]$ are analyst-imposed safeguards designed to prevent implausibly large household-level adjustments.

Sequential scheme with class-aware projection

Let $\mathbf{y}^{(0)} \equiv \mathbf{x}$. For $t = 1, 2, \ldots$:

Step 1: Class freezing. For each k, set $\mathcal{B}_{kj}^{(t)} = \{h : \; a_{k,j-1} \leq y_{hk}^{(t-1)} < a_{k,j}\}$ and build $\mathbf{B}_k^{(t)}$ accordingly. Define the frozen-class total operator at iteration t as

$$\widehat{T}_{kj}^{(t)}(\boldsymbol{\varphi}_k) \; := \; \sum_h w_h \, x_{hk} \, \varphi_{hk} \, \mathbb{1}\{h \in \mathcal{B}_{kj}^{(t)}\},$$

so that enforcing $\widehat{T}_{kj}^{(t)}(\boldsymbol{\varphi}_k) = T_{kj}$ is equivalent to $\mathbf{B}_k^{(t)} \boldsymbol{\varphi}_k = \mathbf{t}_k$. In Step 2 the equalities are imposed with respect to these frozen sets $\{\mathcal{B}_{kj}^{(t)}\}_j$.

Step 2: Bounded equality-constrained QP. Solve (6.3) with $\mathbf{B}_k^{(t)}$ and bounds $[\underline{\varphi}_k, \overline{\varphi}_k]$, obtaining $\boldsymbol{\varphi}_{\cdot k}^{(t)}$. Set $y_{hk}^{(t)} \leftarrow \varphi_{hk}^{(t)} x_{hk}$.

Step 3: Ex-post, class-aware projection. Define $\Pi_{[0.95,1.05]}(x) = \min\{1.05, \max\{0.95, x\}\}$ and identify reclassifiers

$$\mathcal{M}_k^{(t)} = \{h : \; c_{hk}(\mathbf{y}^{(t)}) \neq c_{hk}(\mathbf{y}^{(t-1)})\}.$$

Project their multipliers to $[0.95, 1.05]$ via

$$\varphi_{hk}^{(t)} \leftarrow \Pi_{[0.95,1.05]}(\varphi_{hk}^{(t)}) \quad (h \in \mathcal{M}_k^{(t)}),$$

and update $y_{hk}^{(t)} \leftarrow \varphi_{hk}^{(t)} x_{hk}$ for those h. The near-unity band $[0.95, 1.05]$ is a deliberately tight trust region: it is chosen so that only observations sufficiently close to a bracket threshold can change class after projection, thereby limiting large-scale reclassification within an iteration and stabilising the sequential updates without relaxing the hard range-sum equalities.

Step 4: Selection and stopping. Compute the class-gap diagnostic

$$G^{(t)} \;=\; \sum_{k=1}^{K} \sum_{j=1}^{J_k} \Big(\widehat{T}_{kj}(\mathbf{y}^{(t)}) - T_{kj} \Big)^2, \tag{6.6}$$

retain the best iterate $t^\star = \arg\min_{1 \le s \le t} G^{(s)}$, and stop when no material improvement is observed for a fixed number of iterations or when a maximum number of iterations is reached. Because class membership is recomputed after each update, $G^{(t)}$ is not guaranteed to decrease monotonically across t: reclassification induced by threshold crossings can temporarily increase the class-gap even when the within-class QP is solved exactly. The reported solution is $\tilde{\mathbf{y}}^{(t^\star)}$, where $t^\star = \arg\min_{1 \le s \le t} G^{(s)}$.

Properties and feasibility diagnostics

Each fixed-class sub-problem admits at least one solution. The ex-post projection preserves feasibility since $[0.95, 1.05] \subseteq [\underline{\varphi}_k, \overline{\varphi}_k]$. At iteration t, feasibility under frozen classes holds if and only if $T_{kj} \in \mathcal{I}_{kj}^{(t)}$ for all j, where $\mathcal{I}_{kj}^{(t)}$ is obtained from (6.4) by replacing $\mathcal{B}_{kj}$ with $\mathcal{B}_{kj}^{(t)}$. Infeasibility can be handled by widening $[\underline{\varphi}_k, \overline{\varphi}_k]$ or introducing small slacks.

Implementation notes

Numerically, (6.3) is a small to medium-sized convex QP with simple bounds and few equalities. Conditioning improves if values are scaled before constructing $\mathbf{B}_k$; the solution is invariant once targets and bounds are scaled consistently. Ties at boundaries follow the left-closed convention $a_{k,j-1} \le x < a_{k,j}$. Model 1 delivers a production-friendly reconciliation to BSR range-sum class totals under explicit bounds, but it remains intrinsically class-driven: the constraints depend on endogenous class membership and are therefore piecewise constant in bracket assignment. As a result, within-bracket shape is only weakly disciplined beyond the truncated-sum equalities, and tight reclassification control can induce visible accumulation near cut-offs when micro and benchmark distributions disagree. These features motivate a shift from hard class constraints to smooth minimum-distance objectives that target the full marginal distribution while keeping survey weights fixed. Model 2 therefore replaces the class-total equalities with a composite loss that aligns both the marginal CDF and the cumulative amount path, yielding a continuous optimisation problem in the value adjustments.

6.3.2 Model 2: distribution-aware calibration with composite loss

This subsection formalises the composite calibration criterion and its optimisation. The calibration keeps survey weights fixed and adjusts reported values through multiplicative factors – the object of calibration is the vector of values, not the weights. We consider one instrument at a time and suppress the instrument index. Relative to Model 1, the object of calibration remains value-based (multiplicative adjustments on holdings with fixed survey weights), but identification is shifted from hard bracket equalities to smooth, design-weighted discrepancies between empirical functionals and the BSR-implied target margin.

Households are indexed by h. Let $x_h \geq 0$ denote the observed amount and $w_h > 0$ its fixed survey weight. Let $\varphi_h > 0$ be the adjustment and $y_h = \varphi_h x_h$ the adjusted value. (We use y_h to avoid a notation clash with Gaussian-score variables introduced in Model 3.) Define the active set $\mathcal{H} = \{h : \ x_h > 0\}$, its cardinality $n = |\mathcal{H}|$, and the corresponding total survey weight $W = \sum_{h\in\mathcal{H}} w_h$. Units with $x_h = 0$ are kept at $\varphi_h = 1$ and are excluded from the margin-specific divergences.

From BSR we take a parametric target margin with CDF F and density f. Let $C(y) = \int_0^y t\,\mathrm{d}F(t)$ be the cumulative amount function, $T = \int_0^\infty t\,\mathrm{d}F(t)$ the total, and $s(y) = C(y)/T$ the relative cumulative share. Comparisons are implemented on the log-scale using $\Phi(\cdot)$ and $\phi(\cdot)$, with bandwidths $h_{\text{shape}} > 0$ and $h_{\text{caf}} > 0$. Let $\{y_g\}_{g=1}^m$ be an evaluation grid with weights $\{\omega_g\}$ that sum to one.

Complete composite loss

Define the smoothed, design-weighted functionals

$$\widehat{F}(y_g;\boldsymbol{\varphi}) = \frac{1}{W}\sum_{h\in\mathcal{H}} w_h\,\Phi\!\left(\frac{\log y_g - \log y_h}{h_{\text{shape}}}\right), \qquad \widehat{C}(y_g;\boldsymbol{\varphi}) = \sum_{h\in\mathcal{H}} w_h y_h\,\Phi\!\left(\frac{\log y_g - \log y_h}{h_{\text{caf}}}\right),$$

$$\widehat{s}(y_g;\boldsymbol{\varphi}) = \frac{\widehat{C}(y_g;\boldsymbol{\varphi})}{T}, \qquad y_h = \varphi_h x_h.$$

The three components are

$$L_{\text{shape}}(\boldsymbol{\varphi}) = \sum_{g=1}^m \omega_g\big(\widehat{F}(y_g;\boldsymbol{\varphi}) - F(y_g)\big)^2,$$

$$L_{\text{caf}}(\boldsymbol{\varphi}) = \sum_{g=1}^m \omega_g\big(\widehat{s}(y_g;\boldsymbol{\varphi}) - s(y_g)\big)^2,$$

$$L_{\text{stab}}(\boldsymbol{\varphi}) = \frac{\sum_{h\in\mathcal{H}} w_h\,h_x(x_h)\,(\varphi_h - 1)^2}{\big(\sum_{h\in\mathcal{H}} w_h\,h_x(x_h)\big)\,M^2},$$

where $h_x(x_h) \geq 0$ is a user-specified stabilisation weight and $M^2 = \max\{(\underline{\varphi}_k - 1)^2, (\overline{\varphi}_k - 1)^2\}$. The complete objective is

$$L(\boldsymbol{\varphi}) = L_{\text{shape}}(\boldsymbol{\varphi}) + \lambda_{\text{caf}}\,L_{\text{caf}}(\boldsymbol{\varphi}) + \lambda_{\text{stab}}\,L_{\text{stab}}(\boldsymbol{\varphi}). \tag{6.7}$$

A two-step strategy improves stability and interpretability. Step 1 secures a shape-oriented solution driven by L_{shape} and L_{caf}; Step 2 then searches within a narrow tolerance band around the Step 1 fit for the minimal-adjustment solution, using a gated stabiliser normalised in $[0, 1]$.

Step 1: shape-oriented fit

Step 1 sets $\lambda_{\text{stab}} = 0$ and minimises

$$L_{\text{fit}}(\boldsymbol{\varphi}) = L_{\text{shape}}(\boldsymbol{\varphi}) + \lambda_{\text{caf}} L_{\text{caf}}(\boldsymbol{\varphi}). \tag{6.8}$$

The tuning parameter λ_{caf} governs the relative emphasis placed on matching the relative cumulative amount function $s(\cdot)$ versus matching the marginal distribution $F(\cdot)$. In practice it is selected empirically: we run a small grid search over candidate values and retain the smallest λ_{caf} that delivers an adequate fit in L_{caf} without materially degrading L_{shape} on the evaluation grid. A larger λ_{caf} prioritises matching the cumulative amount path (and thus the implied concentration of holdings), potentially at the cost of a looser fit in the marginal CDF, while a smaller λ_{caf} does the opposite. For parsimony, the reported results use the selected λ_{caf}. Initialise at the total-matching scale $\varphi_h^{(0)} = T / \sum_{h \in \mathcal{H}} w_h x_h$. Let

$$u_{gh} = \frac{\log y_g - \log y_h}{h_{\text{shape}}}, \qquad v_{gh} = \frac{\log y_g - \log y_h}{h_{\text{caf}}}, \qquad y_h = \varphi_h x_h.$$

By direct differentiation, for $h \in \mathcal{H}$,

$$\frac{\partial \widehat{F}(y_g; \boldsymbol{\varphi})}{\partial \varphi_h} = -\frac{w_h}{W} \frac{1}{h_{\text{shape}} \varphi_h} \phi(u_{gh}), \tag{6.9}$$

$$\frac{\partial \widehat{C}(y_g; \boldsymbol{\varphi})}{\partial \varphi_h} = w_h \Big[x_h \, \Phi(v_{gh}) - \frac{x_h}{h_{\text{caf}}} \, \phi(v_{gh}) \Big], \qquad \frac{\partial \widehat{s}(y_g; \boldsymbol{\varphi})}{\partial \varphi_h} = \frac{1}{T} \frac{\partial \widehat{C}(y_g; \boldsymbol{\varphi})}{\partial \varphi_h}. \tag{6.10}$$

This log-scale smoothing and its differentiability are standard devices for stabilising discontinuous criteria (Horowitz, 1992). Hence, the gradient of (6.8) is, for $h \in \mathcal{H}$,

$$\begin{aligned} \frac{\partial L_{\text{fit}}}{\partial \varphi_h} &= 2 \sum_{g=1}^{m} \omega_g \big(\widehat{F}(y_g; \boldsymbol{\varphi}) - F(y_g) \big) \Big(- \frac{w_h}{W} \frac{1}{h_{\text{shape}} \varphi_h} \phi(u_{gh}) \Big) \\ &\quad + \frac{2 \, \lambda_{\text{caf}}}{T} \sum_{g=1}^{m} \omega_g \big(\widehat{s}(y_g; \boldsymbol{\varphi}) - s(y_g) \big) \, w_h \Big[x_h \, \Phi(v_{gh}) - \frac{x_h}{h_{\text{caf}}} \, \phi(v_{gh}) \Big]. \end{aligned} \tag{6.11}$$

Step 2: tie-breaker with gated stabilisation

Let $\boldsymbol{\varphi}^{(1)}$ be a Step 1 solution and record $s_1 = L_{\text{shape}}(\boldsymbol{\varphi}^{(1)})$ and $c_1 = L_{\text{caf}}(\boldsymbol{\varphi}^{(1)})$. Choose small tolerances $\epsilon_{\text{s}}, \epsilon_{\text{c}} > 0$ and set $\tau_{\text{s}} = (1 + \epsilon_{\text{s}}) s_1$, $\tau_{\text{c}} = (1 + \epsilon_{\text{c}}) c_1$. To avoid non-differentiable kinks, use a smoothed hinge $H_\delta(r) = \frac{1}{2}\big(r + \sqrt{r^2 + \delta^2}\big)$ with derivative $H'_\delta(r) = \frac{1}{2}\Big(1 + \frac{r}{\sqrt{r^2 + \delta^2}}\Big)$. Define the relative deviations

$$r_{\text{s}}(\boldsymbol{\varphi}) = \frac{L_{\text{shape}}(\boldsymbol{\varphi}) - \tau_{\text{s}}}{\tau_{\text{s}}}, \qquad r_{\text{c}}(\boldsymbol{\varphi}) = \frac{L_{\text{caf}}(\boldsymbol{\varphi}) - \tau_{\text{c}}}{\tau_{\text{c}}},$$

and the activation gate

$$\mathrm{act}(\boldsymbol{\varphi}) = \exp\Big\{ - k_{\mathrm{act}}\big(H_\delta(r_{\mathrm{s}}(\boldsymbol{\varphi})) + H_\delta(r_{\mathrm{c}}(\boldsymbol{\varphi}))\big)\Big\}, \qquad k_{\mathrm{act}} > 0.$$

Step 2 then minimises the gated stabiliser with a drift penalty

$$L_{\mathrm{tie}}(\boldsymbol{\varphi}) = \lambda_{\mathrm{stab}}\, L_{\mathrm{stab}}(\boldsymbol{\varphi})\, \mathrm{act}(\boldsymbol{\varphi}) + \rho\Big(H_\delta(r_{\mathrm{s}}(\boldsymbol{\varphi}))^2 + H_\delta(r_{\mathrm{c}}(\boldsymbol{\varphi}))^2\Big). \tag{6.12}$$

The coefficients λ_{stab} and k_{act} are tuning parameters rather than model parameters estimated from the data: λ_{stab} controls the strength of shrinkage toward one conditional on being within the Step 1 tolerance region, while k_{act} controls how sharply the activation gate $\mathrm{act}(\boldsymbol{\varphi})$ switches off shrinkage when the fit departs from that region. In implementation, ρ is set sufficiently large that violating the Step 1 tolerances is dominated by the drift penalty, and $(\lambda_{\mathrm{stab}}, k_{\mathrm{act}})$ are selected empirically to achieve the smallest adjustments without materially increasing L_{shape} and L_{caf} beyond τ_{s} and τ_{c}. The stabiliser L_{stab} is normalised in $[0,1]$ by construction and is multiplied by $\mathrm{act}(\boldsymbol{\varphi}) \in (0,1]$, which sharply decreases if either L_{shape} or L_{caf} exceeds its tolerance. Consequently, shrinkage operates only within the near-flat region around $\boldsymbol{\varphi}^{(1)}$, preserving the Step 1 fit while selecting the smallest adjustment vector.

The gradient of the stabiliser is, for $h \in \mathcal{H}$,

$$\frac{\partial L_{\mathrm{stab}}}{\partial \varphi_h} = \frac{2\, w_h\, h_x(x_h)}{\big(\sum_{\ell\in\mathcal{H}} w_\ell h_x(x_\ell)\big)\, M^2}\,(\varphi_h - 1). \tag{6.13}$$

By the chain rule, using (6.13) and (6.11), the Step 2 gradient is

$$\begin{aligned}\frac{\partial L_{\mathrm{tie}}}{\partial \varphi_h} &= \lambda_{\mathrm{stab}}\Big[\mathrm{act}\,\frac{\partial L_{\mathrm{stab}}}{\partial \varphi_h} + L_{\mathrm{stab}}\frac{\partial \mathrm{act}}{\partial \varphi_h}\Big] \\ &\quad + 2\rho\, H_\delta(r_{\mathrm{s}})\, H'_\delta(r_{\mathrm{s}})\,\frac{1}{\tau_{\mathrm{s}}}\,\frac{\partial L_{\mathrm{shape}}}{\partial \varphi_h} + 2\rho\, H_\delta(r_{\mathrm{c}})\, H'_\delta(r_{\mathrm{c}})\,\frac{1}{\tau_{\mathrm{c}}}\,\frac{\partial L_{\mathrm{caf}}}{\partial \varphi_h},\end{aligned} \tag{6.14}$$

with

$$\frac{\partial \mathrm{act}}{\partial \varphi_h} = -k_{\mathrm{act}}\,\mathrm{act}\,\Big(H'_\delta(r_{\mathrm{s}})\,\frac{1}{\tau_{\mathrm{s}}}\,\frac{\partial L_{\mathrm{shape}}}{\partial \varphi_h} + H'_\delta(r_{\mathrm{c}})\,\frac{1}{\tau_{\mathrm{c}}}\,\frac{\partial L_{\mathrm{caf}}}{\partial \varphi_h}\Big).$$

Two-step optimisation is crucial for three reasons. First, it isolates the distributional target from the stabilisation pressure, preventing premature shrinkage that could bias the tail alignment. Second, it supplies a clear identification narrative: Step 1 fixes the global scale and shape relative to F and s, Step 2 then selects the minimal ℓ_2-type deformation consistent with that fit. Third, it reduces numerical instability and multi-modality by guiding the search along a sequence of simpler objectives before activating the stabiliser.

Constraints, implementation, and diagnostics

Box bounds $[\underline{\varphi}_k, \overline{\varphi}_k]$ prevent pathological adjustments. A mild heteroscedastic stabiliser through $h_x(x_h)$ reduces undue shrinkage on very small x_h while down-weighting large holdings. Order preservation in the x-order can be considered as

an option (e.g., a light adjacency penalty or projection onto the isotonic cone as in (Robertson et al., 1988)), though it is not used in the baseline results. Optimisation uses L-BFGS-B with analytical gradients and a two-step continuation: in Step 1 start with $h_{\text{caf}} \approx 1.5\, h_{\text{shape}}$ and shrink both as alignment improves; in Step 2 keep bandwidths fixed and calibrate λ_{stab}, k_{act}, ρ and the tolerances. The tolerances ϵ_{s} and ϵ_{c} (hence τ_{s} and τ_{c}) are chosen as small percentage relaxations around the Step 1 losses, so that Step 2 searches only within a near-tie region of distributional fit; operationally, we set ϵ_{s} and ϵ_{c} to the smallest values that still allow non-trivial shrinkage (i.e. that do not force $\text{act}(\boldsymbol{\varphi})$ to collapse to zero for all feasible iterates). The smoothing parameter δ in H_δ is selected for numerical stability and is set sufficiently small that the hinge closely approximates $\max\{r, 0\}$ while keeping derivatives well behaved.

Diagnostics include overlays of log-densities and of the relative cumulative amount on the evaluation grid, residuals at bracket thresholds, total residuals, and plots of φ_h against x_h to detect humps. The use of design-weighted empirical functionals and smooth indicator approximations follows standard practice in complex surveys and in smoothed criteria for optimisation (Binder, 1983; Lumley, 2004; Horowitz, 1992).

While Model 2 targets each margin in a distribution-aware way, marginal alignment alone does not identify a joint distribution: many dependence structures are compatible with the same set of calibrated marginals. In portfolio data, this is consequential because cross-instrument co-movement affects both concentration and composition in the upper tail. Model 3 addresses this remaining degree of freedom by introducing an explicit dependence penalty, defined on a marginal-free scale via probability integral transforms and Gaussian scores, and activated in a staged manner so that dependence improvements are not purchased at the cost of marginal misfit.

6.3.3 Model 3: multivariate composite calibration with dependence control

This subsection formalises the multivariate calibration that extends Model 2 from calibrated margins to calibrated dependence. Operationally, Model 3 treats the Model 2 output as the starting point: marginal alignment is secured first, and the dependence penalty is then introduced gradually so that improvements in correlation structure do not come at the expense of marginal shape or cumulative-amount fit. Households are indexed by $h = 1, \ldots, n$ and instruments by $k = 1, \ldots, K$. Let $x_{hk} \geq 0$ denote reported holdings, $w_h > 0$ the fixed survey weight, and $\varphi_{hk} > 0$ the multiplicative adjustment. Non-holders are kept fixed, $\varphi_{hk} = 1$ whenever $x_{hk} = 0$. Adjusted values are $y_{hk} = \varphi_{hk} x_{hk}$ and $W = \sum_{h=1}^{n} w_h$.

Each margin k is represented by a parametric distribution F_k with density f_k, with the same grids and smoothing conventions adopted in Model 2. For each margin k, define the active set of holders $\mathcal{H}_k = \{h : x_{hk} > 0\}$ and its total survey weight $W_k = \sum_{h \in \mathcal{H}_k} w_h$. As in Model 2, non-holders are kept at $\varphi_{hk} = 1$ and excluded from the margin-specific CDF and CAF discrepancies.

Relative to Model 2, the marginal components of the objective are unchanged (shape and relative CAF terms by instrument), but the criterion is augmented by a dependence block that targets regime-specific correlations of Gaussian scores under survey weighting, with feasibility and shrinkage handled upstream and with staged activation during optimisation.

Margins, parametric PIT, Gaussian scores, and smooth regime membership

For each instrument k and grid $\{y_{gk}\}$, define smoothed, design-weighted empirical functionals

$$\widehat{F}_k(y_{gk};\boldsymbol{\varphi}_{\cdot k}) = \frac{1}{W_k}\sum_{h\in\mathcal{H}_k} w_h\,\Phi\left(\frac{\log y_{gk}-\log y_{hk}}{h_{\mathrm{shape},k}}\right),$$

$$\widehat{s}_k(y_{gk};\boldsymbol{\varphi}_{\cdot k}) = \frac{1}{T_k}\sum_{h\in\mathcal{H}_k} w_h\,y_{hk}\,\Phi\left(\frac{\log y_{gk}-\log y_{hk}}{h_{\mathrm{caf},k}}\right),$$

with T_k the external instrument total.

Define the parametric PIT and Gaussian scores

$$U_{hk} = F_k(y_{hk}) \quad \text{and} \quad z_{hk} = \Phi^{-1}\big(\,\mathrm{clip}(U_{hk};\epsilon,1-\epsilon)\big),$$

with $\epsilon \in (0,1/2)$ and $\mathrm{clip}(u;a,b) = \min\{\max\{u,a\},\,b\}$. When $U_{hk}\in(\epsilon,1-\epsilon)$,

$$\frac{\partial z_{hk}}{\partial \varphi_{hk}} = \frac{f_k(y_{hk})\,x_{hk}}{\phi(z_{hk})},$$

and the derivative is set to zero when clipped.

Regime-specific dependence is identified using two common value segments across instruments: $[0, 250\mathrm{k})$ and $[250\mathrm{k}, \infty)$. Membership is implemented by smooth windows with break $b = 250{,}000$:

$$m_{hk}^{\mathrm{low}}(b;h_{\mathrm{seg}}) = \Phi\left(\frac{\log b-\log y_{hk}}{h_{\mathrm{seg}}}\right), \qquad m_{hk}^{\mathrm{high}}(b;h_{\mathrm{seg}}) = 1-m_{hk}^{\mathrm{low}}(b;h_{\mathrm{seg}}).$$

Regime-specific correlation targets and dependence loss

For each year and regime $r\in\{\mathrm{low},\mathrm{high}\}$, the target $K\times K$ correlation matrix $R_r^\star$ is estimated on the Gaussian-score scale using survey weights and a pairwise intersection rule. The assembled matrix is projected to the nearest positive definite correlation matrix with minimal adjustment.

Given current adjustments $\boldsymbol{\varphi}$, let $R^{(r)}(\mathbf{z})$ denote the corresponding design-weighted correlation matrix computed on the regime-specific sample. The dependence loss aggregates Frobenius gaps across regimes,

$$L_{\mathrm{dep}}(\boldsymbol{\varphi}) \;=\; \sum_{r\in\{\mathrm{low},\mathrm{high}\}} \big\|R^{(r)}(\mathbf{z}(\boldsymbol{\varphi})) - R_r^\star\big\|_F^2,$$

optionally normalised to keep magnitude $\mathcal{O}(1)$.

Composite objective, constraints, staging, and computation

The multivariate fit objective adds the dependence penalty to the marginal terms:

$$L_{\text{fit}}(\boldsymbol{\varphi}) \;=\; \sum_{k=1}^{K} L_{\text{shape},k}(\boldsymbol{\varphi}_{\cdot k}) \;+\; \lambda_{\text{caf}} \sum_{k=1}^{K} L_{\text{caf},k}(\boldsymbol{\varphi}_{\cdot k}) \;+\; \lambda_{\text{dep}}\, L_{\text{dep}}(\boldsymbol{\varphi}),$$

with box bounds $\varphi_{hk} \in [\underline{\varphi}_k, \overline{\varphi}_k]$ and $\varphi_{hk} = 1$ for $x_{hk} = 0$. Stage 1 minimises L_{fit} with continuation in $(h_{\text{shape},k}, h_{\text{caf},k})$ and cautious ramp-up of λ_{dep}. Stage 2 activates a normalised stabiliser on adjustments, multiplied by an activation gate based on hinge-type penalties on relative increases in the fit components beyond their Stage 1 values.

