BE-PARADIS
Working Paper wp.25.1

March 2025

Income Inequality in Belgium 1985-2022
New evidence from Distributional National Accounts

Bart Capéau, André Decoster, Nabil Sheikh Hassan,
Stijn Van Houtven, Jonas Vanderkelen, Kobe Wolfs

BE-PARADIS formulates a renewed and profound inquiry of existing and available data, concepts, and methods. This allows us:

- To highlight and quantify the relative importance of different drivers in the evolution of inequality and poverty.
- To align Belgium with the international research agenda and its output in the form of DINA's (Distributional National Accounts).
- To enlarge and deepen the conceptual framework of distributional analysis by going beyond mere household disposable income.

The project is funded by Belgian Science Policy, BELSPO.

www.beparadis.be

Income Inequality in Belgium 1985-2022
New evidence from Distributional National Accounts

Bart Capéau[*,‡], André Decoster[‡], Nabil Sheikh Hassan[†],
Stijn Van Houtven[‡] Jonas Vanderkelen[‡], Kobe Wolfs [‡]

This version: March 14, 2025

Abstract

In comparative studies on income inequality, Belgium frequently stands out as an exception: the level of inequality is low and has not increased over the last few decades, contrasting with the experiences of many other Western countries. In this paper, we apply the methodology of the Distributional National Accounts to reconsider the evolution of income inequality in Belgium between 1985 and 2022. Our findings underscore the impact of aligning the distributional information in microdata with National Accounts aggregates. We unveil a previously unobserved rise in income inequality in the aftermath of the financial crisis. Through a decomposition of the Gini coefficient into its constituent income components, it is ascertained that this rise in income inequality is predominantly attributable to a shifting composition of capital incomes. Specifically, the significance of earnings from savings accounts and other fixed-rate assets has diminished significantly in comparison to dividends and undistributed profits.

Keywords: Income inequality, Distributional National Accounts, income distribution

JEL codes: D31, E01

[*] ECARES, ULB

[†] IRES/LIDAM, UCLouvain

[‡] Department of Economics, KU Leuven

Corresponding author: André Decoster. E–mail: `andre.decoster@kuleuven.be`. We gratefully acknowledge financial support from BELSPO under contract number B2/191/P3/BE-PARADIS (for more details on the BE-PARADIS project, see www.beparadis.be). The present study utilizes data from the EU Statistics on Income and Living Conditions (EU-SILC) for Belgium, administrative data from personal income tax records (IPCAL), which were made available by the Belgian statistical office Statbel, and the Household Finance and Consumption Survey (HFCS), which was made available by the European Central Bank. We would like to express our sincere gratitude to Facundo Alvaredo for his advice on constructing DINA, to Patrick Lusyne and Toon Vanheukelom for their assistance with data access and processing, and to the participants of the BE-PARADIS workshops in Leuven (May 2022) and Brussels (December 2024), as well as the 1st Welfare and Policy Conference (WAP) at the Bordeaux School of Economics (May 2023) for valuable comments and suggestions.

1 Introduction

The discourse surrounding income inequality in Belgium is marked by a dichotomy between the prevailing sentiment among the general public, which suggests an increase in inequality, and the paucity of compelling empirical evidence to substantiate this assertion. According to the findings of the OECD's 2020 *Risk that Matter*-survey, 56% of Belgians perceive an increase in income inequality over the past decade (OECD, 2021). However, as summarised by Assal et al. (2023), most of the existing empirical evidence on income inequality in Belgium shows a rather stable trend, making Belgium one of the exceptions compared to most other OECD countries where income inequality has increased over the last decades (OECD, 2011, 2015). These findings are primarily based on after-tax and transfer household disposable income, collected through surveys designed to be as representative as possible of the Belgian population. Notably, income surveys, such as the EU Statistics on Income and Living Conditions (EU-SILC), have played an instrumental role in investigating the evolution of income distribution over the past decades. In addition to providing detailed income information, these surveys also offer a comprehensive set of sociodemographic covariates, which are of significant importance to get insight in the evolution of this income distribution.

But these surveys are not without their limitations. They are subject to unit nonresponse, item nonresponse, and misreporting, particularly at the upper end of the income distribution. Furthermore, there is a discrepancy between the total income aggregate recorded in household surveys and the macroeconomic concept of net national income (hereafter, NNI) recorded by the System of National Accounts. Consequently, extant evidence on income inequality is based on a segment of the total net national income 'pie', available for distribution among residents in a given country. However, reliance on aggregate information from the National Accounts themselves is, for obvious reasons, a non-starter for distributional analysis.

To address this shortcoming, a novel methodology for distributional analysis has been proposed that seeks to integrate the strengths of both approaches. The methodology involves using national accounts to determine the size of the pie to be distributed, supplemented by distributional information from ancillary sources such as surveys and administrative tax records. The impetus for the integration of surveys into national accounts came initially from the late Tony Atkinson (see, for example, Atkinson et al., 1995 section 3.6). However, the detailed elaboration and rapid dissemination of this approach is largely due to the efforts of the *World Inequality Lab* (hereafter WIL), led by Atkinson's student Thomas Piketty.