6.4 Practical implementation

This section describes how the three models are implemented in practice, using the same data inputs and target construction but different calibration criteria. In all cases, weights remain fixed and calibration acts on values via multipliers. Consistently with the notation used in Section 6.3, we use x_{hk} for reported holdings and $y_{hk} = \varphi_{hk} x_{hk}$ for calibrated holdings (with the instrument index suppressed in Model 2).

In terms of computation, runtimes differ markedly across models. Model 1 is fast and runs in a few minutes per wave (roughly 5 minutes), so that all five waves can be processed within about half an hour. By contrast, Models 2 and 3 are substantially more demanding: under the final optimised implementation, Model 2 typically requires 6–8 hours to complete, while Model 3 ranges between 10 and 16 hours depending on the wave, reflecting differences in effective sample size and in the number of active dependence constraints.

All three models share the same computational skeleton: survey weights are held fixed, calibration acts through multiplicative value adjustments under simple box bounds, and optimisation proceeds on common grids that include BSR thresholds (with the same smoothing conventions where distributional terms are used). What changes across models is the target object and, correspondingly, the active constraints or penalties: Model 1 enforces exact bracket totals through a sequential fixed-class QP with reclassification control; Model 2 replaces hard class constraints with a smooth composite marginal loss (shape plus relative CAF) and a staged, gated stabiliser; Model 3 retains the marginal losses of Model 2 and adds a dependence term based on Gaussian-score correlations, again introduced progressively to preserve marginal fit.

Across the three models, the empirical workflow is intentionally modular. The same pre-processing block constructs (*i*) instrument–year targets from BSR/QSA inputs and (*ii*) the evaluation objects used in optimisation (class edges, grids, and smoothing conventions). The only model-specific component is the optimisation block, which differs solely in the objective and constraint structure described in the

Theory section.

Operationally, each model follows the same pipeline: (*i*) build instrument–year targets; (*ii*) initialise multipliers at the total-matching scale and enforce common box bounds; (*iii*) run the model-specific optimiser (QP for Model 1; gradient-based optimisation for Models 2–3) under continuation in smoothing and penalties; (*iv*) record diagnostics and retain the best iterate when the criterion is non-monotone; and (*v*) export calibrated microdata and derived distributional summaries. The subsections below therefore emphasise what changes from one model to the next – i.e. which additional blocks are activated on top of the shared backbone.

6.4.1 Common inputs and target construction

Targets replicate BSR range-sum objects. For each instrument k and year, bracket shares are computed as size-biased truncated first moments from the target marginal $f_k(\cdot;\widehat{\boldsymbol{\theta}}_k)$:

$$s_{kj} \;=\; \frac{\int_{a_{k,j-1}}^{a_{k,j}} t\, f_k(t;\widehat{\boldsymbol{\theta}}_k)\, dt}{\mu_{1k}(\widehat{\boldsymbol{\theta}}_k)}, \qquad \mu_{1k}(\widehat{\boldsymbol{\theta}}_k) \;=\; \int_0^{\infty} t\, f_k(t;\widehat{\boldsymbol{\theta}}_k)\, dt, \qquad T_{kj} \;=\; s_{kj}\, T_k,$$

with T_k the instrument-level macro aggregate. Zeros remain fixed at $\varphi_{hk} = 1$, and class assignment (when used) follows the left-closed rule $a_{k,j-1} \le y_{hk} < a_{k,j}$ for the currently evaluated value.

6.4.2 Model 1 implementation: bounded class-total calibration

Model 1 is the baseline implementation block: it enforces BSR class totals exactly, using frozen class membership within each iteration. The practical loop therefore alternates between a bounded equality-constrained QP solve and an ex-post projection for reclassifiers, with selection by the class-gap diagnostic when threshold crossings prevent monotone improvement.

Initial multipliers are set to $\varphi_{hk}^{(0)} = 1$. At iteration t, classes are frozen using $\mathbf{y}^{(t-1)}$; for each k, solve the fixed-class QP (6.3) under baseline bounds $[\underline{\varphi}_k, \overline{\varphi}_k] = [0.5, 10]$, producing $\boldsymbol{\varphi}_{\cdot k}^{(t)}$ and $\mathbf{y}^{(t)}$. Immediately after the solve, apply ex-post projection for reclassifiers to $[0.95, 1.05]$. Record $G^{(t)}$ in (6.6) and retain the best iterate. Reported outputs are single-instrument calibrations (deposits, listed shares, bonds, fund shares), to isolate the impact of enforcing BSR-style class totals.

Model 1 is deliberately conservative: it reproduces the externally mandated BSR class totals exactly (up to the class reallocation induced by threshold crossings) while preventing implausibly large household-level adjustments through explicit bounds and a reclassification trust region. However, because the constraints only pin down truncated sums by class, the within-class micro shape remains underidentified and solutions may accumulate around cut-offs, even when the class totals are matched tightly. This motivates Model 2, which uses the BSR-implied parametric margin to target distributional shape and the cumulative-amount path directly, rather than

relying on class totals alone.

6.4.3 Model 2 implementation: composite distributional calibration

Model 2 preserves the same backbone inputs (fixed survey weights, external instrument total, and BSR-based target marginal) but replaces hard class equalities with a smooth composite loss evaluated on a common grid. In implementation this is the key shift: the calibration step becomes gradient-based optimisation with analytical derivatives and a two-stage schedule (fit first, then shrink within a near-tie region), while all upstream target construction and downstream diagnostics remain unchanged.

For each instrument, build the evaluation grid as quantiles of the target F plus BSR thresholds, with endpoint up-weighting. Step 1 minimises (6.8) under box bounds $\varphi_h \in [\underline{\varphi}, \overline{\varphi}]$, initialised at the total-matching scale; λ_{caf} is chosen by a small grid search to balance CAF alignment and CDF fit. Step 2 activates stabilisation only within tight tolerances around Step 1 losses, using a smooth gate so shrinkage operates as a tie-breaker among near-equivalent fits.

Diagnostics include density overlays, CAF overlays (with thresholds marked), total residuals, and φ_h–x_h plots.

6.4.4 Model 3 implementation: multivariate dependence-aware calibration

Model 3 proceeds in two blocks. First, estimate regime-specific correlation targets $R^{\star}_{\text{low}}$ and $R^{\star}_{\text{high}}$ on Gaussian scores using weights and the pairwise intersection rule, then project targets to positive definite correlation matrices. Second, calibrate by Stage 1 minimisation of the composite fit objective (margins plus dependence), ramping up λ_{dep} only after marginal discrepancies are small, followed by Stage 2 stabilisation under a gated criterion that preserves Stage 1 achievements in all components.

Diagnostics add before/after regime-specific correlations on Gaussian scores, regime-support checks, and sensitivity of results to clipping ϵ and segmentation bandwidth h_{seg}.

6.5 Results

This section reports real-data results for the three calibration layers developed in this chapter: the baseline class-penalised calibration (Model 1), the distribution-aware univariate calibration (Model 2), and the multivariate, dependence-aware calibration (Model 3). The methods are evaluated in the same reconciled setting: BSR totals are reconciled to QSA, so aggregate targets refer to QSA outstanding amounts, while bracket-level comparisons refer to BSR class totals. Throughout, we keep survey weights fixed and adjust values through multiplicative factors; to avoid confusion

with sampling weights, we refer to these factors as adjustment parameters.

To keep notation consistent across models and to facilitate like-for-like comparisons, we use the following shorthand for microdata states along the DWA pipeline:

- HFCS_{std}: adjusted microdata under the standard Italian DWA procedure up to, but excluding, proportional allocation;
- $\text{HFCS}_{\text{class}}$: adjusted microdata after Model 1 (class-consistent calibration) up to, but excluding, proportional allocation;
- $\text{HFCS}_{\text{dist}}$: adjusted microdata after Model 2 (distribution-aware calibration) up to, but excluding, proportional allocation;
- $\text{HFCS}_{\text{multi}}$: adjusted microdata after Model 3 (multivariate calibration) up to, but excluding, proportional allocation;
- superscript fin denotes the corresponding full-pipeline output after proportional allocation and the subsequent quarterly interpolation/extrapolation.

All calibrations are deterministic conditional on inputs and a fixed seed. In the absence of a household-level ground truth, attaching sampling-style uncertainty to calibrated values would be arbitrary and potentially misleading; we therefore report point estimates only and focus on realised impacts on (*i*) aggregate coverage to QSA, (*ii*) bracket alignment to BSR, and (*iii*) micro-level distributional and dependence diagnostics.

The empirical scope is deposits and three securities components held on custody accounts (listed shares, bonds, and investment fund shares). Deposits are calibrated at the household level (consistent with the DWA unit of analysis), while bracket-level comparisons for deposits and securities are reconstructed on the original BSR taxonomies (deposit-account size and securities-account size), using the mappings described earlier in the thesis. For comparability across time, we use the 2022 break vectors for both deposits and securities in all years.

6.5.1 Model 1: baseline class-consistent calibration

We begin with real-data results for the baseline class-consistent calibration introduced in Neri et al. (2023) and extended here to the BSR-coherent household marginals recovered earlier (deposits in Chapter 4; listed shares, bonds, and investment fund shares in Chapter 5). The resulting estimates form the benchmark against which Models 2 and 3 are evaluated.

The calibration targets class totals by BSR break in each instrument-year cell. Given the data structure, we do not impose direct control on the within-class microdistribution: the hard constraints are range-sum totals, and the decision variables are household-level adjustment parameters. Pipeline effects are therefore evaluated end-to-end under the full DWA procedure (including proportional allocation): the standard pipeline is the current Italian DWA method; the updated pipeline differs

only in the pre-proportional step, where the present calibration replaces the account-based deposits approach while keeping proportional allocation unchanged.

Since Chapters 4–5 recover the marginal distributions, the class totals needed for calibration are obtained by scaling the QSA instrument totals by the model-implied bracket shares. For instrument k and bracket j,

$$\widetilde{T}_{kj} \;=\; \widehat{s}_{kj}(\hat{\boldsymbol{\theta}}_k)\, T_k,$$

where T_k is the QSA outstanding amount and $\widehat{s}_{kj}(\hat{\boldsymbol{\theta}}_k)$ is the share implied by the fitted parameters $\hat{\boldsymbol{\theta}}_k$ on the common break vector.

At the aggregate level, coverage with respect to QSA improves materially relative to pre-calibration, and class totals align closely with the administrative values. Totals by instrument-year match, to a high degree, the external aggregates used for reconciliation. The method also attenuates instances of over-coverage observed prior to proportional allocation (e.g., listed shares in 2022), largely attributable to managed-account adjustments. Figure 6.1 summarises coverage improvements with side-by-side bars for pre- and post-calibration totals by instrument and year.

We then compare $\text{HFCS}_{\text{class}}$ with original-scope BSR totals on the BSR bracket taxonomy – deposits by deposit-account size and securities by securities-account size. In Appendix A, Figure A.7 reports bracket-level totals from $\text{HFCS}_{\text{class}}$ alongside BSR for selected years and instruments. The comparison indicates a close match in most combinations, with remaining discrepancies concentrated at specific breaks. These failures to converge are consistent with sample-composition constraints: the realised microdata may not support allocating the required mass to every bracket without generating spurious participation. This motivates the subsequent emphasis on zero-reporting detection and on smoother distributional targets.

To benchmark the deposits component against the account-based calibration in the standard pipeline, we compare post-calibration distributions for $\text{HFCS}_{\text{class}}$ and HFCS_{std} at the instrument level over time (Figure 6.2). As expected, the standard class-penalisation approach tends to disregard micro-level distribution when adjusting to target totals, accumulating observations at boundaries; this is particularly visible in wave 4 (2020), where the standard approach yields a stepped, boundary-peaked distribution with several pronounced modes. This limitation is one of the main motivations for Models 2 and 3.

We then evaluate account-level alignment against BSR deposit-account bracket totals by reconstructing account-level aggregates from household-level deposits (see Chapter 4). Figure 6.3 plots HFCS_{std} and $\text{HFCS}_{\text{class}}$ bars with the BSR bracket totals overlaid as a reference line. While the standard calibration is tailored to accounts, the household-level procedure performs similarly – and in several brackets slightly better – after account-level reconstruction, consistent with household-level co-movement of multiple accounts and with the dilution effects discussed in the deposits chapter.

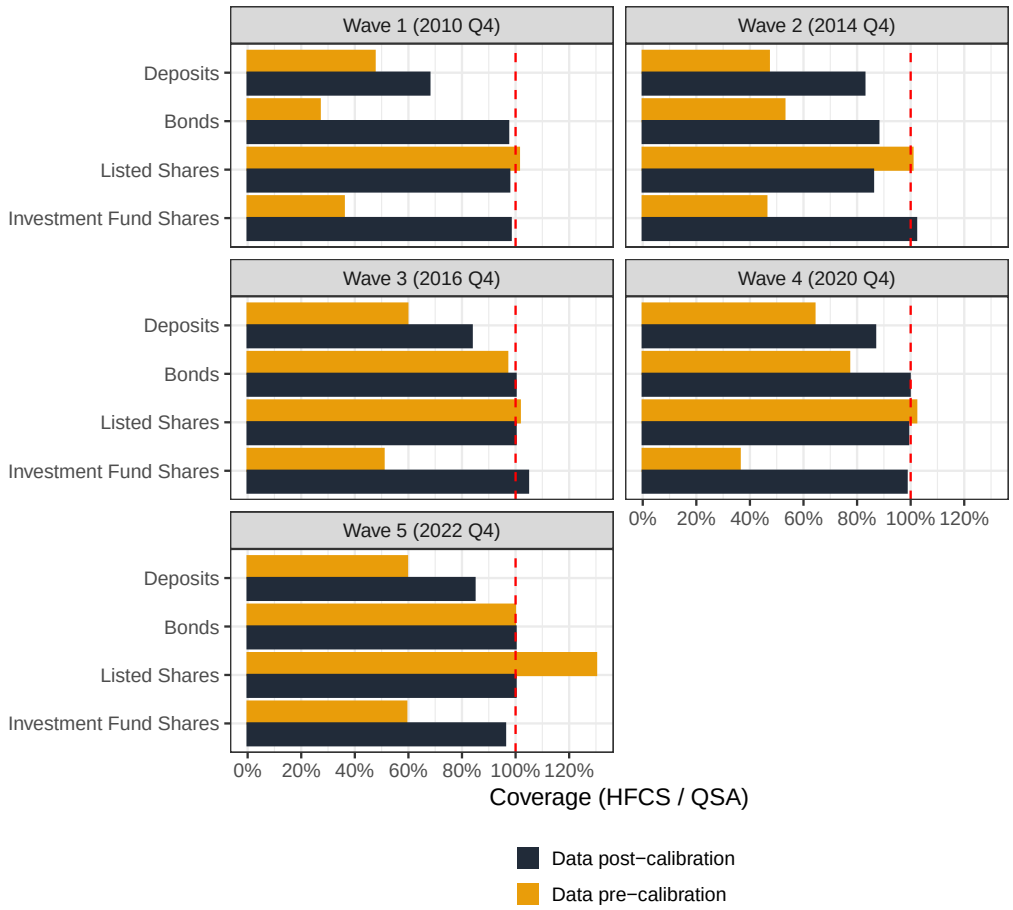

Figure 6.1. Instrument-level coverage ratio before and after calibration, by year.
Notes: For each instrument, bars report totals from the adjusted microdata before and after the baseline calibration. Coverage is computed as the ratio of calibrated totals to QSA totals, in per cent.

At the aggregate level, we assess how these changes translate into portfolio composition (for the instruments analysed in this thesis) by net-wealth group. The adjustment further highlights the distinct composition of the top net-wealth groups, which diversify more of their holdings, while lower net-wealth groups tend to accumulate resources in deposits. In Appendix A, Figure A.8 visualises instrument shares within the household financial portfolio by year and net-wealth group.To quantify end-of-pipeline implications – and to provide a benchmark for Models 2 and 3 – we compare $\text{HFCS}^{\text{fin}}_{\text{std}}$ with $\text{HFCS}^{\text{fin}}_{\text{class}}$ in terms of inequality measures. Figure 6.4 plots the Gini for net wealth over time under the two pipelines.

The new class-aware approach lowers the estimates relative to the standard one, suggesting that boundary accumulation may play a key role. This consideration motivated the development of the following models and is also why this approach was applied in production only to deposits: given their less pronounced tail behaviour, accumulation at the lower bound of the last break is likely to be negligible.

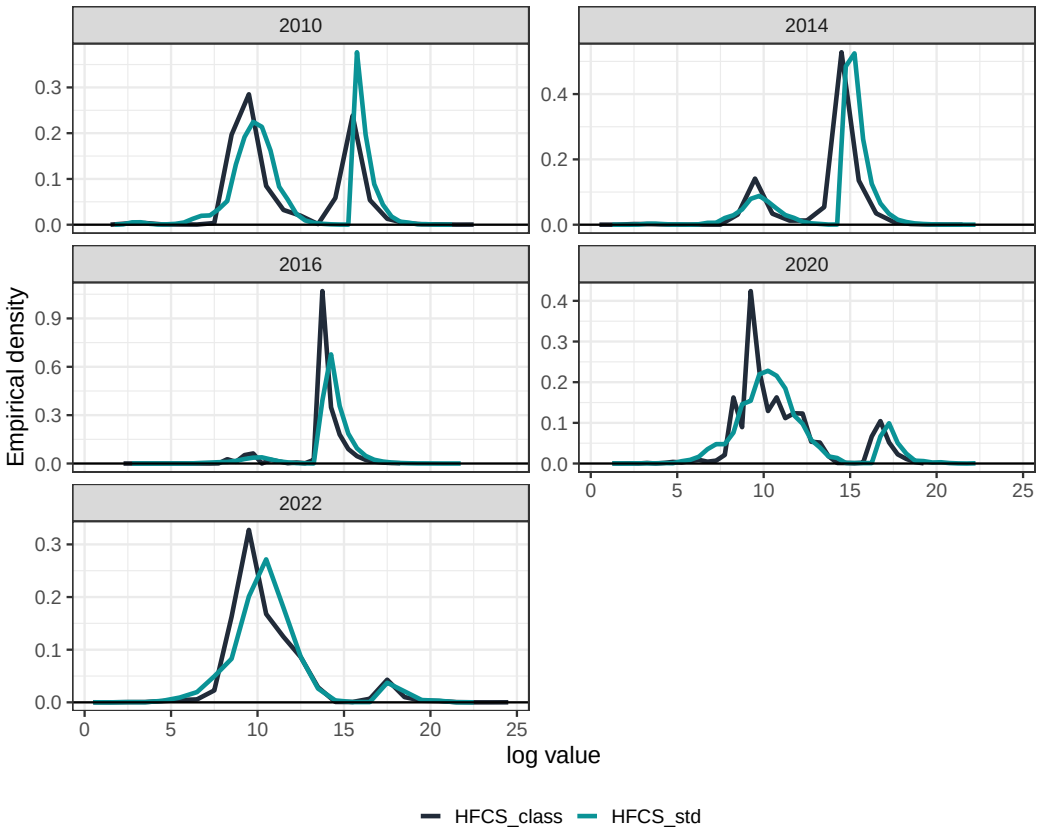

Figure 6.2. Deposits – post-calibration distribution: $\text{HFCS}_{\text{class}}$ vs HFCS_{std}.
Notes: Curves show the empirical PDF estimates for each method in each year.

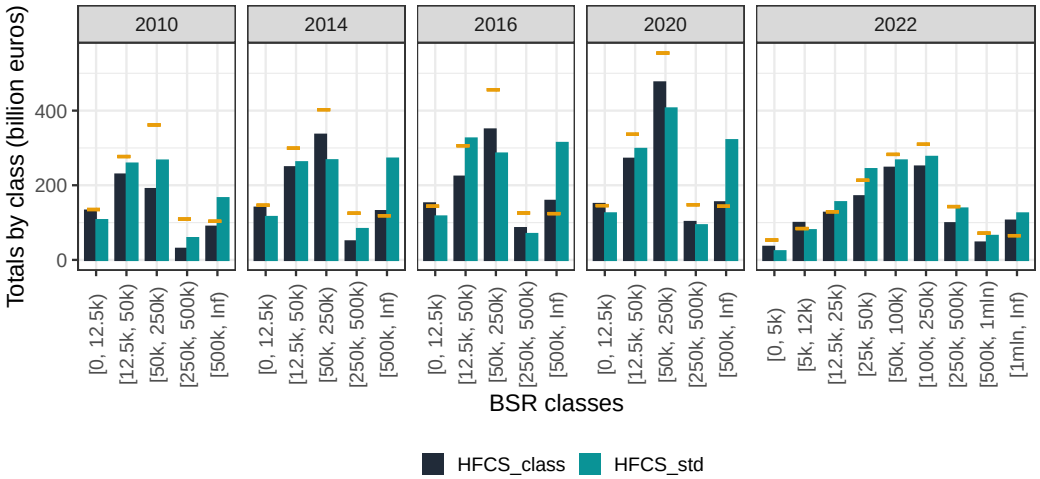

Figure 6.3. Deposits – account-level bracket totals: HFCS_{std} and $\text{HFCS}_{\text{class}}$ vs BSR.
Notes: For each deposit-account bracket, two bars (standard and new) are shown with a reference line at the BSR total. Brackets follow the deposit-account taxonomy used for BSR dissemination.

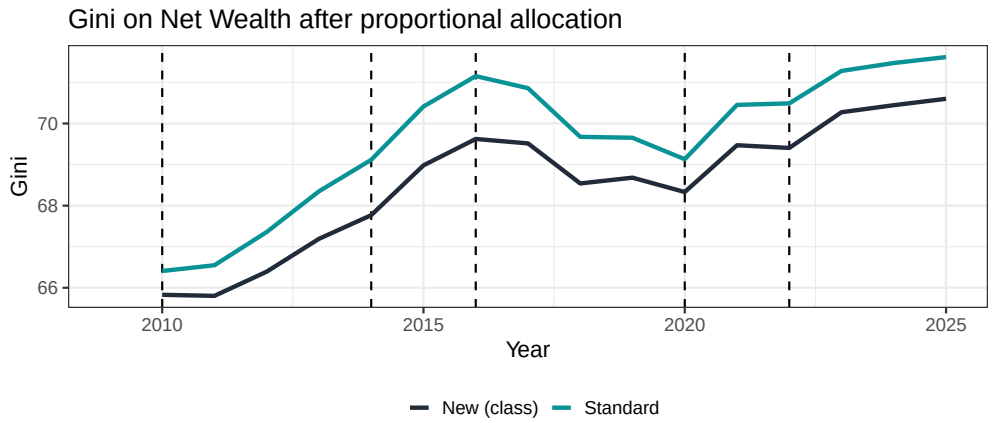

Figure 6.4. Gini on net wealth – standard vs new full pipelines ($\text{HFCS}^{\text{fin}}_{\text{std}}$ and $\text{HFCS}^{\text{fin}}_{\text{class}}$).
Notes: Vertical dashed lines indicate survey waves.

In Appendix A, Figure A.9 extends the comparison to median wealth, top shares, the bottom-50 per cent share, and the share of households above € 1 million. Overall, the class-consistent calibration tends to reduce measured concentration. As discussed below, this is consistent with micro-level boundary piling and with the absence of explicit control on within-class shapes and cross-instrument dependence.

Two micro-level pitfalls are also documented. First, iteration-by-iteration adjustments can cumulate onto a small set of households because each step conditions only on values at t and $t-1$. To prevent excessive leverage, we impose a cap on the cumulative adjustment parameter per household-instrument; the baseline cap is 15. Table 6.1 reports dispersion diagnostics.

In Appendix A, Figure A.10 shows the implied distribution (deposits as example). Second, because the method does not target the final microdistribution within bins, adjusted values tend to gravitate towards class boundaries;in Appendix A, Figure A.11 illustrates boundary-aligned mass in 2022.

6.5.2 Model 2: distribution-aware calibration with composite loss

We next report results for the distribution-aware calibration proposed in this chapter. This univariate version is the building block for the multivariate calibration in Model 3 and provides preliminary evidence on the gains from targeting distributional shapes directly. The method addresses two limitations of Model 1 – accumulation of adjustment parameters across iterations and piling of values at bracket boundaries – by optimising a composite loss on the full marginal distribution and on cumulative amounts. We begin with aggregate coverage to QSA. Figure 6.5 shows pre- and post-calibration totals by instrument and year, and Figure 6.6 compares coverage under $\text{HFCS}_{\text{dist}}$ and $\text{HFCS}_{\text{class}}$.

Table 6.1. Adjustment-weight concentration diagnostics (baseline cap = 15).

Instrument	**Year**	φ_{median}	φ_{p99}	$\varphi_{\max}$
Deposits	2010	1.07	7.14	23.73
Deposits	2014	1.65	15.27	30.88
Deposits	2016	1.40	8.80	15.93
Deposits	2020	1.25	5.62	8.30
Deposits	2022	1.29	4.24	7.43
Bonds	2010	2.11	7.24	9.96
Bonds	2014	2.82	6.92	7.27
Bonds	2016	1.05	1.45	6.45
Bonds	2020	0.93	1.23	8.58
Bonds	2022	1.10	2.13	2.65
Listed Shares	2010	0.81	0.99	2.68
Listed Shares	2014	1.52	7.81	8.81
Listed Shares	2016	0.38	0.94	10.00
Listed Shares	2020	0.67	1.00	3.88
Listed Shares	2022	1.26	5.15	6.53
Investment Funds	2010	0.88	0.99	37.23
Investment Funds	2014	5.05	22.29	25.64
Investment Funds	2016	2.14	6.57	10.00
Investment Funds	2020	1.17	1.87	9.98
Investment Funds	2022	1.49	3.27	3.97

Notes: Weighted diagnostics by instrument and year. "At cap" is the share of households with cumulative adjustment parameter at or above the cap.

Coverage under $\text{HFCS}_{\text{dist}}$ is broadly comparable to the class-consistent benchmark, meeting the design goal of improving the microdistribution without materially degrading totals. In some instances totals exceed QSA (notably 2014 for securities), which may reflect wave-specific comparability issues between HFCS and BSR and interacts with the soft-threshold behaviour discussed earlier. Where pre-calibration gaps are largely removed by $\text{HFCS}_{\text{dist}}$, proportional allocation subsequently operates on negligible residues.

A distinguishing feature of Model 2 is explicit distributional alignment. To keep the presentation compact (and to mirror Model 3), we report PDF overlays as two full-page panels; CDF and CAF diagnostics are deferred to the appendix and are provided in the replication package. In Appendix A, Figures A.12–A.13 overlay kernel-smoothed PDFs of adjusted values ($\text{HFCS}_{\text{dist}}$) on the BSR-implied targets, alongside pre-calibration curves. Some HFCS curves are visibly wiggly relative to the smooth parametric targets; this is expected when the effective sample size for strictly positive holdings is limited and reflects the variance of weighted kernel estimators

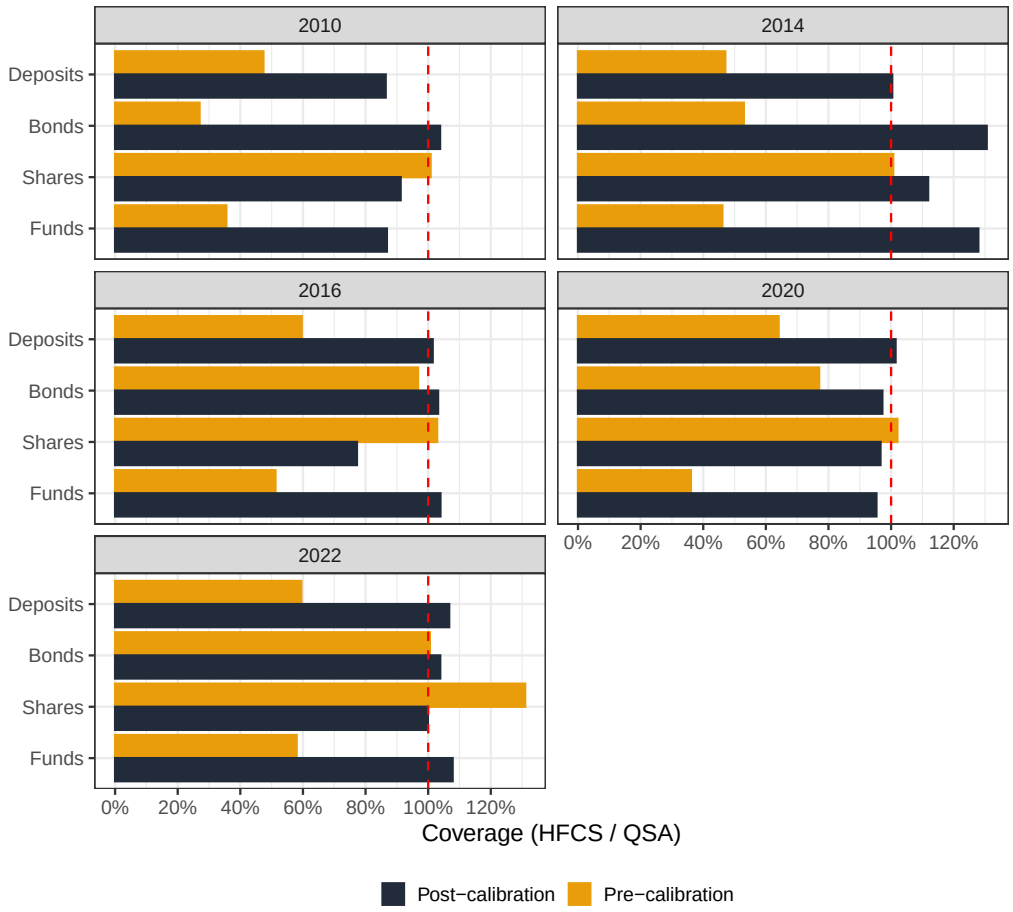

Figure 6.5. Instrument-level totals before and after the distribution-aware calibration, by year.
Notes: For each instrument, bars report totals from the adjusted microdata before and after $\text{HFCS}_{\text{dist}}$. Coverage is the ratio to QSA totals (per cent).

rather than instability of the calibration itself.

We summarise distributional differences across models using quantiles and inequality indicators prior to proportional allocation. Table 6.2 reports selected quantiles by model and instrument for 2022 (excluding zeros where participation is sparse), and Table 6.3 reports the Gini and top-decile share for financial net wealth. Relative to HFCS_{std} and $\text{HFCS}_{\text{dist}}$, the class-consistent baseline tends to compress the top tail by piling mass near the lower bound of the open-ended top bin. Conversely, a purely marginal distributional calibration can move mass into unrelated upper tails across instruments, potentially overstating concentration. These patterns motivate the dependence-aware extension in Model 3.

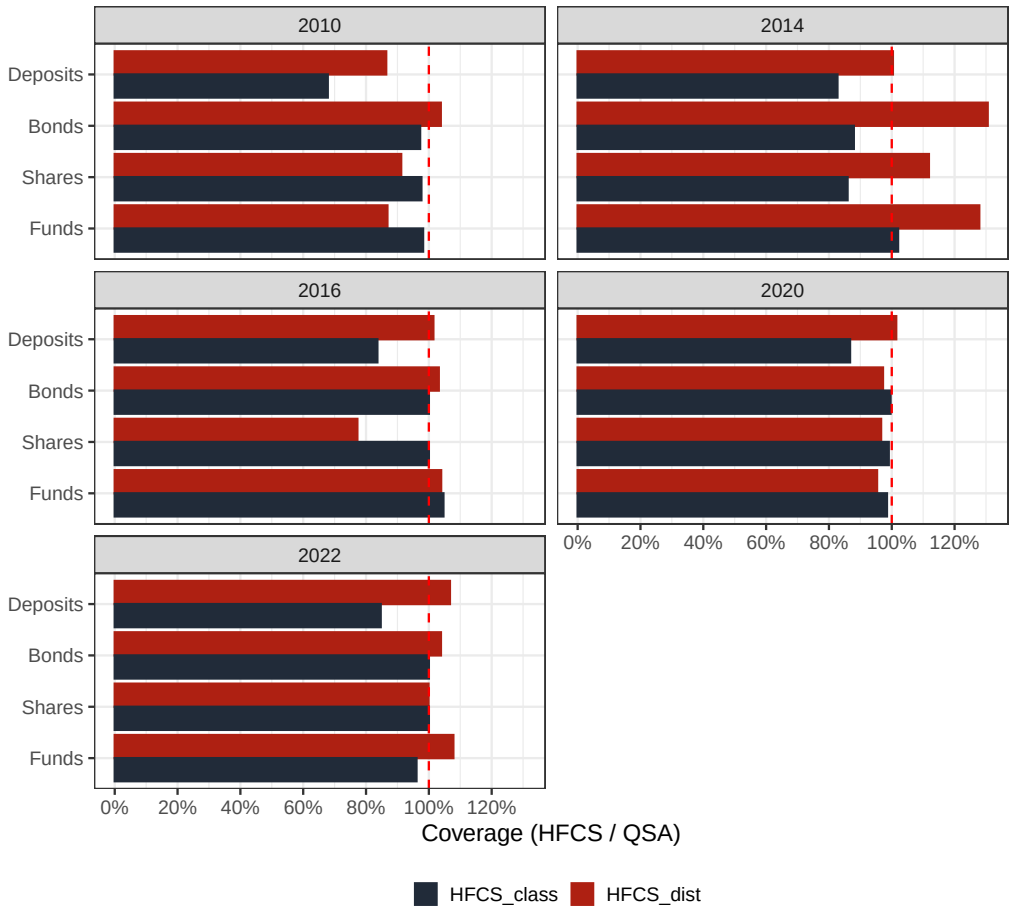

Figure 6.6. Coverage comparison: distribution-aware vs class-consistent calibration.
Notes: For each instrument and year, bars show coverage (per cent) to QSA under $\text{HFCS}_{\text{dist}}$ and $\text{HFCS}_{\text{class}}$. Axes and legend are harmonised across years.

We then compare portfolio composition across models. Figure 6.7 shows instrument shares within the financial portfolio by net-wealth group for 2022 (for readability), complementing the previous tables and Figure A.8.

Figure 6.8 plots the Gini series after proportional allocation and quarterly interpolation/extrapolation for net wealth and financial net wealth under $\text{HFCS}^{\text{fin}}_{\text{std}}$, $\text{HFCS}^{\text{fin}}_{\text{class}}$, and $\text{HFCS}^{\text{fin}}_{\text{dist}}$. The figure shows that the class-consistent calibration consistently lowers the Gini index in all periods relative to the standard approach, whereas the distribution-aware calibration tends to increase it.

This suggests a marked difference between the class-consistent and distribution-aware approaches, despite the latter being, in principle, a "continuous" version of the former. One explanation may relate to the boundary accumulation induced by the class-consistent procedure. By construction, it favours low, uniform adjustment values, which can lead to mass accumulation at class boundaries.

As a result, the top tail may become overly under-represented, which – particularly

Table 6.2. Selected quantiles by model, year, and instrument.

Model	Year	Instrument	p1	p10	p50	p90	p99
$HFCS_{class}$	2022	Deposits	428	3,789	14,057	99,296	463,060
$HFCS_{dist}$	2022	Deposits	267	2,624	19,831	107,918	733,231
$HFCS_{std}$	2022	Deposits	274	2,477	20,001	145,129	474,147
$HFCS_{class}$	2022	Bonds	253	6,008	34,388	172,616	614,915
$HFCS_{dist}$	2022	Bonds	342	5,031	27,601	58,503	1,271,110
$HFCS_{std}$	2022	Bonds	32	713	8,914	59,640	1,704,288
$HFCS_{class}$	2022	Listed Shares	59	937	16,129	190,209	959,732
$HFCS_{dist}$	2022	Listed Shares	58	1,599	17,340	67,111	2,324,526
$HFCS_{std}$	2022	Listed Shares	48	954	16,173	97,536	3,090,592
$HFCS_{class}$	2022	Investment Funds	1,618	10,989	83,372	452,625	3,272,888
$HFCS_{dist}$	2022	Investment Funds	308	20,332	103,932	412,540	2,645,425
$HFCS_{std}$	2022	Investment Funds	2,200	17,886	90,125	500,813	2,504,064

Notes: Models are $HFCS_{std}$, $HFCS_{class}$, and $HFCS_{dist}$. Columns report p1, p10, p50, p90, and p99. Large euro figures are integers. Zeros are excluded where zero mass dominates participation.

Table 6.3. Inequality summary prior to proportional allocation: financial net wealth.

Model	2010		2014		2016		2020		2022	
	Gini	Top 10	Gini	Top 10	Gini	Top 10	Gini	Top 10	Gini	Top 10
$HFCS_{class}$	81.91	76.45	77.58	66.25	81.85	76.88	81.59	74.55	78.15	68.48
$HFCS_{dist}$	82.63	75.63	84.66	76.51	84.87	79.28	84.56	80.58	80.26	72.70
$HFCS_{std}$	80.35	72.44	83.07	74.60	82.59	77.40	79.45	71.44	78.96	69.91

Notes: Gini (per cent) and top-10 per cent share refer to financial net wealth constructed from the three securities instruments.

for markedly skewed instruments such as bonds, listed shares, and investment fund shares – may translate into a lower estimate of inequality in net wealth.

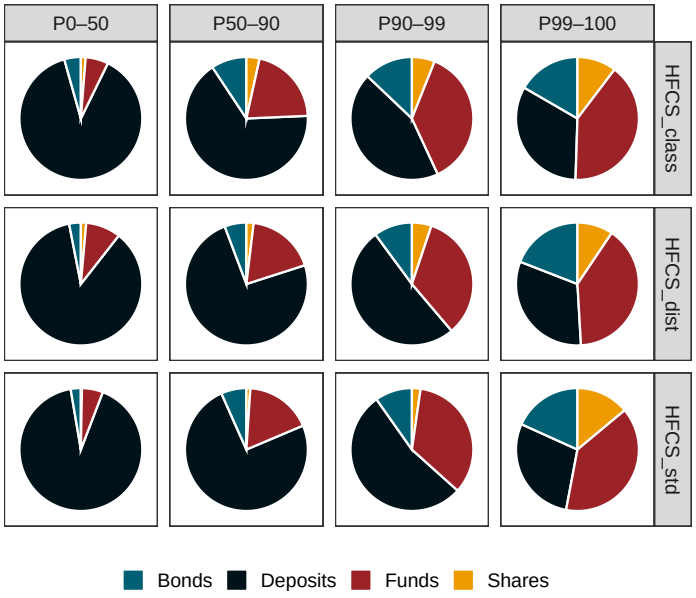

Figure 6.7. Household financial portfolio composition by net-wealth group (2022).
Notes: Pie charts show the shares of deposits, listed shares, bonds, and investment fund shares. Facets place net-wealth groups in columns. Models: HFCS_{std}, $\text{HFCS}_{\text{class}}$, $\text{HFCS}_{\text{dist}}$.

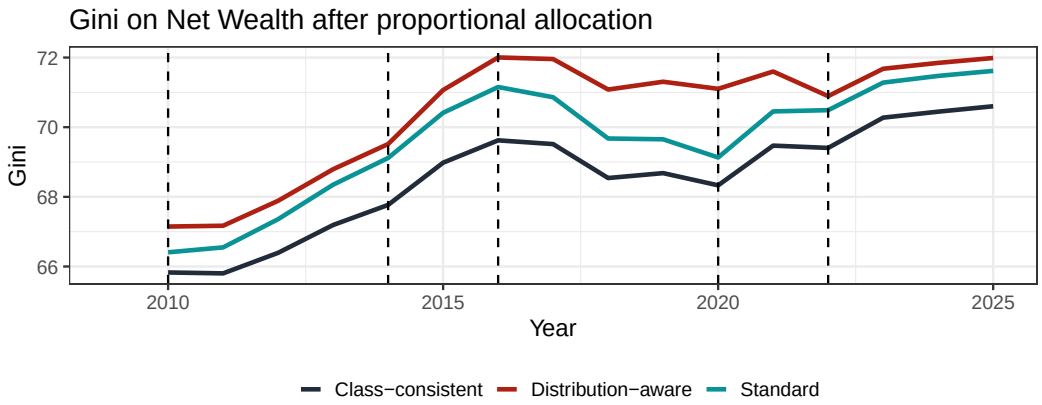

Figure 6.8. Gini of net wealth – standard vs class-consistent vs distribution-aware full pipelines.
Notes: Year-by-year Gini for overall and financial net wealth under $\text{HFCS}^{\text{fin}}_{\text{std}}$, $\text{HFCS}^{\text{fin}}_{\text{class}}$, and $\text{HFCS}^{\text{fin}}_{\text{dist}}$. Vertical dashed lines may indicate survey waves.

In Appendix A, Figure A.14 extends the comparison to additional inequality indicators. The distribution-aware method can be tail-enhancing relative to the other pipelines; without explicit dependence control, marginal adjustments may over-concentrate joint tails for some portfolios, motivating Model 3.

Finally, we assess micro-level credibility relative to Model 1. Figure 6.9 contrasts adjustment-parameter distributions under $\text{HFCS}_{\text{dist}}$ and $\text{HFCS}_{\text{class}}$; Table 6.4 reports dispersion diagnostics; and, in Appendix A, Figure A.15 reports Spearman pre-post

correlations as a deformation diagnostic. Overall, distribution-aware adjustments curb boundary piling and, in several cases, preserve rank ordering more tightly than the class-consistent benchmark.

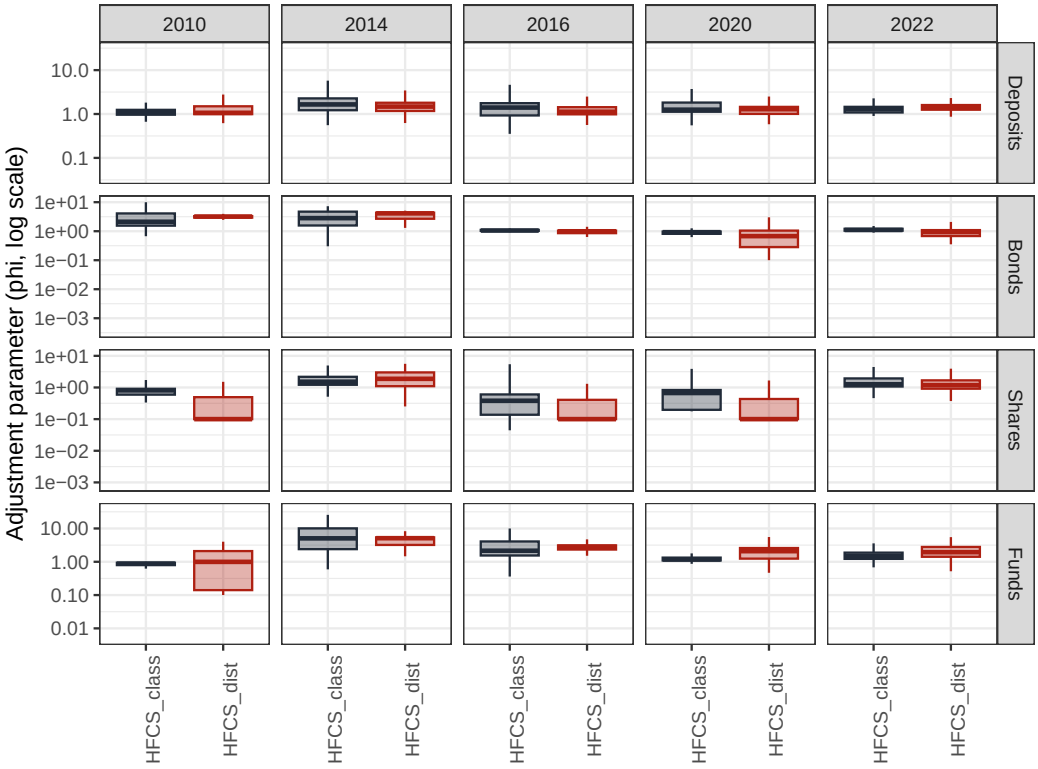

Figure 6.9. Adjustment parameters under $\text{HFCS}_{\text{dist}}$ vs $\text{HFCS}_{\text{class}}$.
Notes: Comparative distributions by instrument and year. Medians and extreme quantiles shown in-panel.

Table 6.4. Adjustment-parameter dispersion diagnostics (cap baseline = 15).

Model	**Year**	**Instrument**	median_{φ}	μ_{φ}	p99_{φ}	$\max_{\varphi}$	σ_{φ}
$\text{HFCS}_{\text{class}}$	2022	Deposits	1.29	1.43	4.24	7.43	0.65
$\text{HFCS}_{\text{dist}}$	2022	Deposits	1.46	1.45	3.25	15.00	0.59
$\text{HFCS}_{\text{class}}$	2022	Bonds	1.10	1.17	2.13	2.65	0.21
$\text{HFCS}_{\text{dist}}$	2022	Bonds	0.94	0.94	3.00	4.63	0.50
$\text{HFCS}_{\text{class}}$	2022	Shares	1.26	1.65	5.15	6.53	0.86
$\text{HFCS}_{\text{dist}}$	2022	Shares	1.18	1.38	3.90	5.41	0.69
$\text{HFCS}_{\text{class}}$	2022	Funds	1.49	1.62	3.27	3.97	0.48
$\text{HFCS}_{\text{dist}}$	2022	Funds	1.94	2.15	5.00	5.50	1.10

Notes: Weighted diagnostics by instrument and year.

In sum, Model 2 improves marginal alignment while maintaining aggregate coverage comparable to the benchmark. Its tail behaviour and pipeline indicators underscore the need for an explicit dependence layer – the focus of Model 3 – to mediate cross-corrections and stabilise joint tails across instruments.

6.5.3 Model 3: multivariate calibration with dependence control

We conclude with real-data results for the multivariate calibration – the final method of the thesis. The objective is to produce HFCS microdata that respect aggregate targets, match marginal distributions, and preserve economically meaningful cross-instrument co-movement. This multivariate layer mediates cross-corrections across instruments and controls the implied dependence structure, a key missing element in current DWA implementations.

As in Models 1 and 2, we keep survey weights fixed and adjust values; zeros remain fixed per instrument. We acknowledge a natural trade-off: by aligning both micro distributions and cross-instrument dependence, we may tolerate small shortfalls in aggregate coverage relative to a method tailored to match totals exactly (Model 1). Our position is that gains in distributional credibility and dependence coherence outweigh small coverage discrepancies, provided the residual gap is sufficiently small that proportional allocation does not undo the improvements.

We begin with aggregate coverage to QSA and bracket-level alignment to BSR. Figure 6.10 displays pre- and post-$\text{HFCS}_{\text{multi}}$ totals by instrument and year. In Appendix A, Figure A.18 complements this analysis, comparing $\text{HFCS}_{\text{multi}}$ bracket totals with BSR. Together, these figures show that the multivariate procedure recovers aggregates with precision comparable to Models 1 and 2 while retaining bracket-level alignment. The 2014 over-coverage observed for securities persists but is attenuated relative to the purely marginal approach.

We then assess marginal alignment with a focus on density shape. In Appendix B, Figures A.16–A.17 overlay kernel-smoothed PDFs of adjusted values ($\text{HFCS}_{\text{multi}}$) on the BSR-implied targets, alongside pre-calibration curves for reference. Relative to Model 2, marginal fit is preserved and in several cases sharpened, while remaining discrepancies are small on the plotting scale used.

We then compare dissemination-relevant summaries across models. Table 6.5 reports selected quantiles by model and instrument in 2022 (excluding zeros where appropriate), and Table 6.6 reports Gini and top-decile shares for financial net wealth prior to proportional allocation. Relative to Model 2, the multivariate procedure delivers a milder concentration profile, consistent with dependence control mediating cross-instrument tail co-movement.

Next, we focus on portfolio composition at the household level across models. Figure 6.11 shows instrument shares within the financial portfolio by year and net-wealth group under HFCS_{std}, $\text{HFCS}_{\text{dist}}$, and $\text{HFCS}_{\text{multi}}$.

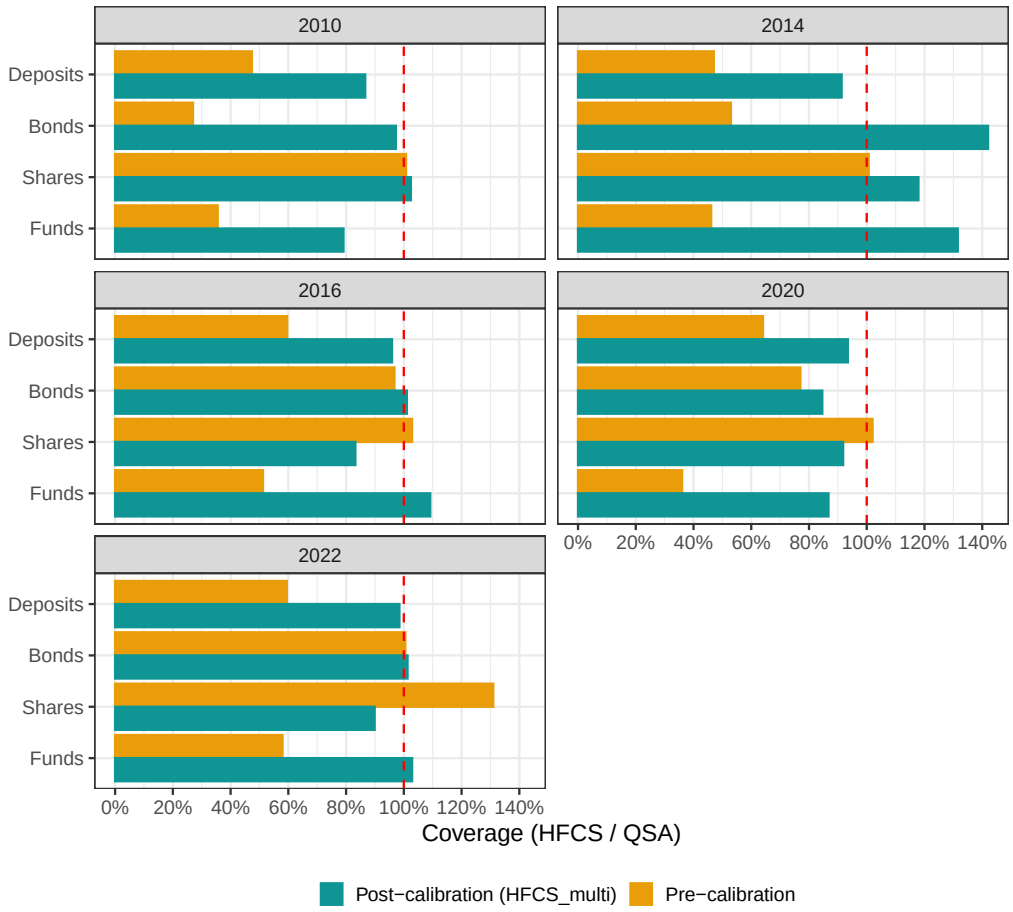

Figure 6.10. Instrument-level totals before and after the multivariate calibration, by year. Notes: For each instrument, bars report totals from the adjusted microdata before and after $\text{HFCS}_{\text{multi}}$. Coverage is the ratio to QSA totals, in per cent. Axes and legend are harmonised across years.