The trigger for this resurgence of interest in Kuznets' seminal contributions arose from the realization that two major developments, the increasing accessibility of survey data, compounded by the rapid advancements in personal computing power, had effectively pushed Kuznets' original contributions somewhat into oblivion. Indeed, in a seminal paper, Kuznets and Jenks (1953) integrated national accounts data with tabulated income tax records to assess the evolution of the top income share in the United States from 1913 to 1948. The resurgence of interest in national accounts data for informing distributional analysis can be regarded as a re-engagement with the origins of this early empirical distributional analysis.

In practice, Atkinson and Piketty (2007, 2010) began with the construction of top income series for an ever-increasing number of countries, culminating in the establishment of the World Top Incomes Database (WTID). However, the project rapidly evolved beyond its initial exclusive focus on the upper tail, metamorphosing into a concerted endeavour to provide estimates of the entire income and wealth distribution. This ambitious research initiative has culminated in the formation of an international research network, known as the World Inequality Lab, and the establishment of a continually expanding and updated World Wealth and Income Database. (hereafter WID, available online: see https://www.wid.world/), an annual report on inequalities (e.g. Chancel et al., 2022), and, most significantly, a unified methodology to construct what is now commonly known as *Distributional National Accounts* (hereafter, DINA).

The endeavour to implement a uniform methodology has evidently been inspired by the standardisation of the national accounts, with the primary objective being to enhance the international comparability of the outcomes. Consequently, this methodology, which is comprehensively described in Alvaredo et al. (2024), constitutes the backbone of this paper. We aim to provide additional evidence on the distribution of total income, as measured by NNI, in Belgium by bridging the gap between the micro- and macro-levels of income information. The methodology involves the linking of the Belgian national accounts from 1985 onwards with survey data from the Socio-Economic Panel (SEP), the European Community Household Panel (ECHP), and the EU Statistics of Income and Living Conditions (EU-SILC). The approach generates distributional information regarding the national account aggregates for Belgium since 1985.

The growing World Inequality Lab-community fostered a rapid propagation of the DINA-methodology. Important applications, among others, are the ones for the United States (Piketty et al., 2018) and France (Garbinti et al., 2018). But also China (Alvaredo et al., 2017), Russia (Novokmet et al., 2018), India (Chancel & Piketty, 2019) and the Middle East (Alvaredo et al., 2019) are covered. Blanchet, Chancel, and Gethin (2022) constructed DINA for 26 European countries, among which also Belgium. This remarkable productivity in the WIL in Paris did not hinder individual countries from delving deeper into their country-specific DINA series. On the contrary, by making all assumptions explicit and as transparent as possible, the WIL laid the foundation stone for this more detailed investigation of country-specific DINA series.[1]

Recent examples of these country-specific studies include Austria (Jestl & List, 2023), Germany

[1] It should also be noted that, in addition to the DINA approach of the WIL, there have been other attempts to combine macroeconomic data with micro data exist in parallel, such as for the US (Fixler et al., 2019) or the EG DNA approach coordinated by the OECD (Zwijnenburg et al., 2021). The latter initiative bears many similarities, yet a significant distinction emerges in the income concept for which they compile distributional results. The EG DNA approach exclusively focuses on the income of the household sector (S14), whereas the DINA approach encompasses income from all sectors.

(Bach et al., 2024), Italy (Guzzardi et al., 2023) and the Netherlands (Bruil et al., 2022). While these studies may appear to be a replication of the WIL-work, they all adopt a more refined approach based on thorough research into and knowledge of the underlying microdata. They produce new DINA series by, for example, using additional (simulated) microdata to distribute some macro aggregates at a more granular level than in the WIL-approach. A notable finding is the discrepancy between both the level and the trend of the inequality figures derived from these augmented DINA series and those reported in the study by Blanchet, Chancel, and Gethin (2022).

This paper contributes to this line of research by producing DINA series for Belgium. In a first step, the objective is to produce a consistent series starting in 1985, utilising all available income surveys for Belgium. Previous results of this research have been reported in Capéau, Decoster, Sheikh Hassan, and Vanderkelen (2023), in which the series start in 1995, and in Capéau et al. (2024). In contrast to the series for Belgium produced by Blanchet, Chancel, and Gethin (2022), this study has relied on more detailed national accounts data on benefits and taxes, and tax-benefit microsimulation models have been employed to determine a detailed picture of tax liabilities and benefits at the micro-level of the household. The results obtained with the DINA methodology challenge the conjecture about low and stable inequality in Belgium, which has increased since the financial crisis. We also highlight the role of income from financial capital in this phenomenon.

The second contribution of this paper is to provide a more in-depth insight into the role of capital income in driving the evolution of income inequality. In order to enhance the distributional information available in income surveys, additional microdatasets and statistical techniques that are now commonly utilised in the relevant literature are employed. Firstly, the top earnings distribution is adjusted using administrative data from personal income tax records, following the method outlined in Blanchet, Flores, and Morgan (2022). Secondly, we employ a Predictive Mean Matching (PMM) imputation method to augment EU-SILC with more reliable and detailed information on capital (incomes) using the Household Finance and Consumption Surveys (HFCS) in combination with a Pareto model to correct for the missing wealthy at the top of the wealth distribution. We then rely on a decomposition of the Gini coefficient by income source to investigate the role of capital income. This decomposition reveals that capital income inequality is a driving factor behind the observed increasing inequality trend. The fixed-interest incomes, which are more prevalent across the population, have undergone a sharp decline, while more unequally distributed components (such as dividends and undistributed profits) have experienced a corresponding increase.