We conclude by showing the impact on the main inequality indicators considered in this thesis. In particular, Figure 6.12 compares the Gini index computed from the DWA microdata after full reconciliation (post proportional allocation), under the different models defined in this chapter.

Table 6.5. Selected quantiles by model, and instrument in 2022.

Model	Year	Instrument	p1	p10	p50	p90	p99
$HFCS_{dist}$	2022	Deposits	267	2,624	19,831	107,918	733,231
$HFCS_{multi}$	2022	Deposits	241	2,346	23,585	115,223	445,253
$HFCS_{std}$	2022	Deposits	274	2,477	20,001	145,129	474,147
$HFCS_{dist}$	2022	Bonds	342	5,031	27,601	58,503	1,271,110
$HFCS_{multi}$	2022	Bonds	350	5,904	22,776	87,721	1,094,554
$HFCS_{std}$	2022	Bonds	100	5,986	35,269	119,713	872,232
$HFCS_{dist}$	2022	Shares	58	1,599	17,340	67,111	2,324,526
$HFCS_{multi}$	2022	Shares	48	954	16,173	97,536	3,090,592
$HFCS_{std}$	2022	Shares	32	713	8,914	59,640	1,704,288
$HFCS_{dist}$	2022	Funds	308	20,332	103,932	412,540	2,645,425
$HFCS_{multi}$	2022	Funds	2,730	17,471	92,029	443,623	2,632,781
$HFCS_{std}$	2022	Funds	2,200	17,886	90,125	500,813	2,504,064

Notes: Models are $HFCS_{std}$, $HFCS_{dist}$, and $HFCS_{multi}$. Columns report p1, p10, p50, p90, and p99. Large euro figures are integers. Zeros are excluded where zero mass dominates participation.

Table 6.6. Inequality summary prior to proportional allocation: financial net wealth.

Model	2010		2014		2016		2020		2022	
	Gini	Top 10	Gini	Top 10	Gini	Top 10	Gini	Top 10	Gini	Top 10
$HFCS_{dist}$	82.63	75.63	84.66	76.51	84.87	79.28	84.56	80.58	80.26	72.70
$HFCS_{multi}$	81.67	73.79	83.39	74.53	82.33	76.05	82.44	75.42	76.72	66.78
$HFCS_{std}$	80.35	72.44	83.07	74.60	82.59	77.40	79.45	71.44	78.96	69.91

Notes: Gini (per cent) and top-10 per cent share refer to financial net wealth constructed from the three securities instruments.

Differences are relevant and, for some periods, exceed 1 percentage point. Moreover, the distribution-aware calibration seems to systematically favour richer portfolios relative to the standard proportional allocation, suggesting a marked difference in the estimated gaps across households. This tendency, however, is more than compensated by the multivariate calibration, which allows for a more careful allocation of the remaining gap by exploiting not only where under-coverage concentrates along the distribution, but also differences in portfolio composition.

Overall, these results suggest that the multivariate calibration can offset distortions associated with "high" holdings of specific instruments (already addressed through the marginals), while placing relatively more weight on portfolios with less balanced compositions, rather than simply on high-net-wealth households. In other words, accounting for cross-instrument dependence allows single instruments to be corrected while explicitly considering the implications for total household wealth.In

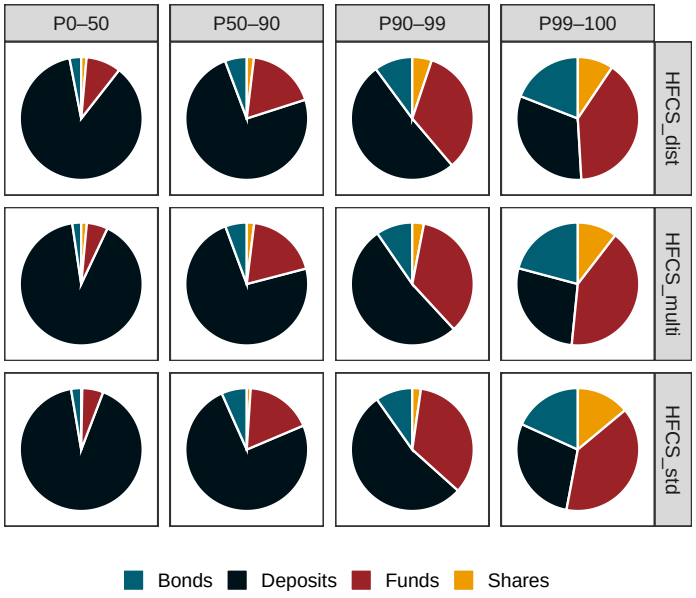

Figure 6.11. Household financial portfolio composition by year and net-wealth group.
Notes: Pie charts show the shares of deposits, listed shares, bonds, and investment fund shares. Facets place years in rows and net-wealth groups in columns. Models compared: HFCS_{std}, $\text{HFCS}_{\text{dist}}$, and $\text{HFCS}_{\text{multi}}$.

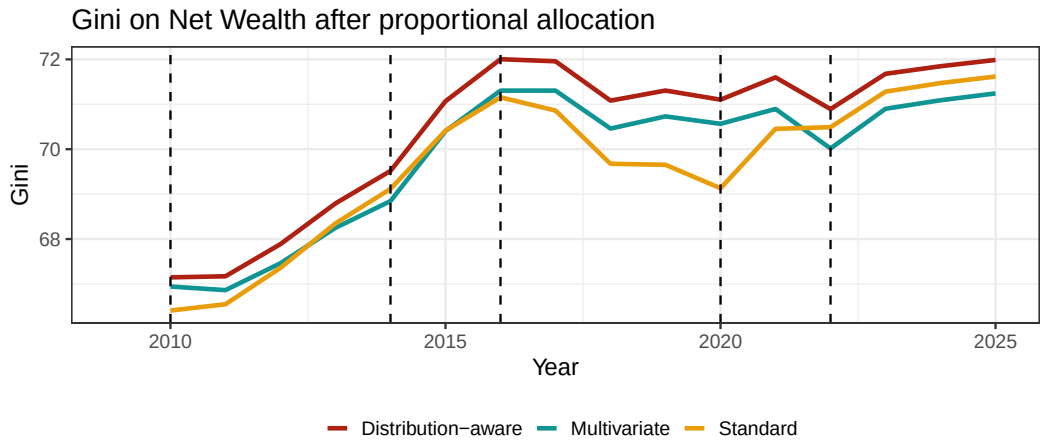

Figure 6.12. Gini of net wealth – standard vs distribution-aware vs multivariate full pipelines.
Notes: Lines report year-by-year Gini for overall net wealth and for financial net wealth under $\text{HFCS}^{\text{fin}}_{\text{std}}$, $\text{HFCS}^{\text{fin}}_{\text{dist}}$, and $\text{HFCS}^{\text{fin}}_{\text{multi}}$. Vertical dashed lines may indicate survey waves.

Appendix A Figure A.19 complements this analysis providing the other relevant inequality indicators between methods.

Turning to micro-level credibility and, crucially, to dependence diagnostics – the distinctive deliverable of Model 3. Figure 6.13 contrasts the distributions of adjustment parameters under $\text{HFCS}_{\text{multi}}$, $\text{HFCS}_{\text{dist}}$, and $\text{HFCS}_{\text{class}}$.

Figure 6.13. Adjustment parameters under $\text{HFCS}_{\text{multi}}$ vs $\text{HFCS}_{\text{dist}}$ vs $\text{HFCS}_{\text{class}}$.
Notes: Comparative distributions of multiplicative adjustment parameters by instrument and year. Report medians and extreme quantiles in-panel.

Table 6.7 reports dispersion diagnostics for $\text{HFCS}_{\text{multi}}$, and Figure A.20, shown in Appendix A, reports rank-based pre-post correlations to quantify deformation under each method.

We finally diagnose the alignment of cross-instrument dependence, the core dimension introduced in Model 3. We show regime-specific correlation matrices on the Gaussian-score scale (low: $[0, 250\text{k})$; high: $[250\text{k}, \infty)$) and quantify gaps to targets.

Figure 6.14 presents, by regime, heatmaps of the implied correlation matrices under $\text{HFCS}_{\text{multi}}$, the targets $R^{\star}$, and their differences. In Appendix A, Figure A.21, and Figure A.22 complement this analysis. Figure A.21 compares regime-specific correlation matrices across $\text{HFCS}_{\text{class}}$, $\text{HFCS}_{\text{dist}}$, and $\text{HFCS}_{\text{multi}}$. Finally, Figure A.22 reports joint-tail co-exceedance rates $P(X_k > q_{k,95}, X_\ell > q_{\ell,95})$ across methods. Together, these diagnostics highlight the practical importance of modelling dependence explicitly: aligning marginals alone can reproduce totals yet misstate regime-specific co-movement, whereas the multivariate calibration closes these gaps – especially in the high-wealth regime where securities co-move more tightly and deposits decouple – while preserving marginal fit.

Table 6.7. Adjustment-parameter dispersion diagnostics (cap baseline = 15) for $\text{HFCS}_{\text{multi}}$.

Model	Year	Instrument	median_ϕ	μ_ϕ	p99_ϕ	$\max_\phi$	σ_ϕ
$\text{HFCS}_{\text{multi}}$	2010	Deposits	1.39	1.39	2.58	5.00	0.28
$\text{HFCS}_{\text{multi}}$	2014	Deposits	1.79	1.67	2.41	4.05	0.21
$\text{HFCS}_{\text{multi}}$	2016	Deposits	1.54	1.47	1.86	3.59	0.06
$\text{HFCS}_{\text{multi}}$	2020	Deposits	1.43	1.33	1.91	5.22	0.10
$\text{HFCS}_{\text{multi}}$	2022	Deposits	1.61	1.53	1.73	2.14	0.05
$\text{HFCS}_{\text{multi}}$	2010	Bonds	3.33	3.18	5.05	10.00	0.64
$\text{HFCS}_{\text{multi}}$	2014	Bonds	3.98	4.09	8.16	8.95	2.54
$\text{HFCS}_{\text{multi}}$	2016	Bonds	1.02	0.95	1.98	3.99	0.10
$\text{HFCS}_{\text{multi}}$	2020	Bonds	0.67	0.69	1.83	2.21	0.18
$\text{HFCS}_{\text{multi}}$	2022	Bonds	0.89	0.84	1.23	2.02	0.03
$\text{HFCS}_{\text{multi}}$	2010	Shares	0.11	0.47	8.66	15.00	2.67
$\text{HFCS}_{\text{multi}}$	2014	Shares	1.74	3.04	11.00	12.00	7.48
$\text{HFCS}_{\text{multi}}$	2016	Shares	0.10	0.20	0.98	7.68	0.16
$\text{HFCS}_{\text{multi}}$	2020	Shares	0.10	0.23	1.50	5.09	0.12
$\text{HFCS}_{\text{multi}}$	2022	Shares	1.04	1.20	2.72	2.86	0.23
$\text{HFCS}_{\text{multi}}$	2010	Funds	0.41	0.87	4.14	5.89	0.99
$\text{HFCS}_{\text{multi}}$	2014	Funds	5.28	5.42	12.00	12.00	8.01
$\text{HFCS}_{\text{multi}}$	2016	Funds	2.78	3.04	5.88	6.38	1.11
$\text{HFCS}_{\text{multi}}$	2020	Funds	2.24	2.12	4.12	4.44	0.62
$\text{HFCS}_{\text{multi}}$	2022	Funds	1.75	1.79	2.15	5.73	0.07

Notes: Weighted diagnostics by instrument and year.

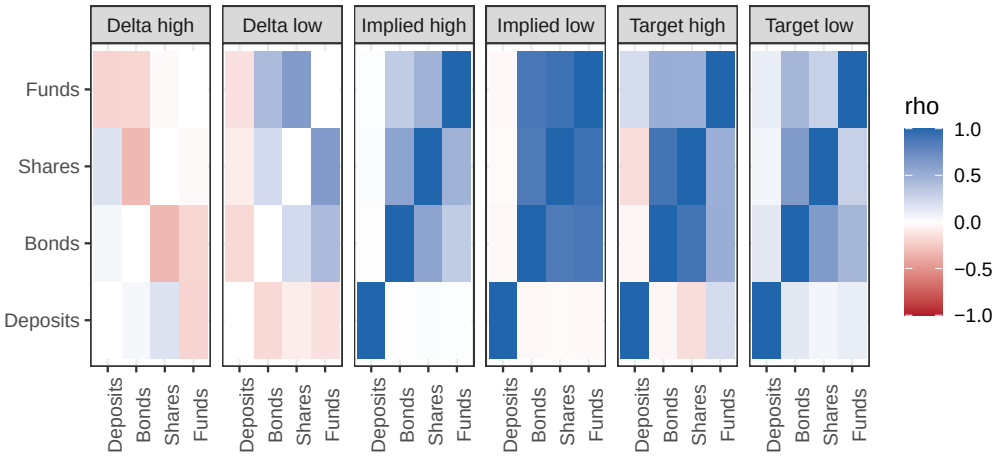

Figure 6.14. Regime-specific correlation alignment: implied vs target matrices.
Notes: Heatmaps show the implied Gaussian-score correlation matrices under $\text{HFCS}_{\text{multi}}$, the targets $R^\star$, and their differences for low and high regimes. Colour scales are fixed across panels; matrices are positive definite by construction.

Chapter 7

Conclusion

This thesis develops a reconciliation procedure within the expanding literature on survey-administrative integration for household wealth. The central aim is to incorporate information from administrative aggregates into survey microdata and to correct several instruments jointly, while preserving, as far as possible and without downgrading coverage excessively, the cross-sectional structure implied by the HFCS and aligning both aggregate totals and target marginal distributions. The resulting pipeline is tailored to the operational constraints of Distributional Wealth Accounts: it delivers microdata that are coherent with quarterly aggregates and bracketed administrative tabulations, and that remain suitable for dissemination and downstream analysis without imposing changes on the survey design weights. The contribution is cumulative. We first construct credible instrument-level targets at the household scale from administrative sources. We then calibrate survey values to these targets under procedures that progressively address aggregate coverage, marginal distributional shape, and cross-instrument dependence. The outcome is a modular reconciliation framework that can be adopted incrementally and audited transparently.

The first building block concerns parametric representation and identification of targets. Chapter 3 examined the distributional families commonly used for financial variables and established a parsimonious yet robust class that accommodates heavy tails. Financial phenomena often exhibit right-skewness, high kurtosis, and concentration of mass at the top – traits that simple lognormal forms may struggle to capture. We therefore adopted and tested splice models with a lognormal body and a Pareto tail, together with a procedure to recover parameters from limited aggregate information. The chapter also introduced a testing framework to discriminate between lognormal and heavier-tailed alternatives, balancing parsimony with the need to fit tail behaviour credibly. Simulation evidence showed that the procedure identifies, with high confidence, both the data-generating class and its parameters across diverse starting scenarios, including cases with thin information. This provides the methodological basis for recovering distributional targets from aggregates when only bracketed or first-moment information is available – a prerequisite for any calibration that aims to control the shape of the microdata rather than only its totals.

The second building block addresses misalignment between the observational perimeters of Banking Supervisory Reports (BSR) and the HFCS. BSR deposits are account-based, whereas the HFCS is household-based. Chapter 4 exploited the account-household links observed in the survey to aggregate BSR deposits to the household level, preserving intra-household dependence across accounts and ensuring that the target perimeter matches the survey object. The procedure leverages the additional HFCS variable collected at the national level on the number of accounts, providing a basic imputation procedure for households with positive holdings and zero-reported number of deposit accounts, to construct a coherent household-level deposit target from account-based aggregates. Simulation results indicated that, under explicit assumptions, the method recovers the household distribution and parameters with high accuracy, both in central regions and in the right tail. For securities, Chapter 5 tackled the range-sum reporting structure by brackets of total securities. The BSR provide, by bracket of total securities, instrument-level totals that are size-biased first moments over accounts with at least one positive holding. We devised a copula-informed disentanglement that uses HFCS information on co-positivity and zero dependence to separate listed shares, bonds, and investment fund shares. The method characterises identification limits when co-positives are sparse in high-value brackets and clarifies the mapping between range-sum totals and instrument marginals under truncation. While identification can be thin in applied settings, an augmented specification that softly anchors the mean to a credible external target stabilised estimation and improved finite-sample performance without forcing the system to hard constraints. These two chapters establish the administrative-data backbone required for a principled reconciliation with the survey.

With targets clarified and placed on a household footing, the third building block develops and compares three calibration strategies for the HFCS microdata, collected in Chapter 6. Model 1 implements a class-consistent, value-based calibration that iteratively matches instrument totals by BSR break, penalising class hopping to ensure stable adjustments. The calibration acts on reported values through multiplicative factors while keeping design weights fixed, thereby avoiding the spillovers that weight calibration would impose on all survey domains. As a benchmark already employed in Italian DWA production, this procedure raises coverage materially and is computationally stable. However, two structural drawbacks emerged for skewed instruments with long tails: first, the tendency of observations to accumulate at class boundaries; second, the under-representation of the far right tail when the top bracket is open-ended. These are direct consequences of steering by discrete breaks and enforcing totals locally without controlling the global shape.

Model 2, defined in Section 6.1.2, addressed these distributional shortcomings. We proposed a composite loss that aligns the full marginal distribution (via a smoothed CDF discrepancy) and the relative cumulative amount function (CAF) of adjusted values, leaving survey weights unchanged. The CAF term anchors totals softly and promotes bracket coherence when thresholds are included in the grid, without reintroducing discrete boundary effects. Optimisation proceeds in two stages: a shape-oriented fit that minimises discrepancies in CDF and CAF under box bounds for multiplicative factors, followed by a gated stabiliser that selects the smallest

deformation consistent with the achieved fit. Zeros remain fixed per instrument to respect participation. Empirically, the distribution-aware calibration delivers coverage comparable to the class-based benchmark while producing a substantially more credible marginal fit to BSR-derived targets. Boundary piling is visibly reduced, and the right tail aligns with the splice-based targets in a way that discrete break matching cannot guarantee. Because proportional allocation subsequently operates on negligible residues, the end-to-end indicators – such as the Gini and top-share measures – reflect these gains and do not revert to the pre-calibration patterns.

Model 3 (Section 6.1.3) extended the approach to the multivariate case. Different configurations of unit-level adjustments can map to the same totals and similar univariate shapes yet imply very different cross-instrument co-movements – a critical dimension for portfolio analysis and dissemination. We therefore introduced a dependence layer that operates on a scale that removes marginal effects by construction. For each instrument, adjusted values are mapped through the parametric PIT associated with the calibrated margin and then to Gaussian scores. This places all instruments on a common standard where correlations have a direct interpretation and can be targeted without confounding from scale or tail differences. Dependence is aligned in value regimes that are economically distinct yet statistically supported. In the baseline, we adopted common breaks – a low segment $[0, 250\text{k})$ and a high segment $[250\text{k}, \infty)$ – implemented via smooth log-scale windows so that gradients remain well behaved. For each regime and year, target correlation matrices are estimated using survey weights on co-positives and projected to the nearest positive definite form with minimal adjustment. The Stage 1 loss augments the marginal terms with regime-specific Frobenius gaps to the targets, introduced conservatively so that marginal alignment is preserved. Stage 2 activates the normalised stabiliser under tolerances that now include a bound on the dependence loss. The outcome is a multivariate calibration that preserves the gains of Section 6.1.2 while aligning cross-instrument dependence in regimes where behaviour differs.

The empirical takeaways are clear and, importantly, consistent across chapters. First, aggregate alignment to quarterly totals improves markedly at each stage, with residual gaps small enough that proportional allocation operates on negligible residues. This ensures that distributional gains achieved by the calibration are not undone in the final step of the pipeline. Second, relative to the class-consistent benchmark, the distribution-aware calibration reduces discrepancies to target CDFs across instruments and years and diminishes boundary piling. These improvements persist after end-to-end processing and are reflected in standard indicators. Third, the multivariate calibration brings the implied dependence of portfolios closer to regime-specific survey-informed targets. Pairwise correlations on the Gaussian-score scale move towards the target matrices; joint-tail co-exceedance rates for large holdings become more realistic; and the shifts in portfolio composition by net-wealth group are interpretable in light of the administrative aggregates. In short, the final method delivers marginal credibility and joint coherence simultaneously, trading a small fraction of aggregate precision – relative to a break-matching method – for a much larger gain in distributional realism and dependence control.

The design choices and their implications are explicit. Results are deterministic conditional on inputs and a fixed seed. In the absence of a household-level ground truth, we have not reported resampling-based uncertainty, preferring to document realised changes in coverage, distributional fit, and dependence alignment. Targets rely on careful HFCS-BSR harmonisation and on parametric margins that, while parsimonious and validated in simulations, remain modelling choices. The dependence layer targets regime-specific Gaussian-score correlations rather than a full copula; this is deliberate, balancing interpretability, signal strength, and computational tractability. Participation is handled upstream, and where false zeros are suspected the quality of corrections depends on the effectiveness of prior identification and imputation. These constraints are not incidental – they reflect the production environment of DWA and the information actually available from administrative sources.

From an operational perspective, the pipeline is modular and auditable as it can be adopted in steps. We emphasise, however, that applying the univariate calibrations of Chapter 6 to several instruments simultaneously without the dependence control of Section 6.1.3 can undermine credibility and stability – for example, by inducing cross-instrument drifts and boundary piling that propagate to dissemination outputs. For multi-instrument deployments, the regime-specific dependence layer should therefore be regarded as the default to preserve portfolio structure and the reliability of aggregate wealth distribution indices.

The scope and limitations are equally explicit. The parametric margins are splice-based and impose a structure that, while empirically sensible and supported by tail testing, is not the only possibility. Alternative bodies and tails could be introduced where diagnostics suggest systematic lack of fit. The dependence layer emphasises pairwise correlations on a transformed scale and does not attempt to encode tail asymmetries or higher-order interactions; this is a controlled restriction aligned with the information content of administrative aggregates and the sampling variability of the survey. Temporal coherence is not enforced by construction, although quarterly interpolation and extrapolation are applied to totals downstream; incorporating mild cross-wave regularisation in the calibration could further stabilise series used for dissemination. Moreover, publicly available documentation on country-specific implementation choices adopted by other national central banks is often insufficiently detailed; consequently, the thesis cannot undertake a systematic cross-country benchmarking of the proposed procedures beyond comparisons at the level of published aggregates. Finally, while computation is efficient enough for the instruments and sample sizes considered, adding instruments or moving to finer segmentation would benefit from compiled kernels and sparse-linear-algebra routines to keep runtimes predictable.

Two main refinements emerge as priorities. First, improving the treatment of false zeros – particularly for securities – would strengthen both the disentanglement in Chapter 5 and the multivariate calibration in Section 6.1.3. Even conservative imputation schemes could reduce bias and sharpen identification, especially in high-value regimes where co-positives are scarce. Second, tightening external anchors for depen-

dence – for instance, through administrative linkages at coarser resolution (microlevel, in the best case scenario) or through structured shrinkage to robust templates – could ease reliance on survey-only correlation signals where support is thin, while keeping the framework auditable. Additional developments that we regard as feasible include: reporting uncertainty via design-based linearisation or targeted bootstrap for selected indicators; experimenting with alternative dependence functionals that emphasise partial correlations or tail co-exceedance rates; enriching regime design where sample support permits; and imposing mild temporal coherence across waves and quarters via penalties that act on changes in the CAF or the dependence gap. Each of these extensions would build on the existing scaffolding rather than replace it.

In conclusion, the thesis delivers a coherent, production-oriented pipeline for reconciling survey microdata with administrative targets in DWA. It integrates household-level target construction for deposits, disentangles securities from range sums, and calibrates microdata to match both marginal shapes and regime-specific dependence – all while safeguarding the practical constraints of dissemination and leaving design weights intact. The resulting procedure is a credible, more reliable alternative to current practice, with clear diagnostics and a controlled set of tuning parameters. With focused work on false zeros and dependence anchors, it can transition from research to production and support more informative, distribution-aware wealth statistics that better reflect both aggregates and the distributional structure of household portfolios.

Appendix A

Figures

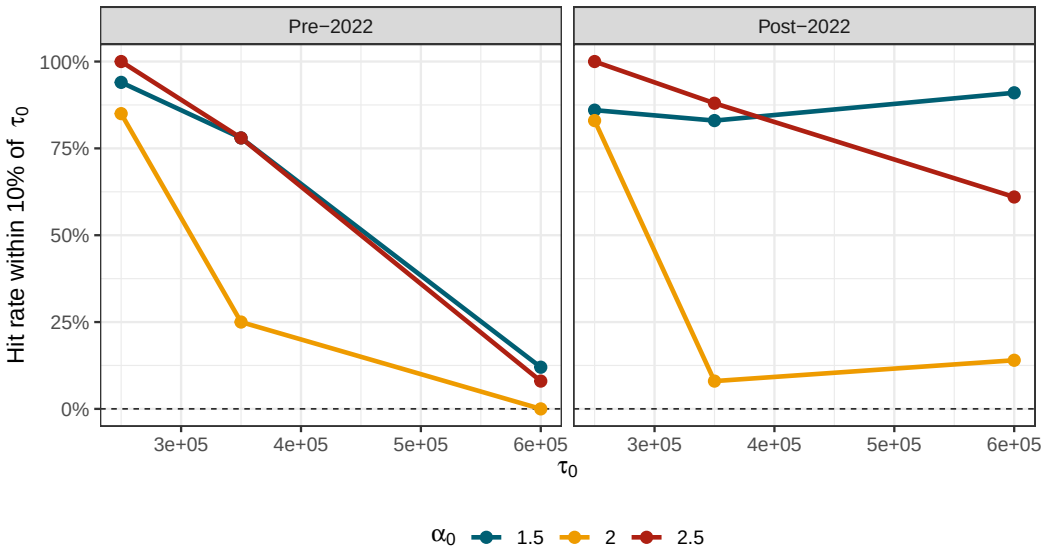

Figure A.1. PLN threshold identification: hit rate within 10 per cent of the true τ_0, by α and bracket scheme.

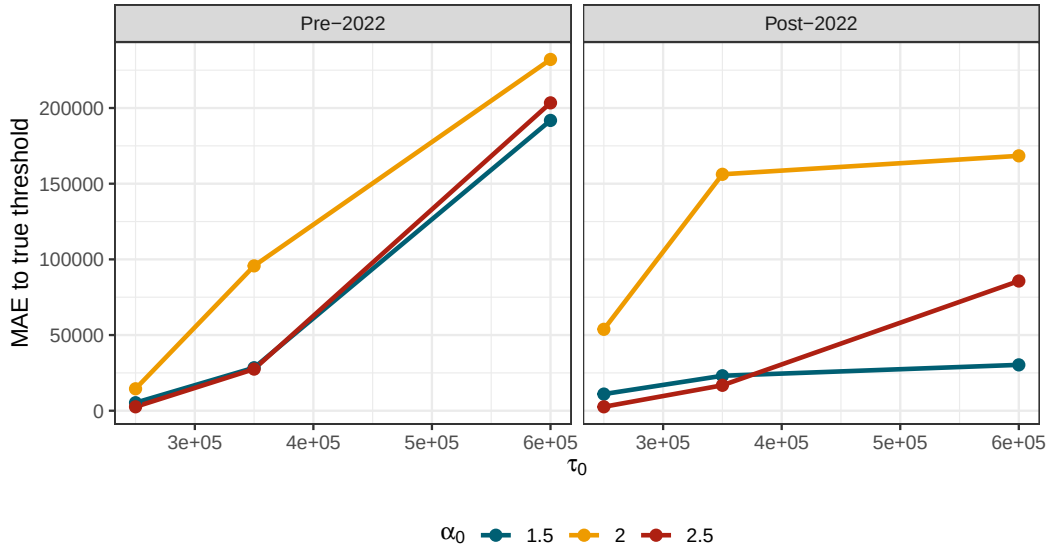

Figure A.2. PLN threshold identification: mean absolute error to the true τ_0, by α and bracket scheme.

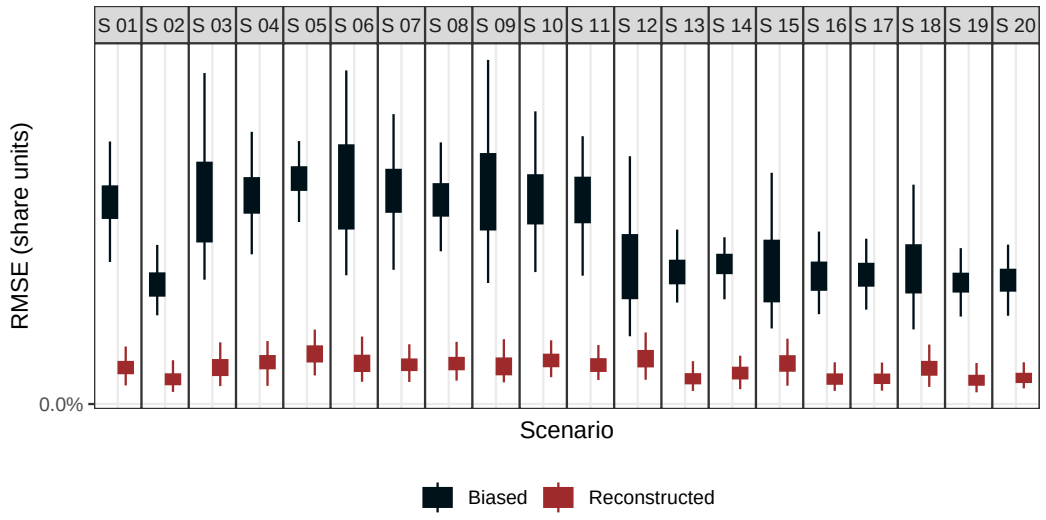

Figure A.3. RMSE by scenario: Before (Biased) vs After (Reconstructed).
Notes: Boxplots over $B = 100$ replications. "Before (Biased)" compares biased pseudo-HFCS household shares to truth; "After (Reconstructed)" compares reconstructed shares to truth.

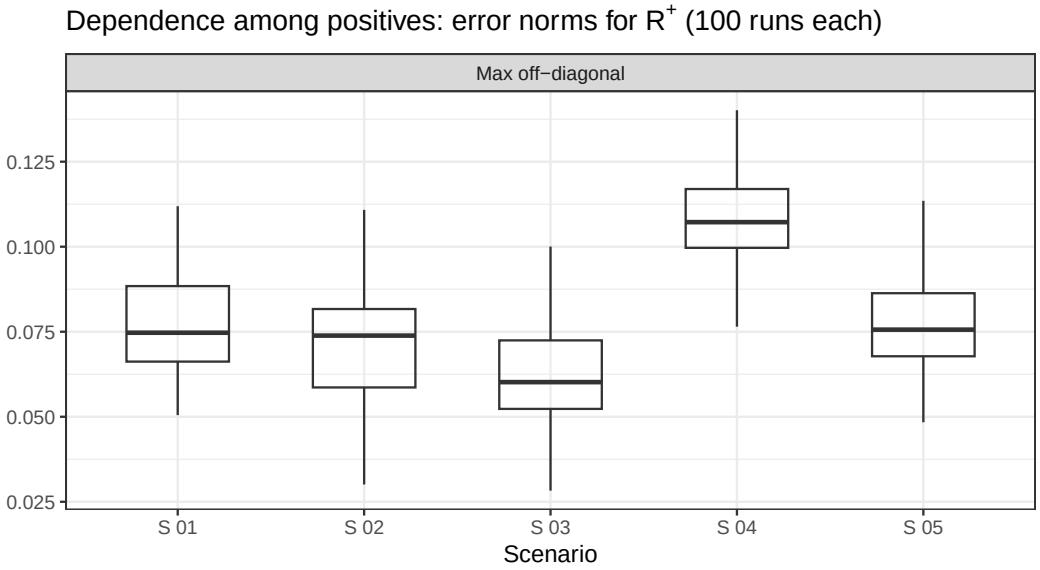

Figure A.4. Base scenario (100 runs): dependence among positives – maximum off-diagonal error across runs.

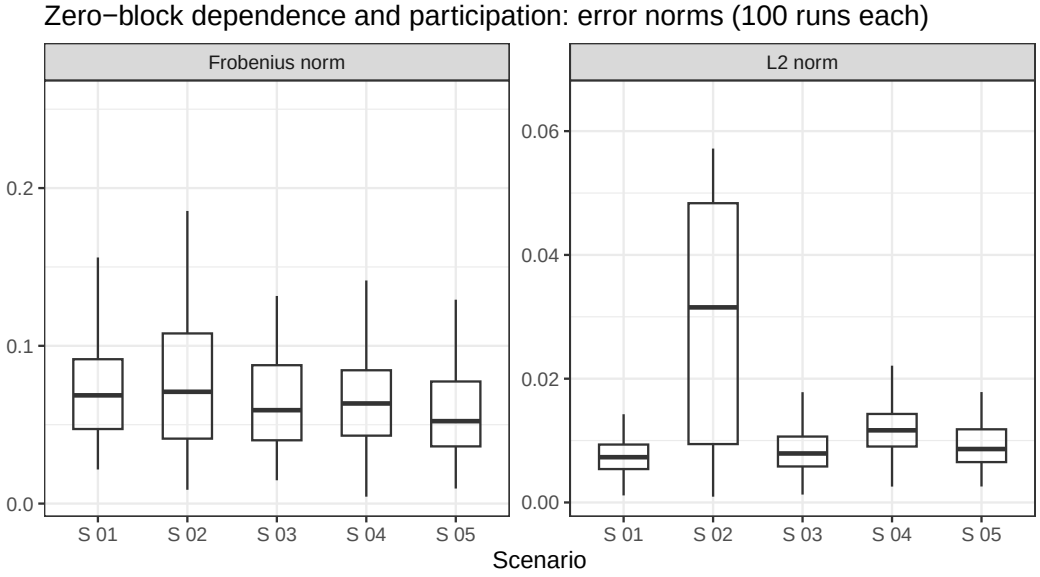

Figure A.5. Five-scenario panel (100 runs each): zero-block dependence and participation – Frobenius error for R^0 and error norms for $\{\pi_k(j)\}$.

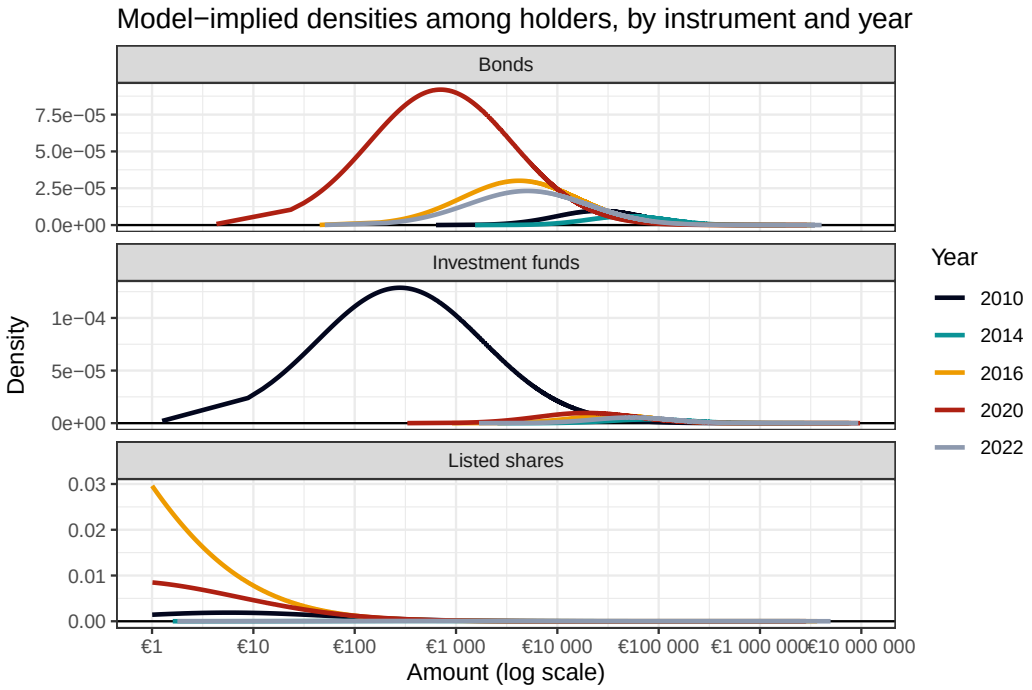

Figure A.6. Yearly densities (holders only) by instrument; log scale on the x-axis; rescaled with peaks equal to 1.

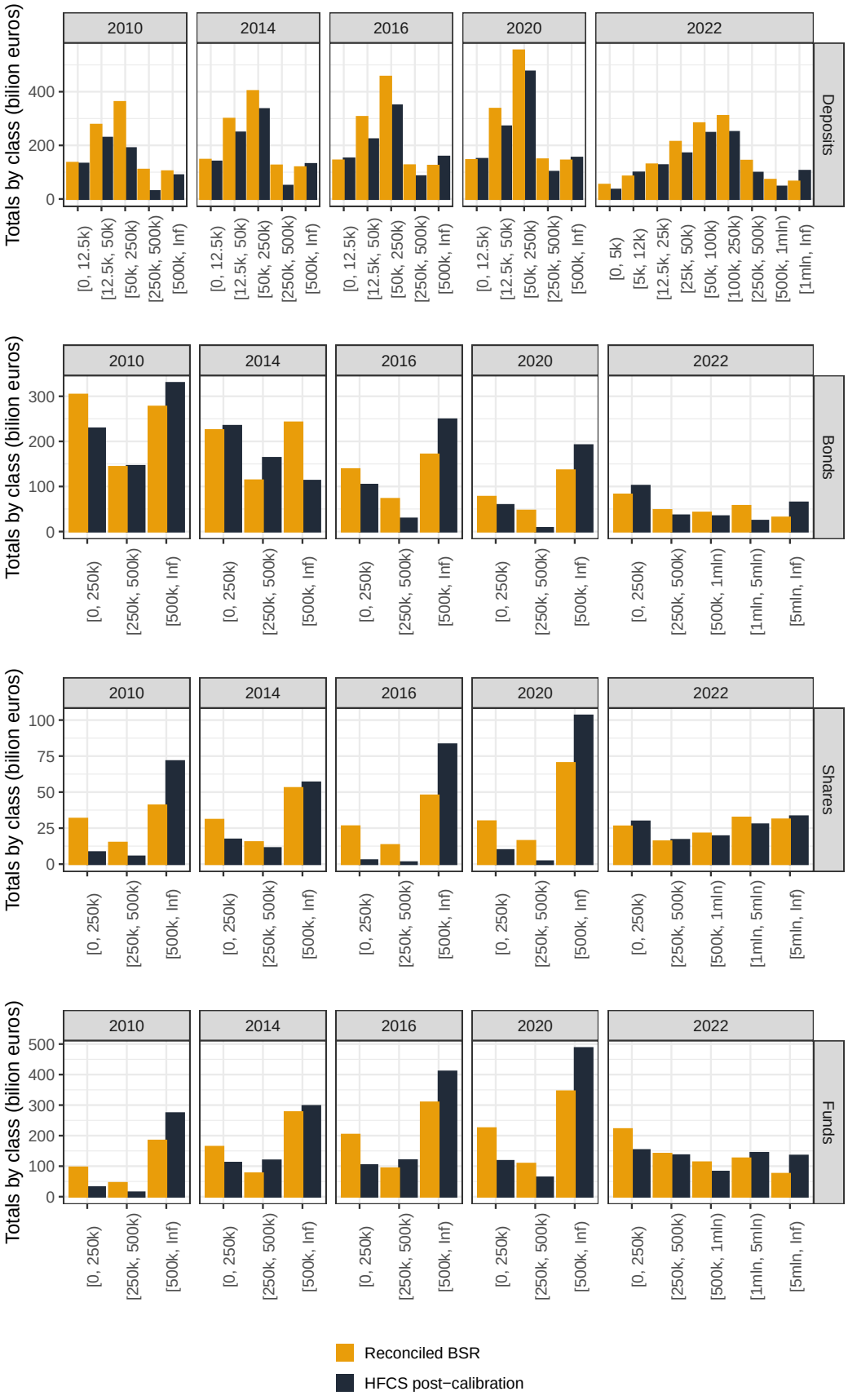

Figure A.7. Bracket-level totals: $\text{HFCS}_{\text{class}}$ vs BSR, by instrument and year.

Notes: For each bracket, the plot displays side-by-side bars for BSR totals and for totals implied by $\text{HFCS}_{\text{class}}$. Deposits are compared by deposit-account size; listed shares, bonds, and investment fund shares are compared by securities-account size. Binning and axes are harmonised across years.

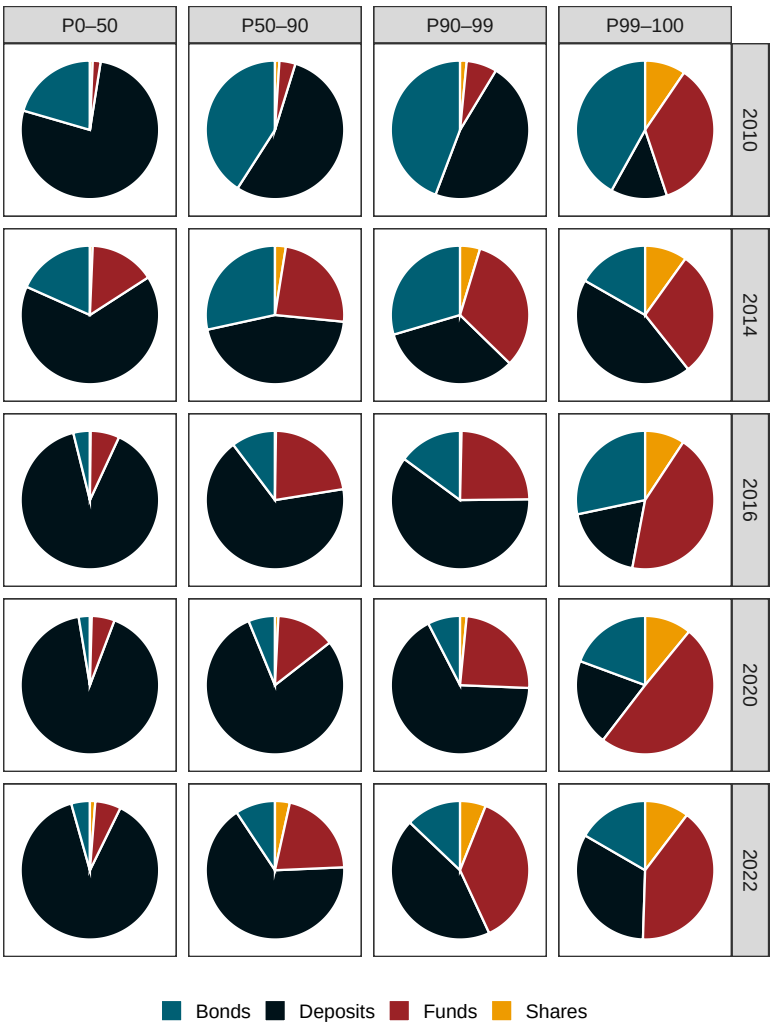

Figure A.8. Household financial portfolio composition by year and net-wealth group.
Notes: Pie charts show the shares of deposits, listed shares, bonds, and investment fund shares in the financial portfolio. Facets place years in rows and net-wealth groups in columns.

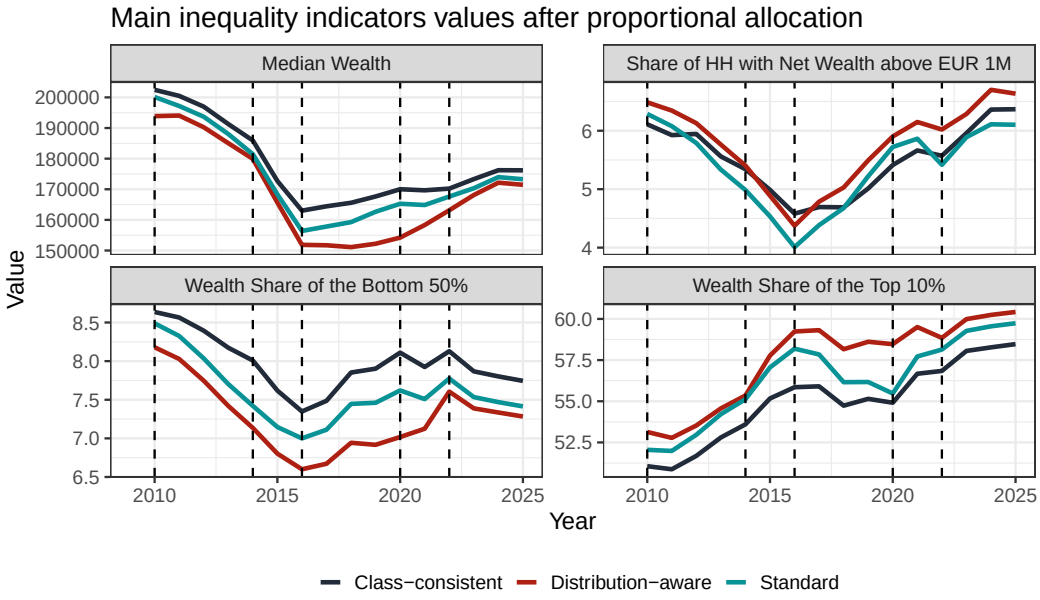

Figure A.9. Selected inequality indicators over time under the two full pipelines.
Notes: Panels report median wealth (thousands of euros), top-10 and top-5 per cent shares, bottom-50 per cent share, and the share of households above € 1 million.

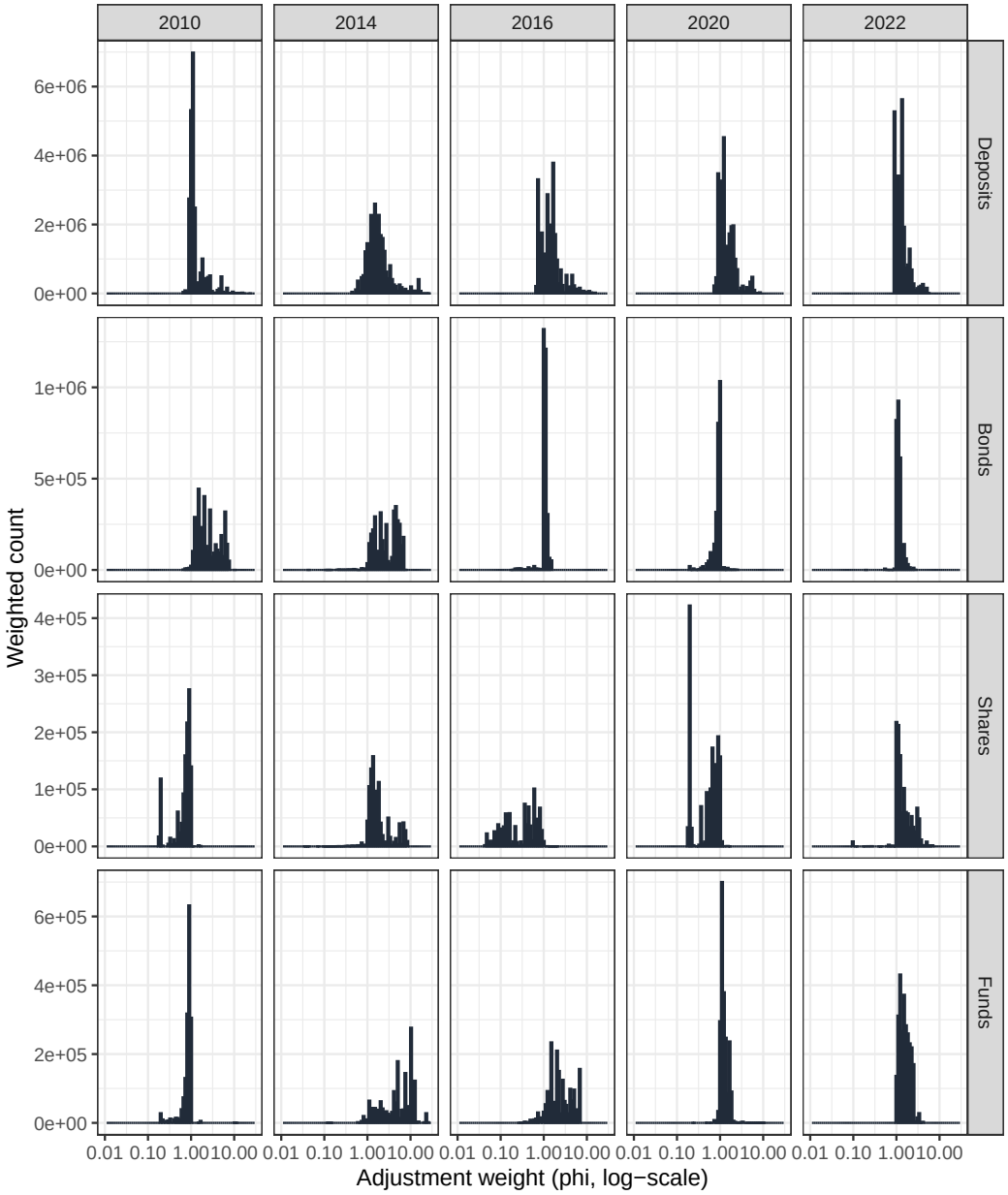

Figure A.10. Distribution of adjustment parameters (deposits shown as example).
Notes: The panel reports the histogram and selected quantiles of adjustment parameters after the baseline calibration. Values above 15 indicate accumulation across iterations.