The structure of the paper is as follows. Section 2 provides a detailed examination of the national accounts, with a particular focus on the evolution of net national income and the various underlying income components. Section 3 then discusses the DINA methodology, i.e. the process of incorporating distributional information into the national accounts. This section also provides an explanation of the various datasets used to obtain the necessary distributional information. The subsequent Section 4 provides a summary of the findings on income inequality in Belgium between 1985 and 2022. To ensure a consistent coverage of this extended period, the results in this section are based exclusively on the distributional information from income surveys. In contrast, Section 5 leverages the recent availability of administrative tax records and wealth surveys to enhance the precision of the distributional information. As these data sources are more recent, the results in this section commence from 2009 onwards.

2 Levels and trends of aggregate income concepts in the National Accounts

The essence of the DINA approach is to distribute all of an economy's income, not just the portion that is captured in household surveys through respondents' answers. The income aggregate that is distributed is net national income (*hereafter* NNI, code $B.5_{n,S1}$ in the national accounts).[2] In this section we first describe the evolution of NNI compared to GDP (subsection 2.1). Then we decompose the evolution of NNI into the five institutional sectors (subsection 2.2) and into different income components (subsection 2.3). In subsection 2.4 we explain how NNI can be studied at different stages of the macroeconomic circular flow of income, leading to four different income concepts in the DINA methodology. By default, we refer to NA aggregates in real terms by deflating the nominal values with the GDP deflator with base year 2015. When we use nominal values, we mention them explicitly.

2.1 From Gross Domestic Product to net national income

net national income (NNI) differs from gross domestic product (GDP) by subtracting depreciation (from "gross" to "net") and adding net foreign income (NFI) (from "domestic" to "national"). In Figure 1 we show the cumulative growth - in the form of an index starting at 100 in 1985 - of GDP, NNI and depreciation.[3] Unless explicitly stated, we refer to both GDP and NNI as measured at 'market prices', in the sense that we include net taxes on production and imports in both aggregates.[4]

Figure 1 and Table 1 show that over the period considered (1985 to 2022) the growth rate of NNI is

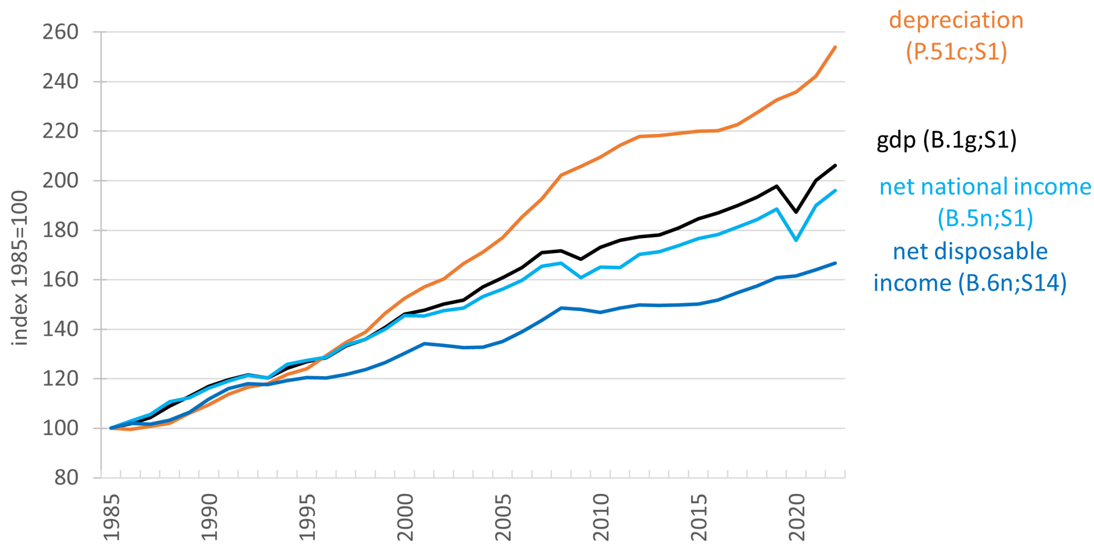
[2]We use the standard NA codes when referring to specific aggregates. Most codes consist of a letter followed by a number. We use subscripts to denote net ('n') or gross ('g'), resources ('R') and uses ('U'), and the institutional sector. The five institutional sectors in the NA are: the non-financial corporate sector 'S11', the financial corporate sector 'S12', the government sector 'S13', the household sector 'S14' and the NPISH sector 'S15'. The total domestic economy is referred to as 'S1' and the rest of the world as 'S2'. A brief introduction to the concepts of national accounts can be found in the Appendix section A.

[3]The Belgian national accounts follow the ESA 2010 standard and in this standard they are publicly available from 1995 onwards. We have used the national accounts in the ESA 1994 standard to extrapolate these series back to 1985.