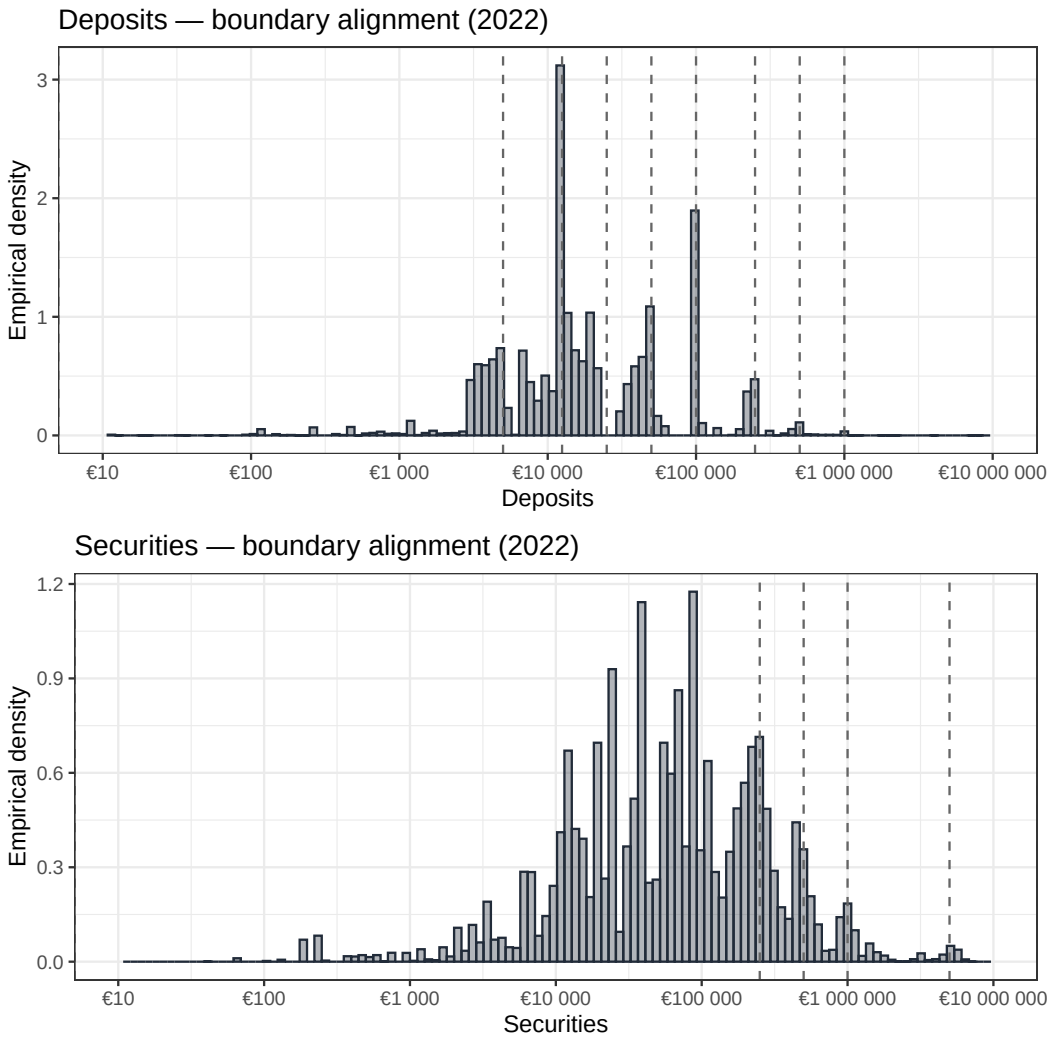

Figure A.11. Boundary concentration in 2022 (wave 5).
Notes: Elevated mass near the breaks signals boundary piling.

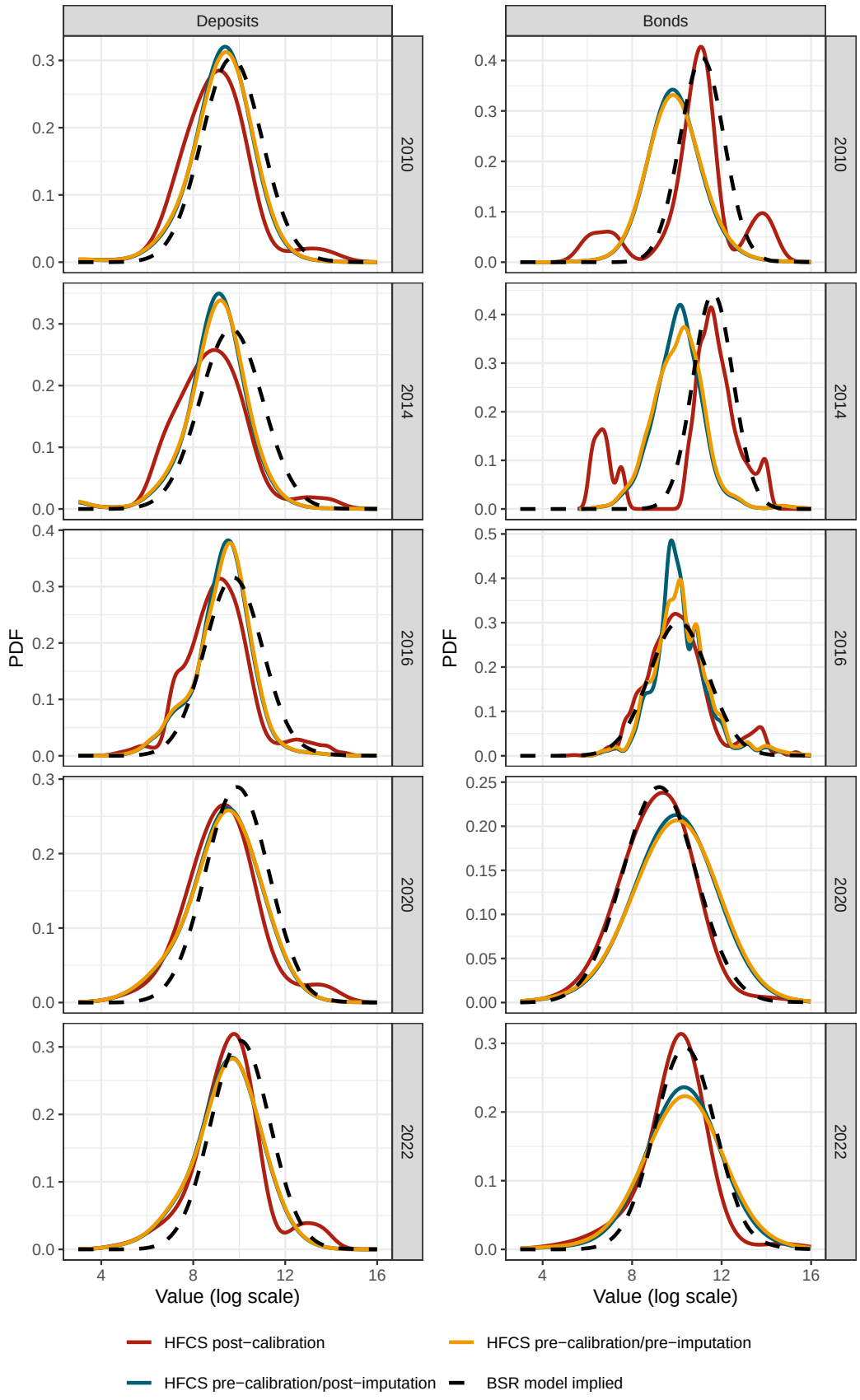

Figure A.12. PDF overlays (log scale), panel 1: Deposits and Bonds.
Notes: Curves show $\text{HFCS}_{\text{dist}}$ post-calibration (solid), pre-calibration references (solid, distinct colours), and BSR model-implied targets (dashed). Axes and bandwidths are harmonised across years.

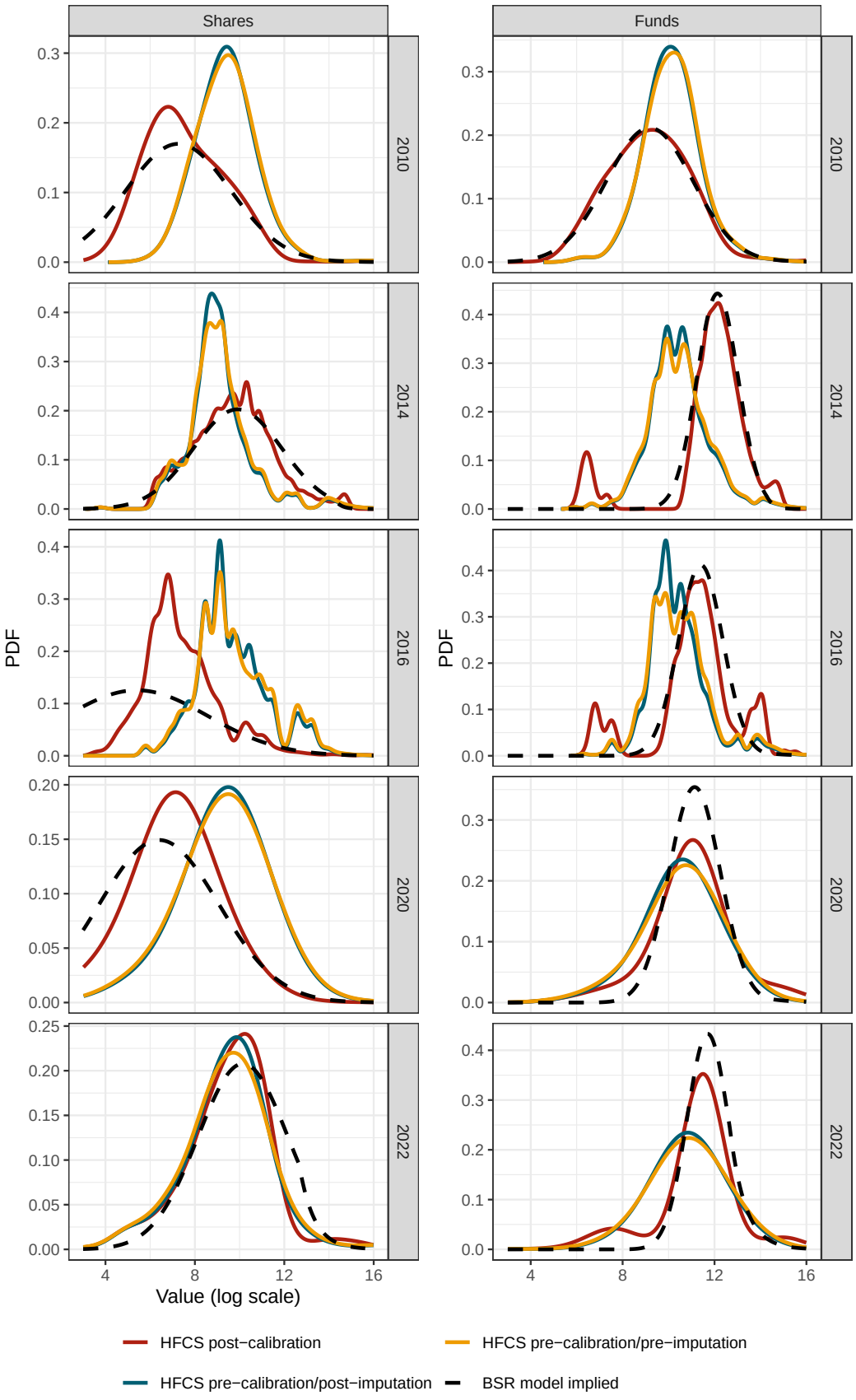

Figure A.13. PDF overlays (log scale), panel 2: Listed shares and Investment fund shares. Notes: As in Figure A.12.

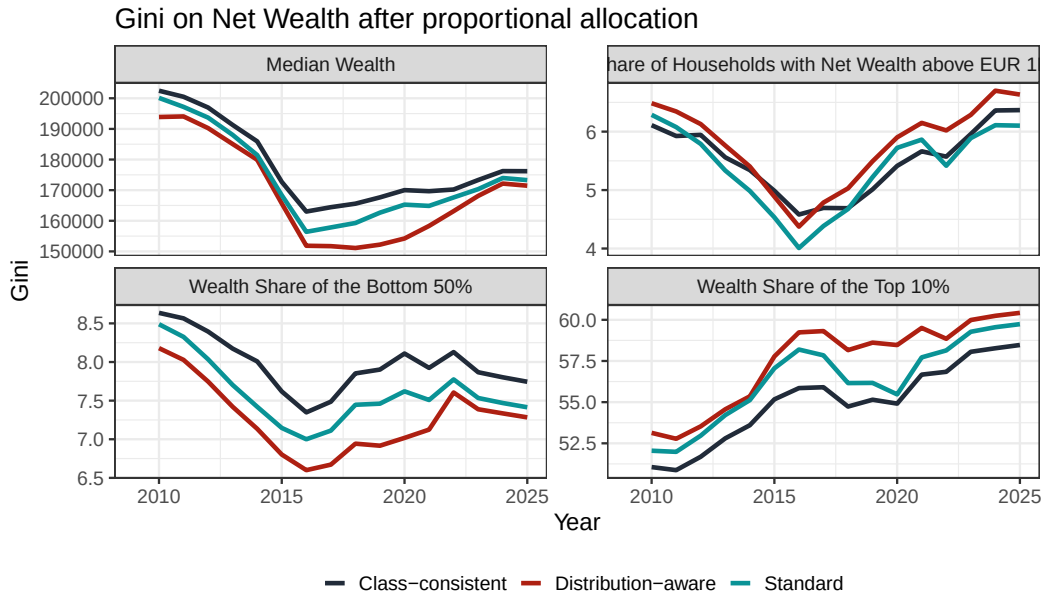

Figure A.14. Selected inequality indicators over time under the three full pipelines.
Notes: Panels report median wealth (thousands of euros), top-10 per cent shares, bottom-50 per cent share, and the share of households above € 1 million. Models: $\text{HFCS}^{\text{fin}}_{\text{std}}$, $\text{HFCS}^{\text{fin}}_{\text{class}}$, $\text{HFCS}^{\text{fin}}_{\text{dist}}$.

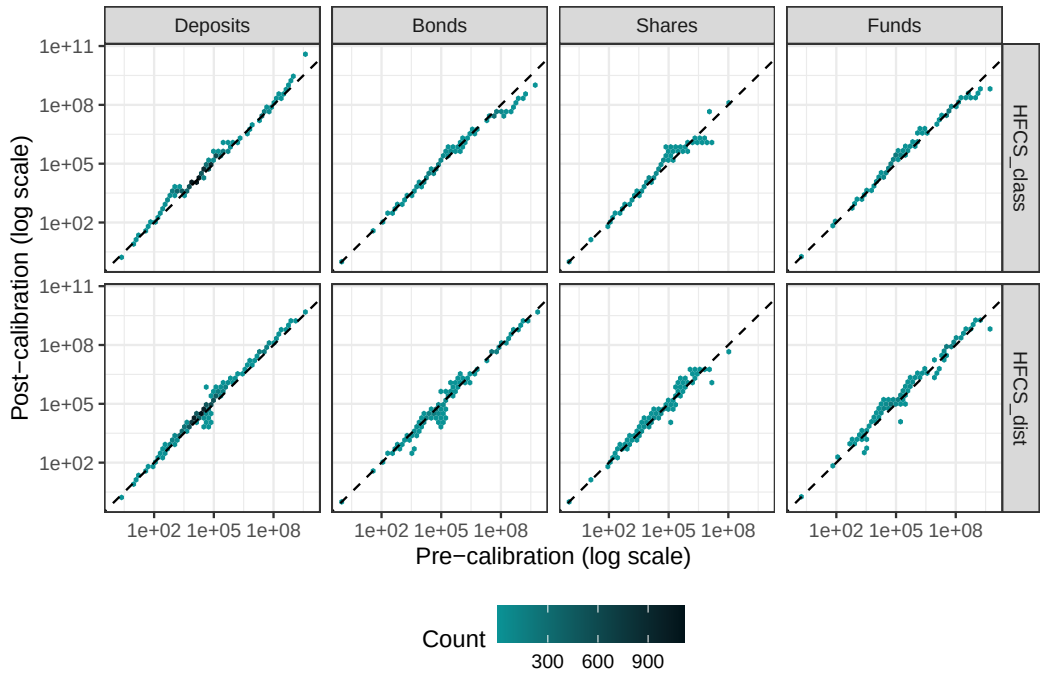

Figure A.15. Pre-post correlations under $\text{HFCS}_{\text{dist}}$ and $\text{HFCS}_{\text{class}}$.
Notes: Panels report Spearman correlations between pre- and post-calibration values and hexbin plots for selected instruments and years. Identical axes and binning aid comparison.

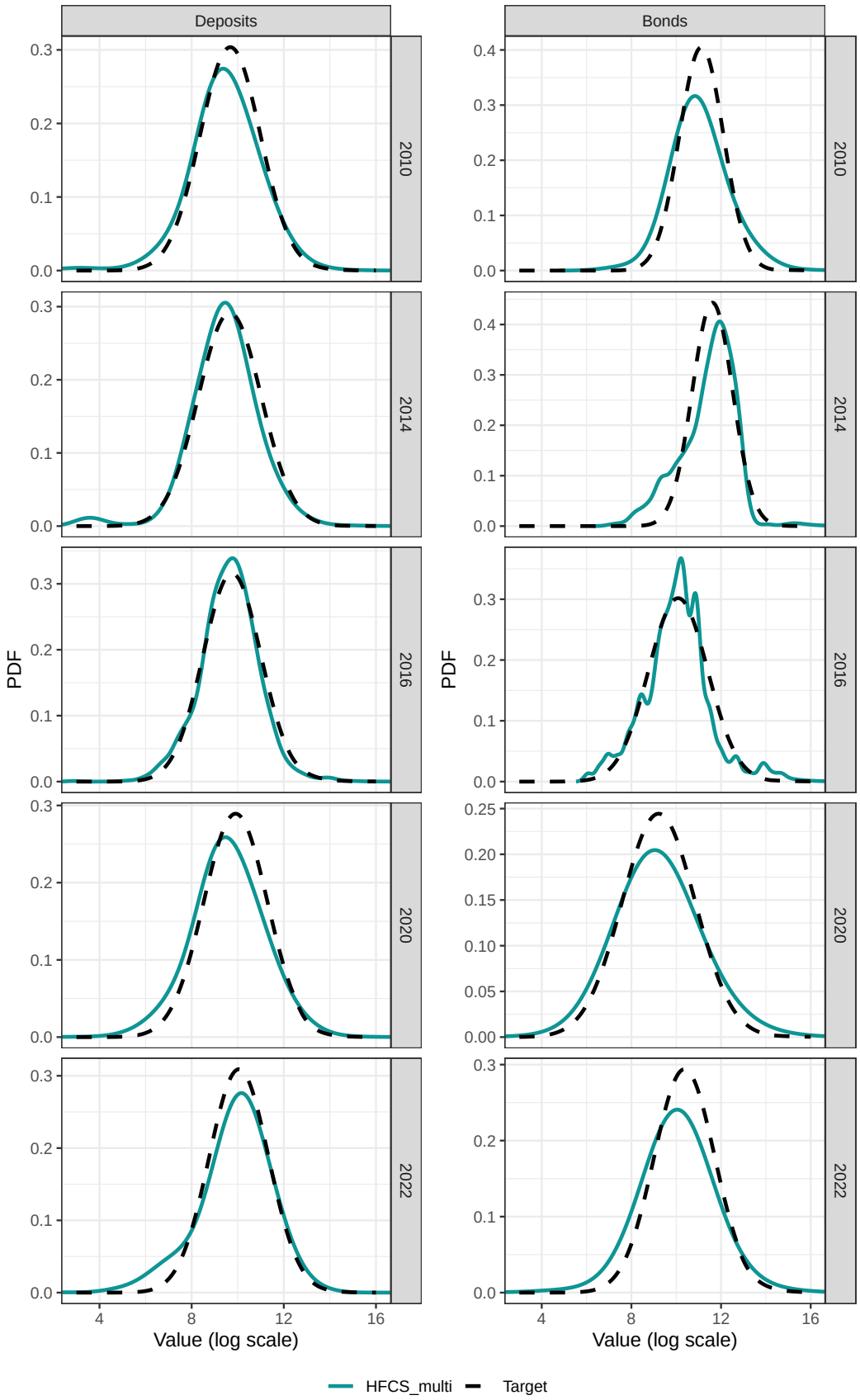

Figure A.16. PDF overlays (log scale), panel 1: Deposits and Bonds.
Notes: Curves show $\text{HFCS}_{\text{multi}}$ post-calibration (solid), pre-calibration/pre-imputation and post-imputation references (solid, distinct colours), and BSR model-implied targets (dashed). Axes and bandwidths are harmonised across years.

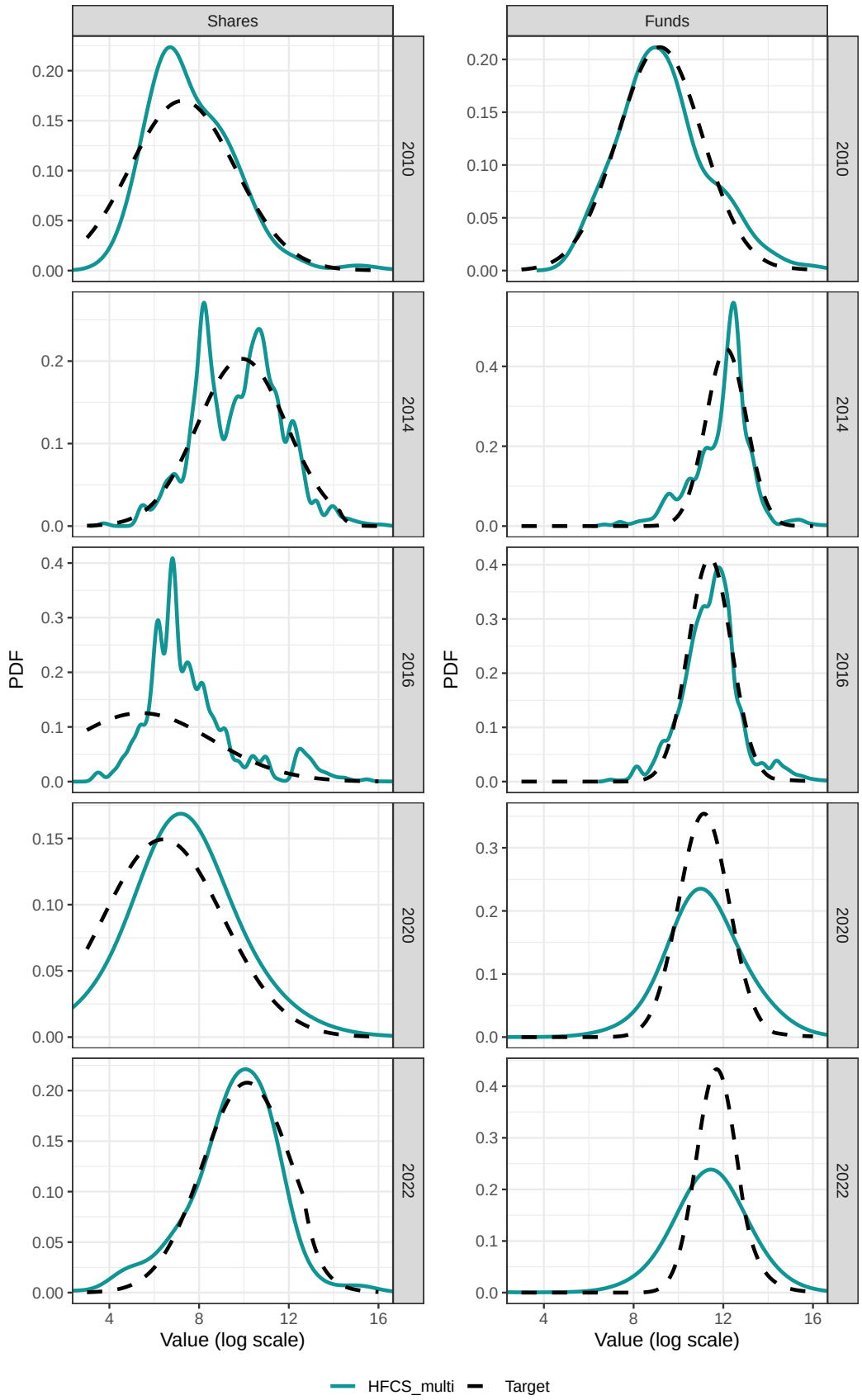

Figure A.17. PDF overlays (log scale), panel 2: Shares and Investment Fund shares.
Notes: As in Figure A.16.

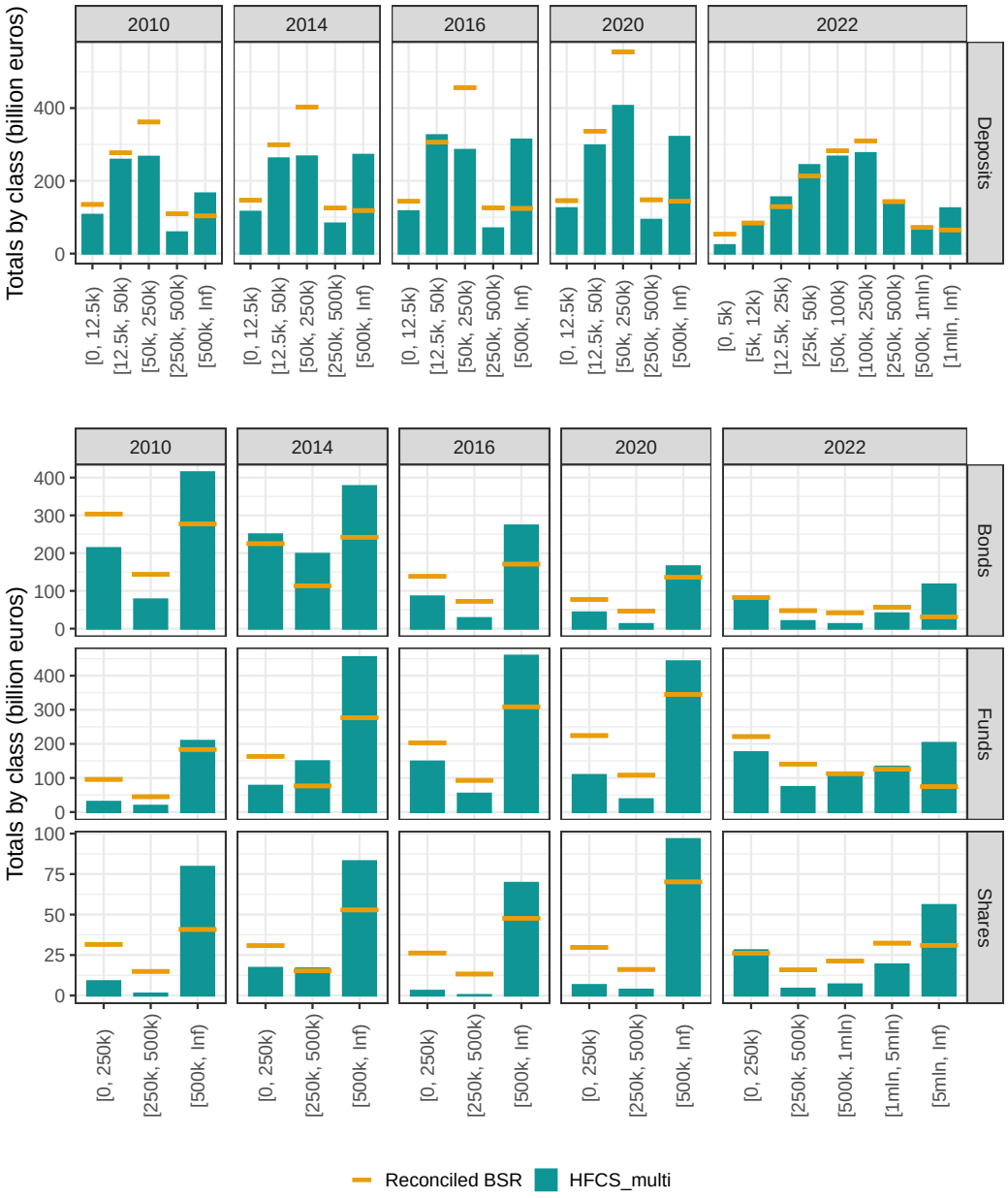

Figure A.18. Bracket-level totals: $\text{HFCS}_{\text{multi}}$ vs BSR, by instrument and year.

Notes: For each bracket, side-by-side bars compare BSR totals and $\text{HFCS}_{\text{multi}}$ totals. Deposits are compared by deposit-account size; listed shares, bonds, and investment fund shares are compared by securities-account size. Binning and axes are harmonised across years.

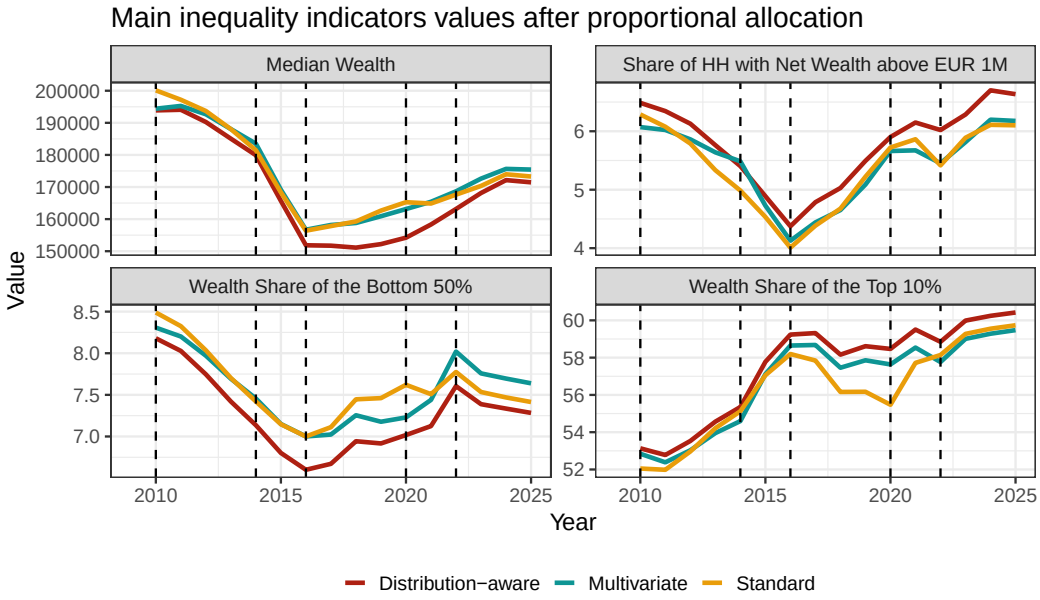

Figure A.19. Selected inequality indicators over time under the three full pipelines.
Notes: Panels report median wealth (thousands of euros), top-10 and top-5 per cent shares, bottom-50 per cent share, and the share of households above € 1 million. Models are $\text{HFCS}^{\text{fin}}_{\text{std}}$, $\text{HFCS}^{\text{fin}}_{\text{dist}}$, and $\text{HFCS}^{\text{fin}}_{\text{multi}}$.

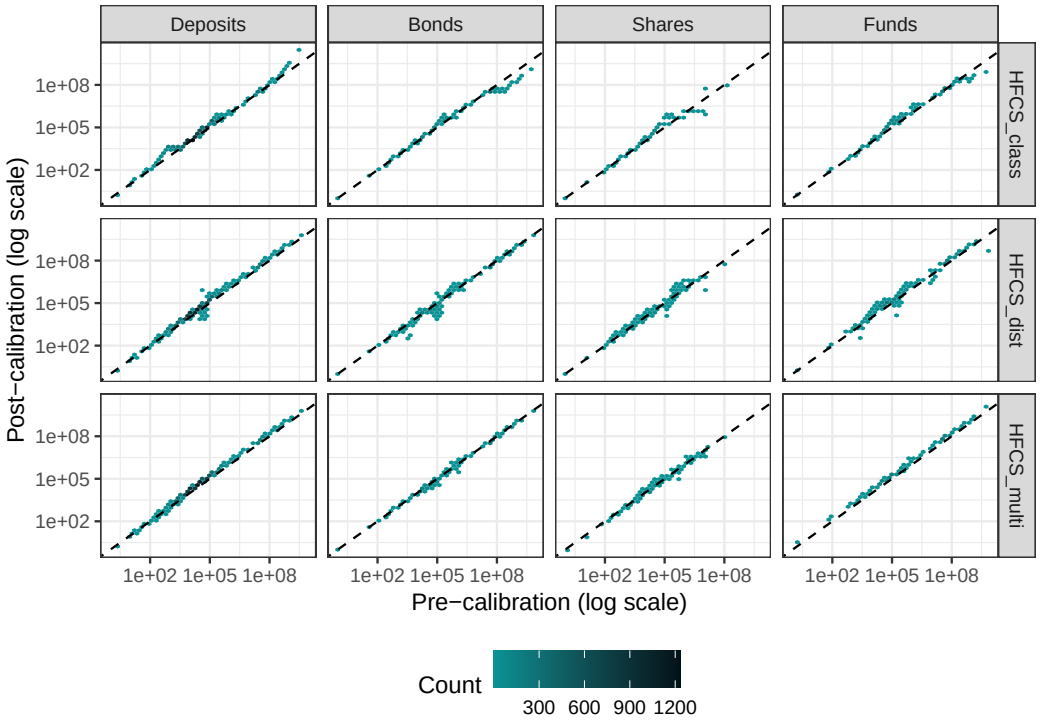

Figure A.20. Pre-post correlations under $\text{HFCS}_{\text{multi}}$, $\text{HFCS}_{\text{dist}}$, and $\text{HFCS}_{\text{class}}$.
Notes: Panels report Spearman correlations between pre- and post-calibration values and hexbin plots for selected instruments and years. Identical axes and binning are used.

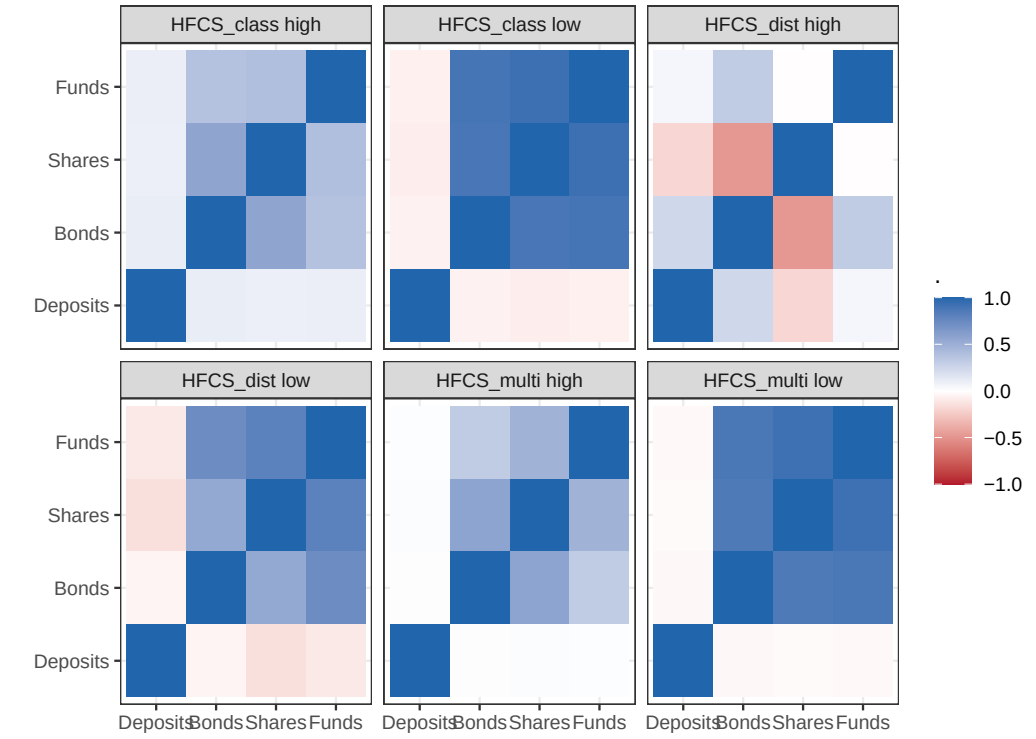

Figure A.21. Regime-specific dependence across methods.
Notes: Heatmaps of implied Gaussian-score correlation matrices for $\mathrm{HFCS}_{\mathrm{class}}$, $\mathrm{HFCS}_{\mathrm{dist}}$, and $\mathrm{HFCS}_{\mathrm{multi}}$ in low and high regimes. Differences to targets $R^{\star}$ can be shown in an inset or separate panels.

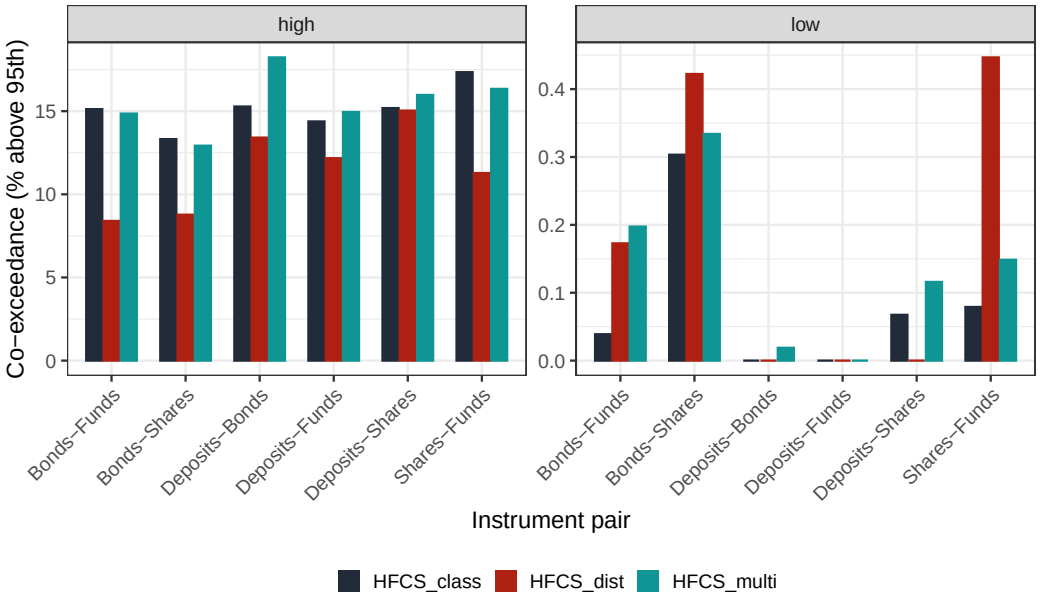

Figure A.22. Joint-tail co-exceedance rates across methods.
Notes: Panels report $P(X_k > q_{k,95}, X_\ell > q_{\ell,95})$ for selected pairs, by year and method. Quantiles $q_{k,95}$ refer to the method-specific marginal; identical axes aid comparison.

Appendix B

Tables

Table B.1. Total deposits by assets range following HFCS account split: HFCS vs BSR – pre-2022.

Year	Account Range	Total		Percentage	
		HFCS	**BSR**	**HFCS**	**BSR**
2010	[0, 12.5k)	98,722	135,035	30.77	13.69
	[12.5k, 50k)	127,401	276,970	39.71	28.09
	[50k, 250k)	66,899	361,736	20.85	36.68
	[250k, 500k)	11,028	109,408	3.44	11.09
	[500k, ∞)	16,773	103,494	5.23	10.49
	Total	320,825	985,657	-	-
2014	[0, 12.5k)	104,543	146,477	32.60	13.41
	[12.5k, 50k)	113,505	299,332	35.40	27.41
	[50k, 250k)	74,130	402,663	23.12	36.87
	[250k, 500k)	16,685	125,126	5.20	11.46
	[500k, ∞)	11,773	118,503	3.67	10.85
	Total	320,637	1,092,103	-	-
2016	[0, 12.5k)	105,862	143,836	27.12	12.44
	[12.5k, 50k)	139,879	306,033	35.84	26.48
	[50k, 250k)	93,475	456,129	23.95	39.46
	[250k, 500k)	30,474	125,808	7.81	10.88
	[500k, ∞)	20,648	124,124	5.29	10.74
	Total	390,340	1,155,933	-	-
2020	[0, 12.5k)	115,431	145,541	17.94	10.96
	[12.5k, 50k)	186,106	336,515	28.92	25.34
	[50k, 250k)	189,869	554,188	29.51	41.74
	[250k, 500k)	26,429	147,933	4.11	11.14
	[500k, ∞)	125,635	143,674	19.52	10.82
	Total	643,472	1,327,852	-	-
2022	[0, 5k)	37,440	53,288	6.50	3.94
	[5k, 12.5k)	77,292	84,059	13.42	6.22
	[12.5k, 25k)	88,322	129,148	15.34	9.56
	[25k, 50k)	107,032	213,257	18.59	15.78
	[50k, 100k)	104,149	282,593	18.09	20.90
	[100k, 250k)	91,276	309,728	15.85	22.92
	[250k, 500k)	39,361	142,913	6.84	10.58
	[500k, 1mln)	18,494	71,694	3.21	5.30
	[1mln, ∞)	12,397	65,132	2.15	4.82
	Total	575,767	1,351,812	-	-

Notes: Values in € millions; shares in per cent.

Table B.2. Total listed shares by securities account range: HFCS vs BSR – pre-2022.

Year	Account Range	Total		Percentage	
		HFCS	**BSR**	**HFCS**	**BSR**
2010	[0k, 250k)	17,656	31,543	56.93	36.17
	[250k, 500k)	7,968	14,891	25.69	17.07
	[500k, ∞)	5,391	40,775	17.38	46.76
	Total	31,015	87,209	-	-
2014	[0k, 250k)	10,852	30,784	58.22	31.13
	[250k, 500k)	3,473	15,257	18.63	15.43
	[500k, ∞)	4,315	52,841	23.15	53.44
	Total	18,640	98,882	-	-
2016	[0k, 250k)	14,536	26,267	45.58	30.13
	[250k, 500k)	4,915	13,244	15.41	15.19
	[500k, ∞)	12,443	47,670	39.01	54.68
	Total	31,894	87,182	-	-
2020	[0k, 250k)	29,537	29,726	44.80	25.63
	[250k, 500k)	6,234	16,095	9.46	13.87
	[500k, ∞)	30,157	70,181	45.74	60.50
	Total	65,927	116,001	-	-
2022	[0k, 250k)	19,077	26,172	13.43	20.67
	[250k, 500k)	6,040	15,828	4.25	12.50
	[500k, 1mln)	6,247	21,270	4.40	16.80
	[1mln, 5mln)	12,881	32,319	9.07	25.53
	[5mln, ∞)	97,781	31,014	68.85	24.50
	Total	142,026	126,603	-	-

Notes: Values in € millions; shares in per cent.

Table B.3. Total bonds by securities account range: HFCS vs BSR – pre-2022.

Year	Account Range	Total		Percentage	
		HFCS	**BSR**	**HFCS**	**BSR**
2010	[0k, 250k)	96,465	303,645	69.20	41.93
	[250k, 500k)	16,022	143,483	11.49	19.81
	[500k, ∞)	26,913	277,047	19.31	38.26
	Total	139,400	724,175	-	-
2014	[0k, 250k)	105,025	224,974	76.95	38.75
	[250k, 500k)	13,916	113,599	10.20	19.57
	[500k, ∞)	17,535	242,040	12.85	41.69
	Total	136,477	580,613	-	-
2016	[0k, 250k)	89,143	138,410	70.60	36.26
	[250k, 500k)	12,528	72,438	9.92	18.98
	[500k, ∞)	24,597	170,902	19.48	44.77
	Total	126,268	381,750	-	-
2020	[0k, 250k)	82,049	77,129	62.86	29.72
	[250k, 500k)	13,617	46,458	10.43	17.90
	[500k, ∞)	34,860	135,892	26.71	52.37
	Total	130,526	259,479	-	-
2022	[0k, 250k)	85,651	82,149	56.96	31.60
	[250k, 500k)	13,318	47,696	8.86	18.34
	[500k, 1mln)	12,824	42,070	8.53	16.18
	[1mln, 5mln)	14,332	56,918	9.53	21.89
	[5mln, ∞)	24,240	31,163	16.12	11.99
	Total	150,366	259,996	-	-

Notes: Values in € millions; shares in per cent.

Table B.4. Total investment fund shares by securities account range: HFCS vs BSR – pre-2022.

Year	Account Range	Total		Percentage	
		HFCS	**BSR**	**HFCS**	**BSR**
2010	[0k, 250k)	38,601	95,865	58.82	29.53
	[250k, 500k)	10,884	45,034	16.58	13.87
	[500k, ∞)	16,142	183,682	24.60	56.59
	Total	65,627	324,581	-	-
2014	[0k, 250k)	54,991	163,384	62.65	31.62
	[250k, 500k)	13,131	76,520	14.96	14.81
	[500k, ∞)	19,648	276,853	22.39	53.58
	Total	87,770	516,757	-	-
2016	[0k, 250k)	48,969	203,059	57.57	33.58
	[250k, 500k)	16,698	92,943	19.63	15.37
	[500k, ∞)	19,399	308,735	22.80	51.05
	Total	85,067	604,738	-	-
2020	[0k, 250k)	60,091	224,284	60.93	33.11
	[250k, 500k)	15,965	108,048	16.19	15.95
	[500k, ∞)	22,574	345,111	22.89	50.94
	Total	98,631	677,443	-	-
2022	[0k, 250k)	87,404	221,323	57.13	32.79
	[250k, 500k)	19,260	140,671	12.59	20.84
	[500k, 1mln)	19,714	112,548	12.88	16.68
	[1mln, 5mln)	12,198	125,504	7.97	18.60
	[5mln, ∞)	14,428	74,878	9.43	11.09
	Total	153,004	674,924	-	-

Notes: Values in € millions; shares in per cent.

Table B.5. Simulation scenarios for oracle recovery.

Scenario	Model	Breaks	α	τ
S 01	LN	pre-2022	-	-
S 02	LN	post-2022	-	-
S 03	PLN	pre-2022	1.5	250,000
S 04	PLN	pre-2022	2	250,000
S 05	PLN	pre-2022	2.5	250,000
S 06	PLN	pre-2022	1.5	350,000
S 07	PLN	pre-2022	2	350,000
S 08	PLN	pre-2022	2.5	350,000
S 09	PLN	pre-2022	1.5	600,000
S 10	PLN	pre-2022	2	600,000
S 11	PLN	pre-2022	2.5	600,000
S 12	PLN	post-2022	1.5	250,000
S 13	PLN	post-2022	2	250,000
S 14	PLN	post-2022	2.5	250,000
S 15	PLN	post-2022	1.5	350,000
S 16	PLN	post-2022	2	350,000
S 17	PLN	post-2022	2.5	350,000
S 18	PLN	post-2022	1.5	600,000
S 19	PLN	post-2022	2	600,000
S 20	PLN	post-2022	2.5	600,000

Notes: DGP is assumed known. μ and σ are fixed at 8.5 and 1.5, respectively. PLN additionally varies α and τ. Bracket schemes mirror BSR pre-2022 and post-2022 deposit bins.

Table B.6. Bracket-share fit: MAE across brackets by scenario (median and inter-decile range over replications).

Scenario	Biased pseudo-HFCS		Reconstructed	
	Median MAE	**IQR$_{10,90}$**	**Median MAE**	**IQR$_{10,90}$**
S 01	5.0%	1.6%	0.8%	0.5%
S 02	2.9%	1.1%	0.5%	0.3%
S 03	4.7%	2.4%	0.8%	0.6%
S 04	5.2%	1.8%	1.0%	0.6%
S 05	5.7%	1.2%	1.2%	0.7%
S 06	5.0%	3.0%	1.0%	0.8%
S 07	5.3%	2.0%	0.9%	0.6%
S 08	5.2%	1.8%	1.0%	0.6%
S 09	4.9%	2.8%	0.9%	0.6%
S 10	5.0%	2.2%	1.0%	0.6%
S 11	5.1%	1.8%	0.9%	0.6%
S 12	3.0%	1.9%	0.9%	0.4%
S 13	3.2%	1.0%	0.5%	0.3%
S 14	3.3%	0.8%	0.6%	0.3%
S 15	2.9%	1.6%	0.8%	0.4%
S 16	3.0%	1.4%	0.6%	0.3%
S 17	3.1%	1.0%	0.5%	0.3%
S 18	3.0%	1.7%	0.8%	0.4%
S 19	2.9%	1.2%	0.5%	0.4%
S 20	2.9%	1.0%	0.6%	0.3%

Table B.7. Bootstrap dispersion measures for PLN estimates in Table 4.4.

Year	Objective	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\tau}$
2010	Augmented	SE: 0.07 CI: [9.44, 9.75]	SE: 0.06 CI: [1.23, 1.48]	SE: 0.18 CI: [1.46, 2.33]	SE: 194,346 CI: [250,000, 1,000,000]
2010	Non-augmented	SE: 0.17 CI: [8.60, 9.21]	SE: 0.08 CI: [1.43, 1.72]	SE: 0.39 CI: [1.51, 3.03]	SE: 332,376 CI: [250,000, 1,000,000]
2014	Augmented	SE: 0.04 CI: [9.51, 9.70]	SE: 0.03 CI: [1.32, 1.42]	SE: 0.13 CI: [1.30, 1.80]	SE: 299,526 CI: [250,000, 1,000,000]
2014	Non-augmented	SE: 0.19 CI: [8.78, 9.36]	SE: 0.10 CI: [1.43, 1.71]	SE: 0.16 CI: [1.32, 1.84]	SE: 375,681 CI: [250,000, 1,000,000]
2016	Augmented	SE: 0.03 CI: [9.57, 9.71]	SE: 0.03 CI: [1.28, 1.43]	SE: 0.05 CI: [1.49, 1.70]	SE: 54,761 CI: [250,000, 500,000]
2016	Non-augmented	SE: 0.08 CI: [9.04, 9.30]	SE: 0.04 CI: [1.42, 1.56]	SE: 0.06 CI: [1.53, 1.73]	SE: 125,227 CI: [250,000, 500,000]
2020	Augmented	SE: 0.09 CI: [9.65, 10.00]	SE: 0.07 CI: [1.30, 1.56]	SE: 0.37 CI: [1.33, 2.97]	SE: 327,323 CI: [250,000, 1,000,000]
2020	Non-augmented	SE: 0.51 CI: [8.27, 10.18]	SE: 0.72 CI: [1.21, 3.72]	SE: 0.33 CI: [1.26, 2.44]	SE: 379,239 CI: [100,000, 1,000,000]
2022	Augmented	SE: 0.07 CI: [9.88, 10.24]	SE: 0.10 CI: [1.21, 1.68]	SE: 0.43 CI: [1.55, 3.11]	SE: 336,864 CI: [100,000, 1,000,000]
2022	Non-augmented	SE: 0.32 CI: [8.72, 10.01]	SE: 0.58 CI: [1.29, 3.48]	SE: 0.65 CI: [1.58, 3.44]	SE: 376,982 CI: [100,000, 1,000,000]

Notes: Dispersion is computed from B household-cluster bootstrap re-runs of the full reconstruction-and-estimation pipeline (mapping, re-aggregation, and PLN estimation) conditional on BSR totals. "SE" is the standard deviation of the bootstrap estimates. "CI" is the central percentile interval (2.5th and 97.5th percentiles) of the bootstrap distribution.