[4]The main reason for including net taxes on production and imports in NNI is that indirect taxes appear as the main component - next to net property income, which is often negative - of the net primary income balance of the government sector S13 ($B.5_{n,S13}$). Therefore, in order to allocate the total NNI, net indirect taxes must be included. In essence, these indirect taxes only introduce a wedge between factor prices (or basic prices) and consumer prices and do not affect the evolution of NNI or GDP in volume terms. Of course, there may be a distributional impact of price changes, and therefore we need to pay attention to the distribution of net indirect taxes among households (see subsection 3.2, and Alvaredo et al., 2024 p. 56-59 for an in depth discussion).

lower than that of GDP (1.8% compared to 2.0%). The main reason for this is that depreciation grows at an annual rate of 2.6%, thus increasing the share of depreciation in GDP: from 15.9% of GDP in 1985 to 19.6% in 2019. It is worth noting that carrying out a distributional analysis on NNI implies neglecting possible distributional effects of this increasing importance of depreciation.

Figure 1: evolution of main aggregates in national accounts, 1985-2022



Note: variables are deflated with the GDP-deflator with base year 2015.

The role of the second element in the difference between GDP and NNI, net foreign income, is less clear. Overall, the share of NFI in GDP fell from 2.1% to 1.5%, with average annual growth barely half that of GDP (1.1% compared to 2.0%). But the annual movements are highly volatile, with, for example, a fall - in nominal terms - from €5.6 billion in 2010 to €602 million in 2011 and a rebound to almost €9 billion in 2012.[5] By using NNI as the reference income to be distributed, we include NFI in the distributional analysis.

Finally, Figure 1 and the bottom row of Table 1 also provide information on the NA concept that is closest to - but not identical with - the income concept most often used to present a distributional picture based on survey information: household disposable income. The annual growth rate of net disposable income of households (in real terms) is even lower than the annual growth rate of NNI: 1.4% compared with 1.8% for NNI and 2.0% for GDP, which is reflected in a decline in the share of disposable income of households in GDP from 64.2% to 52.2%.

Apart from the fact that, with disposable income, we have moved from a pre-tax to a post-tax and transfer concept, the difference is also due to the fact that for disposable income of households we restrict the income measurement to the household sector (labeled S14 in the NA). In fact, important parts of NNI, such as undistributed profits, do not reach households: in 2022, net disposable income

[5]This high volatility is the main reason why we omit NFI from Figure 1.

Table 1: gdp, nni, depreciation and household disposable income, 1985-2022

	levels in bn €		in % of gdp		average annual growth (%)
	1985	2022	1985	2022	1985-2022
Gross domestic product (S1)	225.7	465.4	100.0	100.0	2.0
Depreciation (S1)	35.9	91.3	15.9	19.6	2.6
Net foreign income	4.8	7.2	2.1	1.5	1.1
net national income (S1)	194.6	381.3	86.2	81.9	1.8
Net disposable income (S14)	145.8	243.1	64.6	52.2	1.4

Note: incomes in levels are deflated with the GDP-deflator with base year 2015.
Source: own calculations based on download from Nbb.Stat on 19.02.2024.

of households amounts to €243 bn or 63.8% of the €381 bn of NNI.

2.2 net national income by institutional sector

To further emphasize the importance of this, we now turn to a decomposition of NNI into its constituent income components. Above, we defined NNI starting from the production of value added as recorded in account $II.1.1$ (the *Generation of income account* in the NA) and applying the operations to move from gross to net and from domestic to national. But to study distributional issues on the basis of the NA, it is more useful to start with the NNI as the sum of all net primary incomes 'received' as resources by all five institutional sectors and registered with code $B.5_n$ in account $II.1.2$, the *Allocation of primary incomes account*. Even before the allocation of these income components to households or individuals in a genuine DINA exercise, changes in the relative importance of the five institutional sectors may already indicate divergent developments of the different income components of NNI.

Table 2: Decomposition of NNI into primary income of the institutional sectors

	levels in bn €		in % of gdp		average annual growth (%)
	1985	2022	1985	2022	1985-2022
Corporate non financial (S11)	11.5	41.3	5.9	10.8	3.5
Corporate financial (S12)	3.4	6.9	1.8	1.8	1.9
Government (S13)	0.8	37.4	0.4	9.8	11.1
Households (S14)	178.9	295.5	91.9	77.5	1.4
Non-profit (S15)	0.03	0.17	0.02	0.04	5.0
net national income (S1)	194.6	381.3	100.0	100.0	1.8

Note: incomes in levels are deflated with the GDP-deflator with base year 2015.
Source: own calculations based on download from Nbb.Stat on 19.02.2024.

In Table 2 we decompose NNI into primary income - that is, before any redistribution through taxes or transfers has taken place - of the five institutional sectors. Overall, the share of primary income of the household sector in NNI shrinks from 91.9% to 77.5% between 1985 and 2022. The first explanation for this declining share lies in the growth rate of primary income in the corporate sector, which is much higher than that of the household sector: 3.5% (S11) and 1.9% (S12) against 1.4% for households (S14). Especially in the period after the financial crisis (2010-2022), this divergence in growth between primary income of households (0.8%) and that of the non-financial corporate sector (6.1%) is striking. In 2022, the primary income of the corporate sector (sectors S11 and S12 combined) amounts to € 48.2 bn. These are mainly undistributed profits, the allocation of which among households will play an important role in the distributional analysis below. The second explanation for the shrinking share of the household sector in the primary income aggregate NNI is the significant increase in the share of the government sector in total primary income. This is due to the decline in interest payments by the government sector.[6] In any case, considering only the primary income of households leaves increasingly important parts of national income outside the scope of the analysis: the gap between net primary income of households and NNI increases from € 15.7 bn in 1985 to € 85.8 bn in 2022.