Table B.8. Portfolio composition by securities-account quintile group and wave for HFCS households with $A > 0$.

Wave	Quintile group	p_B	p_L	p_F	p_{BL}	p_{BF}	p_{LF}	p_{BLF}
	Q1 (0–20%)	59.59	17.99	16.27	2.75	0.23	3.07	0.11
	Q2 (20–40%)	69.49	8.19	12.65	1.98	5.71	1.98	0.00
1	Q3 (40–60%)	61.15	9.09	13.50	6.24	5.62	3.30	1.10
	Q4 (60–80%)	45.07	2.22	20.16	12.19	12.58	2.05	5.72
	Q5 (80–100%)	30.41	2.82	13.17	12.07	17.48	7.15	16.91
	Q1 (0–20%)	54.91	24.19	17.63	2.30	0.67	0.29	0.00
	Q2 (20–40%)	64.40	7.41	14.82	6.07	4.18	1.09	2.03
2	Q3 (40–60%)	60.90	2.62	14.78	9.08	8.18	3.77	0.67
	Q4 (60–80%)	51.48	1.27	19.58	6.03	13.67	3.88	4.10
	Q5 (80–100%)	32.86	1.31	21.02	8.88	16.11	3.97	15.86
	Q1 (0–20%)	58.31	23.17	15.44	0.96	1.31	0.29	0.52
	Q2 (20–40%)	51.39	5.48	29.33	1.23	5.56	5.86	1.15
3	Q3 (40–60%)	55.08	1.97	22.73	4.11	12.80	2.86	0.47
	Q4 (60–80%)	44.05	3.75	22.76	4.49	12.63	7.02	5.30
	Q5 (80–100%)	27.86	2.64	19.73	11.97	13.14	4.00	20.65
	Q1 (0–20%)	43.90	33.67	8.32	2.60	6.69	1.31	3.52
	Q2 (20–40%)	48.86	19.69	22.16	3.83	2.66	0.93	1.86
4	Q3 (40–60%)	32.57	11.32	25.36	7.05	8.82	12.28	2.61
	Q4 (60–80%)	32.09	8.28	32.75	9.61	8.36	6.06	2.84
	Q5 (80–100%)	20.03	3.20	24.85	8.44	16.29	10.89	16.30
	Q1 (0–20%)	46.46	20.62	22.73	6.71	1.02	1.95	0.52
	Q2 (20–40%)	41.67	3.04	36.35	4.25	8.71	4.30	1.68
5	Q3 (40–60%)	41.50	6.44	38.49	3.36	3.14	3.58	3.49
	Q4 (60–80%)	30.46	2.19	23.38	12.39	18.42	8.86	4.31
	Q5 (80–100%)	9.59	1.64	37.02	5.67	16.21	7.97	21.91

Note: HFCS diagnostics by quintiles of A used to inform smoothing of $\pi_k(\cdot)$ and $\lambda_k(\cdot)$. Calibration uses the fixed BSR partition $\mathcal{J}$ to compute simulated shares $\widetilde{s}_{kj}(\boldsymbol{\theta})$.

Table B.9. Median and IQR of the absolute error of bracket shares by scenario and margin.

Scenario	Margin	Median	IQR
S 01	Margin 1	0.26	0.46
S 01	Margin 2	0.26	0.46
S 01	Margin 3	0.22	0.32
S 02	Margin 1	0.42	0.60
S 02	Margin 2	0.34	0.61
S 02	Margin 3	0.29	1.00
S 03	Margin 1	0.12	0.17
S 03	Margin 2	0.08	0.10
S 03	Margin 3	0.11	0.11
S 04	Margin 1	0.32	0.53
S 04	Margin 2	0.30	0.66
S 04	Margin 3	0.50	1.07
S 05	Margin 1	0.40	0.78
S 05	Margin 2	0.35	0.71
S 05	Margin 3	0.30	0.87

Table B.10. Real-data overview by year and margin: marginal parameters, model-implied mean vs anchor with standard estimation (no mean anchor).

Year	Instrument	Model	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\alpha}$	$\widehat{\tau}$	$\widehat{\mu}^H$	gap Δ_μ
2010	Bonds	PLN	9.47	1.89	4.03	960,657	55,312	-1.24
2010	Listed shares	PLN	7.65	2.28	4.03	960,657	20,226	-1.29
2010	Investment funds	PLN	9.14	2.32	4.03	960,657	58,773	-1.22
2014	Bonds	PLN	7.25	2.48	2.84	1,213,844	20,820	-2.05
2014	Listed shares	PLN	5.92	2.50	2.84	1,213,844	7,039	-2.61
2014	Investment funds	PLN	8.05	2.56	2.84	1,213,844	40,904	-2.03
2016	Bonds	PLN	7.61	2.26	4.51	967,793	18,807	-1.88
2016	Listed shares	PLN	7.08	1.92	4.51	967,793	7,245	-2.49
2016	Investment funds	PLN	9.15	2.23	4.51	967,793	55,599	-1.75
2020	Bonds	PLN	10.54	2.76	1.34	95,246	65,766	-0.37
2020	Listed shares	PLN	8.98	1.54	1.34	95,246	33,611	-0.76
2020	Investment funds	PLN	14.45	0.76	1.34	95,246	303,996	-0.06
2022	Bonds	PLN	9.47	1.65	1.50	1,236,891	56,933	-0.48
2022	Listed shares	PLN	8.81	1.90	1.50	1,236,891	44,350	-0.81
2022	Investment funds	PLN	10.97	1.18	1.50	1,236,891	136,101	-0.61

Bibliography

Aitchison, J. and Brown, J. *The Lognormal Distribution: With Special Reference to Its Uses in Economics*. Monographs / Department of Applied Economics, University of Cambridge. Cambridge University Press, 1957. URL `https://books.google.it/books?id=ZyeHOQEACAAJ`.

Aitchison, J. "The statistical analysis of compositional data". *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

Alstadsæter, A., Johannesen, N., and Zucman, G. "Tax evasion and inequality". *American Economic Review*, 109(6):2073–2103, 2019.

Anderson, T. W. and Darling, D. A. "A test of goodness of fit". *Journal of the American statistical association*, 49(268):765–769, 1954.

Andrews, D. W. "Tests for parameter instability and structural change with unknown change point". *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Andrews, D. W. and Ploberger, W. "Optimal tests when a nuisance parameter is present only under the alternative". *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

Andridge, R. R. and Little, R. J. "A review of hot deck imputation for survey non-response". *International statistical review*, 78(1):40–64, 2010.

Bach, S., Thiemann, A., and Zucco, A. "Looking for the missing rich: Tracing the top tail of the wealth distribution". *International Tax and Public Finance*, 26(6): 1234–1258, 2019.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data". *The Journal of Machine Learning Research*, 9:485–516, 2008.

Basu, A., Shioya, H., and Park, C. *Statistical inference: the minimum distance approach*. CRC press, 2011.

Beckman, R. J., Baggerly, K. A., and McKay, M. D. "Creating synthetic baseline populations". *Transportation Research Part A: Policy and Practice*, 30(6):415–429, 1996.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. *Statistics of Extremes: Theory and Applications*. Wiley, 2004. ISBN 0471976474. Pagination: 522.

Binder, D. A. "On the variances of asymptotically normal estimators from complex surveys". *International Statistical Review/Revue Internationale de Statistique*, pages 279–292, 1983.

Bobasu, A., Di Nino, V., and Osbat, C. "The impact of the recent inflation surge across households". *Economic Bulletin Articles*, 3, 2023.

Brown, L. D., Cai, T. T., and DasGupta, A. "Interval estimation for a binomial proportion". *Statistical science*, 16(2):101–133, 2001.

Buono, D., Elliott, D., Bikker, R., Frölich, M., Gatto, R., Guardalbascio, B., Hauf, S., Infante, E., Moauro, F., Oltmanns, E., et al. *European Statistical System (ESS) guidelines on temporal disaggregation, benchmarking and reconciliation*. European Union, 2018.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. "Bootstrap-based improvements for inference with clustered errors". *The review of economics and statistics*, 90(3): 414–427, 2008.

Cannon, A. J. "Multivariate quantile mapping bias correction: an n-dimensional probability density function transform for climate model simulations of multiple variables". *Climate dynamics*, 50(1):31–49, 2018.

Cantarella, M., Neri, A., and Ranalli, M. G. "Mind the wealth gap: a new allocation method to match micro and macro statistics for household wealth". *arXiv preprint arXiv:2101.01085*, 2021.

Cantarella, M., Neri, A., and Ranalli, M. G. "Estimating the distribution of household wealth: methods for adjusting survey data estimates using national accounts and rich list data". *Review of Income and Wealth*, 70(3):551–580, 2024.

Cardot, H., Goga, C., and Shehzad, M.-A. "Calibration and partial calibration on principal components when the number of auxiliary variables is large". *Statistica Sinica*, pages 243–260, 2017.

Carroll, R. J. and Hall, P. "Optimal rates of convergence for deconvolving a density". *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.

Chang, T. and Kott, P. S. "Using calibration weighting to adjust for nonresponse under a plausible model". *Biometrika*, 95(3):555–571, 2008.

Chen, H.-S., Lai, K., and Ying, Z. "Goodness-of-fit tests and minimum power divergence estimators for survival data". *Statistica Sinica*, pages 231–248, 2004.

Chernozhukov, V., Fernández-Val, I., and Melly, B. "Inference on counterfactual distributions". *Econometrica*, 81(6):2205–2268, 2013.

Chotikapanich, D., Griffiths, W. E., Hajargasht, G., Karunarathne, W., and Rao, D. P. "Using the gb2 income distribution". *Econometrics*, 6(2):21, 2018.

Clauset, A., Shalizi, C. R., and Newman, M. E. "Power-law distributions in empirical data". *SIAM review*, 51(4):661–703, 2009.

Clopper, C. J. and Pearson, E. S. "The use of confidence or fiducial limits illustrated in the case of the binomial". *Biometrika*, 26(4):404–413, 1934.

Coles, S., Heffernan, J., and Tawn, J. "Dependence measures for extreme value analyses". *Extremes*, 2(4):339–365, 1999.

Cooray, K. and Ananda, M. M. "Modeling actuarial data with a composite lognormal-pareto model". *Scandinavian Actuarial Journal*, 2005(5):321–334, 2005.

Cressie, N. and Read, T. R. "Multinomial goodness-of-fit tests". *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 46(3):440–464, 1984.

Cuturi, M. "Sinkhorn distances: Lightspeed computation of optimal transport". *Advances in neural information processing systems*, 26, 2013.

Dagum, E. B. and Cholette, P. A. *Benchmarking, temporal distribution, and reconciliation methods for time series*. Springer, 2006.

Davies, J. B. and Shorrocks, A. F. "The distribution of wealth". *Handbook of income distribution*, 1:605–675, 2000.

Davies, R. B. "Hypothesis testing when a nuisance parameter is present only under the alternative". *Biometrika*, 74(1):33–43, 1987.

Davison, A. C. and Hinkley, D. V. *Bootstrap methods and their application*. Number 1. Cambridge university press, 1997.

Dekkers, A. L., Einmahl, J. H., and De Haan, L. "A moment estimator for the index of an extreme-value distribution". *The Annals of Statistics*, pages 1833–1855, 1989.

Demarta, S. and McNeil, A. J. "The t copula and related copulas". *International statistical review*, 73(1):111–129, 2005.

Deming, W. E. and Stephan, F. F. "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known". *The Annals of Mathematical Statistics*, 11(4):427–444, 1940.

Denton, F. T. "Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization". *Journal of the american statistical association*, 66(333):99–102, 1971.

Deville, J.-C. and Särndal, C.-E. "Calibration estimators in survey sampling". *Journal of the American statistical Association*, 87(418):376–382, 1992.

Deville, J.-C., Särndal, C.-E., and Sautory, O. "Generalized raking procedures in survey sampling". *Journal of the American statistical Association*, 88(423): 1013–1020, 1993.

Embrechts, P., Lindskog, F., and McNeil, A. "Modelling dependence with copulas and applications to risk management". *Handbook of heavy tailed distributions in finance*, 8(1):329–384, 2003.

Embrechts, P., Puccetti, G., and Rüschendorf, L. "Model uncertainty and var aggregation". *Journal of Banking & Finance*, 37(8):2750–2764, 2013.

Engel, J., Riera, P. G., Grilli, J., and Sola, P. *Developing reconciled quarterly distributional national wealth: Insight into inequality and wealth structures*. Number 2687. ECB Working Paper, 2022.

Fan, J. "On the optimal rates of convergence for nonparametric deconvolution problems". *The Annals of Statistics*, pages 1257–1272, 1991.

Friedman, J., Hastie, T., and Tibshirani, R. "Sparse inverse covariance estimation with the graphical lasso". *Biostatistics*, 9(3):432–441, 2008.

Frigessi, A., Haug, O., and Rue, H. "A dynamic mixture model for unsupervised tail estimation without threshold selection". *Extremes*, 5(3):219–235, 2002.

Gabaix, X. "Power laws in economics and finance". *Annu. Rev. Econ.*, 1(1):255–294, 2009.

Gastwirth, J. L. "The estimation of the lorenz curve and gini index". *The review of economics and statistics*, pages 306–316, 1972.

Genest, C. and Rémillard, B. "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models". In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 1096–1127, 2008.

Goga, C. and Shehzad, M. A. "Pak. j. statist. 2014 vol. 30 (4), 429-438 a note on partially penalized calibration". *Pak. J. Statist*, 30(4):429–438, 2014.

Graf, M. and Nedyalkova, D. "Modeling of income and indicators of poverty and social exclusion using the generalized beta distribution of the second kind". *Review of Income and Wealth*, 60(4):821–842, 2014.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. "A kernel two-sample test". *The journal of machine learning research*, 13(1):723–773, 2012.

Griffiths, W. E. and Hajargasht, G. "Gmm estimation of mixtures from grouped data: an application to income distributions". *Department of Economics-Working Papers Series*, 1148, 2012.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T. "Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods". *Hydrology and Earth System Sciences*, 16(9):3383–3390, 2012.

Guggemos, F. and Tillé, Y. "Penalized calibration in survey sampling: Design-based estimation assisted by mixed models". *Journal of statistical planning and inference*, 140(11):3199–3212, 2010.

Guillou, A. and Hall, P. "A diagnostic for selecting the threshold in extreme value analysis". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):293–305, 2001.

Hajargasht, G. and Griffiths, W. E. "Pareto–lognormal distributions: Inequality, poverty, and estimation from grouped income data". *Economic Modelling*, 33: 593–604, 2013.

Hajargasht, G., Griffiths, W. E., Brice, J., Rao, D. P., and Chotikapanich, D. "Inference for income distributions using grouped data". *Journal of Business & Economic Statistics*, 30(4):563–575, 2012.

Hall, P. and Horowitz, J. L. "Bootstrap critical values for tests based on generalized-method-of-moments estimators". *Econometrica: Journal of the Econometric Society*, pages 891–916, 1996.

Hansen, B. E. "Inference when a nuisance parameter is not identified under the null hypothesis". *Econometrica: Journal of the econometric society*, pages 413–430, 1996.

Hansen, B. E. "Sample splitting and threshold estimation". *Econometrica*, 68(3): 575–603, 2000.

Hansen, L. P. "Large sample properties of generalized method of moments estimators". *Econometrica: Journal of the econometric society*, pages 1029–1054, 1982.

Haziza, D. and Beaumont, J.-F. "Construction of weights in surveys: A review". *Statistical Science*, 32(2):206–226, 2017. ISSN 08834237, 21688745. URL `http://www.jstor.org/stable/26408226`.

Heffernan, J. E. and Tawn, J. A. "A conditional approach for multivariate extreme values (with discussion)". *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(3):497–546, 2004.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F. "A trend-preserving bias correction–the isi-mip approach". *Earth System Dynamics*, 4(2): 219–236, 2013.

Higham, N. J. "Computing the nearest correlation matrix—a problem from finance". *IMA journal of Numerical Analysis*, 22(3):329–343, 2002.

Horowitz, J. L. "A smoothed maximum score estimator for the binary response model". *Econometrica: journal of the Econometric Society*, pages 505–531, 1992.

Horowitz, J. L. "Bootstrap critical values for tests based on the smoothed maximum score estimator". *Journal of Econometrics*, 111(2):141–167, 2002.

Infante, L., Loschiavo, D., Neri, A., Spuri, M., and Vercelli, F. "The heterogeneous impact of inflation across the joint distribution of household income and wealth". *Bank of Italy Occasional Paper*, (817), 2023.

Joe, H. "Asymptotic efficiency of the two-stage estimation method for copula-based models". *Journal of multivariate Analysis*, 94(2):401–419, 2005.

Joe, H. *Dependence modeling with copulas*. CRC press, 2014.

Kaplan, G., Moll, B., and Violante, G. L. "Monetary policy according to hank". *American Economic Review*, 108(3):697–743, 2018.

Kleiber, C. and Kotz, S. *Statistical size distributions in economics and actuarial sciences*. John Wiley & Sons, 2003.

Kott, P. S. "Using calibration weighting to adjust for nonresponse and coverage errors". *Survey Methodology*, 32(2):133, 2006.

Kuhn, H. W. and Tucker, A. W. "Nonlinear programming". In *Traces and emergence of nonlinear programming*, pages 247–258. Springer, 1951.

Ledford, A. W. and Tawn, J. A. "Statistics for near independence in multivariate extreme values". *Biometrika*, 83(1):169–187, 1996.

Ledoit, O. and Wolf, M. "A well-conditioned estimator for large-dimensional covariance matrices". *Journal of multivariate analysis*, 88(2):365–411, 2004.

Limpert, E., Stahel, W. A., and Abbt, M. "Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question". *BioScience*, 51(5): 341–352, 2001.

Lindsay, B. G. "Composite likelihood methods". In *Statistical Inference from Stochastic Processes: Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference Held August 9-15, 1987, with Support from the National Science Foundation and the Army Research Office*, volume 80, page 221. American Mathematical Soc., 1988.

Lindskog, F., McNeil, A., and Schmock, U. "Kendall's tau for elliptical distributions". In *Credit risk: Measurement, evaluation and management*, pages 149–156. Springer, 2003.

Loschiavo, D., Mariani, G., Moscatelli, M., Neri, A., Porreca, E., and Tullio, F. "Enhancing the shiw weighting methodology with external data sources", 2024.

Lucy, L. B. "An iterative technique for the rectification of observed distributions". *Astronomical Journal, Vol. 79, p. 745 (1974)*, 79:745, 1974.

Lumley, T. "Analysis of complex survey samples". *Journal of statistical software*, 9: 1–19, 2004.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., et al. "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user". *Reviews of geophysics*, 48(3), 2010.

McDonald, J. B. and Xu, Y. J. "A generalization of the beta distribution with applications". *Journal of Econometrics*, 66(1-2):133–152, 1995.

Nelsen, R. B. *An introduction to copulas*. Springer, 2006.

Neri, A., Spuri, M., and Vercelli, F. "Combining survey and administrative data to estimate the distribution of household deposits". *Bank of Italy Occasional Paper*, (802), 2023.

Neri, A., Spuri, M., and Vercelli, F. "Distributional wealth accounts: methods and preliminary evidence". *Bank of Italy Occasional Paper*, (836), 2024.

Newey, W. K. and McFadden, D. "Large sample estimation and hypothesis testing". *Handbook of econometrics*, 4:2111–2245, 1994.

Patil, G. P. and Rao, C. R. "Weighted distributions and size-biased sampling with applications to wildlife populations and human families". *Biometrics*, pages 179–189, 1978.

Patton, A. J. "Modelling asymmetric exchange rate dependence". *International economic review*, 47(2):527–556, 2006.

Peyré, G., Cuturi, M., et al. "Computational optimal transport: With applications to data science". *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Potter, F. J. "A study of procedures to identify and trim extreme sampling weights". In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, volume 225230. American Statistical Association Washington, DC, 1990.

Puccetti, G. and Rüschendorf, L. "Computation of sharp bounds on the expected value of a supermodular function of risks with given marginals". *Communications in Statistics-Simulation and Computation*, 44(3):705–718, 2015.

Rao, J., Wu, C., and Yue, K. "Some recent work on resampling methods for complex surveys". *Survey methodology*, 18(2):209–217, 1992.

Rao, J. N. and Wu, C. "Resampling inference with complex survey data". *Journal of the american statistical association*, 83(401):231–241, 1988.

Reed, W. J. and Jorgensen, M. "The double pareto-lognormal distribution—a new parametric model for size distributions". *Communications in Statistics-Theory and Methods*, 33(8):1733–1753, 2004.

Richardson, W. H. "Bayesian-based iterative method of image restoration". *Journal of the optical society of America*, 62(1):55–59, 1972.

Rizzo, M. L. and Székely, G. J. "Energy distance". *wiley interdisciplinary reviews: Computational statistics*, 8(1):27–38, 2016.

Robertson, T., Wright, F., and Dykstra, R. *Order Restricted Statistical Inference*. Probability and Statistics Series. Wiley, 1988. ISBN 9780471917878. URL `https://books.google.it/books?id=sqZfQgAACAAJ`.

Rosenblatt, M. "Remarks on a multivariate transformation". *The annals of mathematical statistics*, 23(3):470–472, 1952.

Rust, K. F. and Rao, J. "Variance estimation for complex surveys using replication techniques". *Statistical methods in medical research*, 5(3):283–310, 1996.

Särndal, C.-E. "The calibration approach in survey theory and practice". *Survey methodology*, 33(2):99–119, 2007.

Scollnik, D. P. "On composite lognormal-pareto models". *Scandinavian Actuarial Journal*, 2007(1):20–33, 2007.

Self, S. G. and Liang, K.-Y. "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions". *Journal of the American Statistical Association*, 82(398):605–610, 1987.

Shepp, L. A. and Vardi, Y. "Maximum likelihood reconstruction for emission tomography". *IEEE transactions on medical imaging*, 1(2):113–122, 1982.

Shorrocks, A. and Wan, G. "Ungrouping income distributions: Synthesising samples for inequality and poverty analysis". Technical report, WIDER Research Paper, 2008.

Shorrocks, A. F. "Ranking income distributions". *Economica*, 50(197):3–17, 1983.

Sklar, M. "Fonctions de répartition à n dimensions et leurs marges". In *Annales de l'ISUP*, volume 8, pages 229–231, 1959.

Stephens, M. A. "Edf statistics for goodness of fit and some comparisons". *Journal of the American statistical Association*, 69(347):730–737, 1974.

Székely, G. J. and Rizzo, M. L. "Energy statistics: A class of statistics based on distances". *Journal of statistical planning and inference*, 143(8):1249–1272, 2013.

Themeßl, M. J., Gobiet, A., and Leuprecht, A. "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models". *International Journal of Climatology*, 31(10):1530–1544, 2011.

Tsukahara, H. "Semiparametric estimation in copula models". *Canadian Journal of Statistics*, 33(3):357–375, 2005.

Vardi, Y. "Nonparametric estimation in the presence of length bias". *The Annals of Statistics*, 10(2):616–620, 1982.

Vardi, Y. "Empirical distributions in selection bias models". *The Annals of Statistics*, pages 178–203, 1985.

Varin, C., Reid, N., and Firth, D. "An overview of composite likelihood methods". *Statistica Sinica*, pages 5–42, 2011.

Vercelli, F. "Financial wealth in italy: evidence from banking supervisory reports". *Bank of Italy Occasional Paper*, (865), 2024.

Vermeulen, P. "How fat is the top tail of the wealth distribution?". *Review of Income and Wealth*, 64(2):357–387, 2018.

Virkar, Y. and Clauset, A. "Power-law distributions in binned empirical data". *The Annals of Applied Statistics*, 8(1), mar 2014. ISSN 1932-6157. doi: 10.1214/13-aoas710. URL `http://dx.doi.org/10.1214/13-AOAS710`.

Vuong, Q. H. "Likelihood ratio tests for model selection and non-nested hypotheses". *Econometrica: journal of the Econometric Society*, pages 307–333, 1989.

Wang, J. C. "Sample distribution function based goodness-of-fit test for complex surveys". *Computational Statistics & Data Analysis*, 56(3):664–679, 2012.

Williamson, P., Birkin, M., and Rees, P. H. "The estimation of population microdata by using data from small area statistics and samples of anonymised records". *Environment and Planning A*, 30(5):785–816, 1998.

Wu, C. and Sitter, R. R. "A model-calibration approach to using complete auxiliary information from survey data". *Journal of the American Statistical Association*, 96(453):185–193, 2001.