2.3 net national income by income component

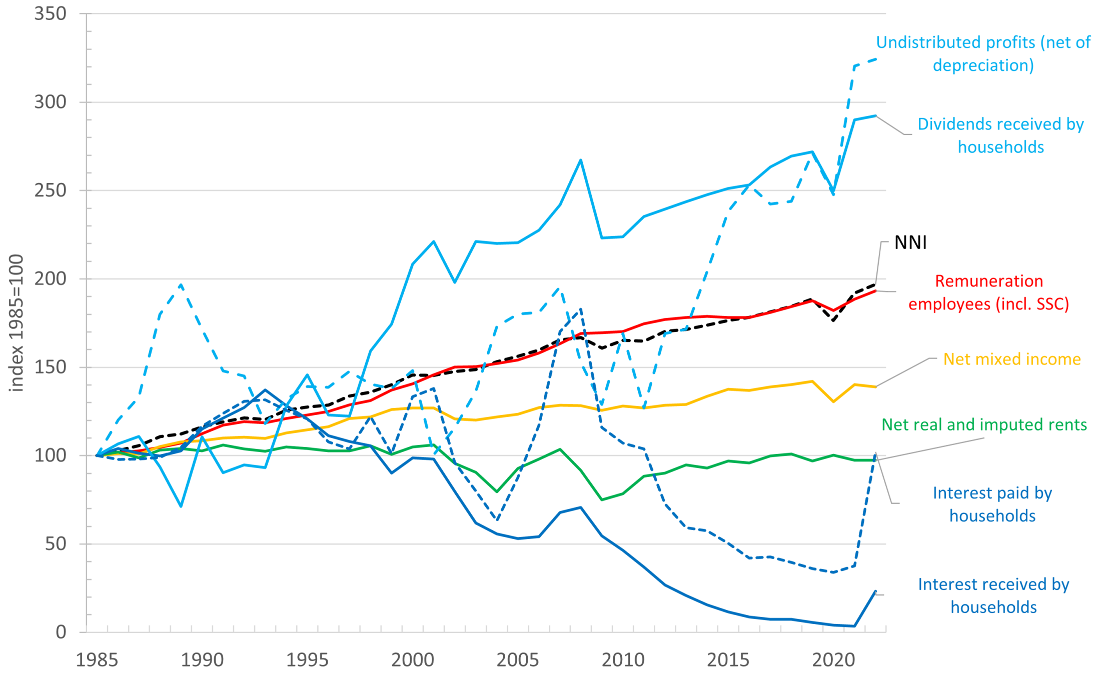
Figure 2 and Table 3 make the above two findings more explicit by breaking down NNI into separate income components.[7] In Figure 2 we show the cumulative growth of NNI as the black dotted line, to compare with the other lines for the different components of NNI. The corresponding annual growth rates, for the whole period and for two subperiods, are shown in Table 3. We decompose NNI into labour income, mixed income for the self-employed, and various forms of capital income. In the latter, we distinguish income from real and imputed rents, financial income (in the form of interest, dividends, and other compensations), and undistributed profits. The bottom line, labeled "other income", consists of the balance of net primary income of the government sector $S13$. Since the incomes in the lines above, taken from account II.1.2, are net of (net) indirect taxes, we have to add them to arrive at the total NNI measured at market prices (see footnote 4 and footnote 6).

Over the whole period, employee compensation has grown at about the same pace as NNI (1.8%), resulting in a more or less stable share of this income component in NNI (61.8% in 1985 to 60.7% in 2022). This high share will translate into an important role of changes in the distribution of compensation of employees to explain changes in the overall level of inequality. Note that the stability of this income share over the whole period masks a different pattern before and after the

[6] Although the net operating surplus of the government sector ($B.2_{n,S13}$) is close to zero, the balance of primary incomes $B.5_{n,S13}$ in account $II.1.2$ is not. The resource side of $S13$ consists mainly of net indirect tax receipts (see footnote 4); the expenditure side consists of interest paid on government debt. The spectacular decline in the latter, which is not matched by a corresponding decline in revenues, is the driving force behind the huge increase in the net primary income balance of $S13$.

[7] The breakdown in Figure 2 and Table 3 is a simplified version of a more detailed breakdown of NNI available for 2022 in Appendix Section D.

Figure 2: evolution of main income components of national income, 1985-2022



Note: all income components are deflated with the GDP-deflator with base year 2015.

Table 3: main income aggregates in net national income, 1985-2022

		levels in bn €		in % of NNI		average annual growth (%)		
		1985	2022	1985	2022	1985 - 2022	1985 - 2007	2010 - 2022
+	Remuneration employees	120.3	231.6	61.8	60.7	1.8	2.3	1.0
+	Net mixed income	19.7	27.2	10.1	7.1	0.9	1.1	0.6
+	Net real & imputed rents	9.5	9.2	4.9	2.4	-0.1	0.2	1.8
+	Financial income received by househ.	29.4	27.5	15.1	7.2	-0.2	0.7	-1.0
	+ Interests	21.9	5.1	11.2	1.3	-3.9	-1.8	-5.7
	+ Dividends	5.8	17.0	3.0	4.5	2.9	4.1	2.2
	+ Other financial income	5.5	9.3	2.8	2.4	1.4	3.4	-2.1
−	Interests paid by households	3.8	3.9	2.0	1.0	0.1	2.4	-0.4
+	Net undistributed profits	14.9	48.2	7.7	12.6	3.2	3.1	5.6
+	Other income	0.8	37.6	0.4	9.9	11.1	17.5	2.6
=	net national income	194.5	381.3	100.0	100.0	1.8	2.3	1.4

Note: incomes in levels are deflated with the GDP-deflator with base year 2015. 'Net' refers to net of depreciation. 'Other income' primarily consists of net primary income of the government sector (S13) and is mainly driven by net indirect taxes minus interest payments, as explained in footnote 4 and footnote 6.

financial crisis. For the period 2010-2022, the growth of compensation of employees is significantly lower than that of NNI: 1.0% versus 1.4%.

In contrast to this stability in the share of compensation of employees, the average annual growth of 1.8% for NNI masks strikingly divergent developments for most of the other components of income. First, the share of *financial income* received by households in total NNI more than halves from 15.1% in 1985 to 7.2% in 2022. The decline in financial income from €29.4 bn in 1985 to €27.5 bn in 2022 results in a slightly negative average annual growth of -0.2%. Figure 2 clearly visualizes that this decline in financial income is driven by the steeply declining curve of 'interest received by households'. In fact, this form of income almost disappeared by 2020 (less than €1 bn) and only recently started to grow again. The average annual growth of interest was negative up to -3.9%, with an even sharper decline after the financial crisis (-5.7%).[8] Note that the other side of the coin, namely the interest paid by households, looks somewhat different. Especially in the years when the interest rate rose (2000-2001 and 2007-2008), the increase in 'interest paid' is spectacular. In absolute terms, for example, interest paid rose from €2.0 bn in 2004 to €6.3 bn in 2008, only to fall back to €4.0 bn in 2009.

The spectacular fall in interest received differs from the evolution of other forms of household financial income. The share of dividends received by households rose from 3.0% to 4.5% of NNI (or from €5.8 bn to €17.0 bn), resulting in a growth rate much higher than that of NNI (2.9% against 1.8%). The share of other financial income, which consists mainly of investment income attributable to insurance policyholders or collective investment fund shareholders and income from pension rights, remained more or less constant. The *composition* of the financial income received by households has thus undergone a fundamental change during the period under review.

Second, Table 3 and Figure 2 show a very high growth rate of *undistributed profits.* This component of NNI tripled from €14.9 bn to €48.2 bn and its share in NNI increased from 7.7% to 12.6%. Especially in the aftermath of the financial crisis, the annual growth rate of undistributed profits of 5.6% far exceeded the growth of NNI. Figure 2 also shows that this component is the most volatile of all income components. Unlike distributed profits in the form of dividends, these undistributed profits do not explicitly trickle down to households. However, since they are integrated into the comprehensive distributional picture produced by DINA, their sharp increase can be expected to have an impact on inequality.

Third, the negative growth of real and imputed rents (-0.1% compared to an average growth of NNI of 1.8%) is somewhat surprising. The often cited rising prices of residential property do not seem to be reflected in correspondingly higher net imputed rents for homeowners, at least not after taking depreciation into account. This has to do with the rapidly increasing depreciation rate already mentioned in Table 1. Accelerating depreciation also explains why the average annual growth of net mixed income is lower than that of NNI (0.9% compared to 1.8%).

[8]From 2010 to 2020, the average annual growth was about -21.4%.

Finally, ‘other income’, which consists mainly of the balance of net primary income of the general government sector S13, shows the highest annual growth rate. The increase from €0.8 bn in 1995 to €37.6 bn in 2022 is driven by two forces. On the one hand, it is partly the flip side of the coin of declining interest payments received by households, since this steep decline is obviously also reflected in declining interest payments on government debt, with a positive effect on the primary income balance of S13. On the other hand, it also reflects the increasing reliance on net production taxes as a source of government revenue.

We conclude that important changes have taken place in the composition of net national income. Undistributed profits play a much more important role in NNI, certainly since the financial crisis. Households’ financial income now consists mainly of dividend income rather than interest. We expect these structural changes to have non-negligible distributional implications.

2.4 Four DINA-concepts of net national income

In line with the DINA methodology of WIL (Alvaredo et al., 2024), we focus on net national income as the focal concept for the cake to be distributed among the inhabitants of a country. However, the same NNI appears at different stages in the sequence of the transition from the pre-tax market income paid to the factors of production to the disposable income of households, supplemented by the in-kind benefits of public consumption. We summarise the different possibilities in a simplified form in Table 4.[9]

Table 4: Four different DINA income concepts

	All incomes going to the production factors labour and capital
+	All other income not directly attributable to labour or capital (government income $B.5_{n,S13}$ and non-profit sector income $B.5_{n,S15}$)
=	**Pre-tax factor income (DINA income concept 1)**
−	social insurance contributions
+	social benefits acting as income replacement (pension, unempl., sickness and invalidity benefits)
=	**Pre-tax post-replacement income (DINA income concept 2)**
−	all forms of taxes on income (personal income taxes, corporate income taxes ...)
+	all remaining social benefits (e.g. social assistance, child benefits ...)
=	**Post-tax disposable income (DINA income concept 3)**
+	social benefits in kind (e.g. expenditures on health, education ...)
+	collective consumption (e.g. expenditures on defense, police, general public services ...)
+	government balance
=	**Post-tax national income (DINA income concept 4)**

[9] See the left part of tables A.9 to A.12 in the Appendix section D for detailed information on how the four DINA concepts are derived from the national accounts.

There are two pre-tax concepts, *pre-tax factor income* (hereafter also referred to as DINA1) and *pre-tax replacement income* (hereafter also referred to as DINA2), and two post-tax concepts, *post-tax disposable income* (hereafter DINA3) and the *post-tax national income* (hereafter DINA4). Except for DINA3, all DINA concepts aggregate to net national income ($B.5_{n,S1}$) measured at market prices. In the following subsections the four perspectives are explained in more detail and the analysis of the evolution of the underlying components over time is continued.

Concept DINA1: Pre-tax factor income

Since *Pre-tax factor income* (or DINA income concept 1) is constructed from aggregates behind balances of primary income of each institutional sector of the economy, we have already described its evolution and the main components in subsection 2.3 (see Table 3 and Figure 2). This income concept is a pre-tax concept. This implies that there is no redistribution of any kind at this stage: no direct tax, no social insurance contribution, no cash benefit, and no use of public spending. Note, however, that since we have chosen to measure NNI at market prices, net indirect taxes (on production and imports) are included.[10]

As there is no income tax or redistributive activity in pre-tax factor income, this concept is close to factor income. As the DINA exercise distributes this aggregate over a population of households or individuals, people who derive their income mainly from benefits (such as many pensioners) appear at this stage with a very low pre-tax factor income. Because of the high rate of home ownership in Belgium, many pensioners have at least a limited amount of capital income in the form of (imputed) rents, but their labour income is close to zero.[11]. Distributional analysis based on DINA1 is therefore sensitive to the age structure of the population, since, other things being equal, the ageing of the population will be reflected in a growing proportion of the population with pre-tax labour income close to zero. Therefore, the description of the evolution of pre-tax factor inequality is often limited to the sub-population of potentially active individuals, e.g. between 20 and 65 years of age.

Concept DINA2: Pre-tax post-replacement income

Pre-tax post-replacement income is the second DINA income concept. It is based on pre-tax factor income by subtracting social contributions and adding some social benefits. Table 5 shows in a simplified way the construction of pre-tax post-replacement income or DINA2 with the level and relative weight of its main aggregates in 1985 and 2022 and the annual growth rates in the right part of the table.[12]

[10]In Table 3, these net indirect taxes are listed under the label 'other income' in the bottom row, since this 'other income' consists mainly of the balance of the net primary income of the government sector $S13$, of which net indirect taxes are an important element.

[11]In the national accounts, imputed rents from owner-occupied housing belong to the operating surplus of the household sector, denoted as $B.2_{n,S14}$

[12]The left side of Table A.10 in Appendix Section D explains in detail how pre-tax post-replacement income is constructed from the National Accounts.

Table 5: Main components of pre-tax post-replacement income - 1985 and 2022

		levels in bn €		in % of NNI		average annual growth (%)
		1985	2022	1985	2022	1985 - 2022
Pre-tax factor income (= NNI)		194.5	381.3	100.0	100.0	1.8
−	Social contributions ($D61_{U,S14}$)	43.0	85.9	22.1	22.5	1.9
+	Cash social security benefits ($D621_{R,S14}$)	28.1	55.6	14.4	14.6	1.9
+	Other social benefits ($D622_{R,S14}$)	10.5	21.5	5.4	5.6	2.0
+	Balance social contrib. and repl.inc.*	4.4	8.8	2.3	2.3	1.9
=	Pre-tax post-replacement income (=NNI)	194.5	381.3	100.0	100.0	1.8

Notes: incomes in levels are deflated with the GDP-deflator with base year 2015. See Table A.10 in Appendix for a detailed breakdown of all income components.

* This balance has no real-world counterpart in the financial accounts of the social security administration. It is the balance obtained from subtracting all benefits retained in this table, from all revenues retained in this table, such that in total we are back at net national income.

Source: own calculations based on download from Nbb.Stat on 19.02.2024.

The social benefits included in this pre-tax psot-replacement income are mainly labour-related benefits: pensions, unemployment benefits, sickness and disability benefits and career break benefits. In this sense, pre-tax post-replacement income attempts to approximate a concept of income that takes into account the aspects of the safety net that are mainly motivated by insurance arguments, to be distinguished from explicitly redistributive motives.[13] Child allowances, for example, are therefore only included in the perimeter of the third DINA income concept.[14] Note that in the bottom line of Table 5 we also add the gap (positive or negative) between total contributions and total benefits, so that pre-tax psot-replacement income equals total NNI.

Concept DINA3: Post-tax disposable income

The third DINA income concept, *post-tax disposable income* or DINA3, is a post-tax and transfer concept: starting from *Pre-tax post-replacement income*, it subtracts all remaining taxes (such as direct taxes on income, property or corporate profits, etc., but also indirect taxes on production and imports) and adds all remaining cash social benefits (such as child allowances).

[13] Obviously, the empirical division between redistributive and insurance elements in social security contributions and benefits is less straightforward than the neat conceptual distinction suggests. Many explicit redistributive elements, such as minimum floors and caps on the benefit side, undermine the insurance element of replacement incomes such as unemployment benefits, pensions or sickness and disability benefits. Allocating them to pre-tax post-replacement income, because it precedes the explicit redistributive role of the welfare state, is therefore a simplification.

[14] In Belgium, child allowances were included in the social security perimeter before their transfer to the regions in 2015. Therefore, in the national accounts up to 2014, they were included in $D621_{R,S14}$ *social security cash benefits* and $D622_{R,S14}$ *other social benefits*. We have removed child allowances from $D621_{R,S14}$ and $D622_{R,S14}$ and added them to $D623_{R,S14}$, *cash social benefits*.

Part of government revenue is not spent on benefits, but on 'in kind' benefits such as education or subsidised health care, and on collective services such as general government administration, police, defence, etc. As this government expenditure is not yet added - this will be done in the fourth DINA concept of income - this third DINA aggregate is no longer equal to NNI. In Table 6 this can be seen by comparing the bottom line (post-tax disposable income) with the top line (pre-tax post-replacement income or NNI). In 2022, post-tax disposable income is €277.1 bn, or €104 bn less than the NNI of €381.3 bn.[15]

Table 6: Main components of post-tax disposable income - 1985 and 2022

		levels in bn €		in % of NNI		average annual growth (%)
		1985	2022	1985	2022	1985 - 2022
Pre-tax post-replacement income (= NNI)		194.5	381.3	100.0	100.0	1.8
−	Balance social contrib. and repl.inc.	4.4	8.8	2.3	2.3	1.9
−	Taxes on production and imports ($D2_{R,S13}$)	27.0	58.9	13.9	15.4	2.1
−	Subsidies on products and production ($D3_{R,S13}$)	-6.4	-17.7	-3.3	-4.6	2.8
−	Taxes on income ($D51_{R,S13}$)	38.5	74.4	19.8	19.5	1.8
−	Other current taxes ($D59_{R,S13}$)	0.9	2.1	0.5	0.5	2.1
+	social assistance benefits in cash ($D623_{R,S14}$)*	7.1	13.9	3.7	3.6	1.8
=	Post-tax disposable income (≠ NNI)	137.2	268.7	70.5	70.5	1.8

Notes: incomes in levels are deflated with the GDP-deflator with base year 2015. See Table A.11 in Appendix for a detailed breakdown of all income components.

* Child allowances in 1985 were removed from $D621_{R,S14}$ and $D622_{R,S14}$ in Table 5, and added to $D623_{R,S14}$ in this table, as explained in footnote 14.

Source: own calculations based on download from Nbb.Stat on 19.02.2024.

Concept DINA4: Post-tax national income

The fourth and final DINA income concept, DINA4 or *Post-tax national income*, brings us back to NNI. As mentioned above, 'benefits in kind' were not included in DINA3. DINA4 starts from DINA3 and includes this expenditure. A distributional analysis based on post-tax national income will therefore require assumptions about how these public expenditures are distributed.[16] We present a simplified version of the construction in Table 7. Note that since total government revenue and

[15] Given its construction, it is tempting to compare the concept of post-tax disposable income with the survey concept of 'household disposable income', the latter being the income left to households after they have paid taxes on their income and received benefits. However, it should be noted that the scope of this third DINA income concept is wider than that of the household sector in the national accounts (S14) and should therefore not be read as 'disposable income of sector S14' in the national accounts.

[16] In national accounts, a distinction is made between, on the one hand, government expenditures that can still be considered as 'personal' or 'individualised' (e.g. education or health care), which are called 'social transfers in kind', and, on the other hand, the so-called pure public goods such as government administration or defence, which are called 'collective consumption'.

expenditure do not necessarily balance, and since we want to distribute total NNI, we add a line at the bottom of the table with the residual and label it *government surplus/deficit*.

Table 7: Main components of post-tax national income - 1985 and 2022

		levels in bn €		in % of NNI		average annual growth (%)
		1985	2022	1985	2022	1985 - 2022
Post-tax disposable income ($\neq$ NNI)		137.2	268.7	70.5	70.5	1.8
+	Individ. consumption expenditures ($P31_{U,S13;S15}$)	31.7	77.7	16.3	20.3	2.4
+	Collective consumption expenditures ($P32_{U,S13}$)	23.5	38.4	12.1	10.1	1.3
+	Government surplus/deficit	2.1	-3.3	1.1	-0.9	
=	Post-tax national income (NNI)	194.5	381.3	100.0	100.0	1.8

Notes: incomes in levels are deflated with the GDP-deflator with base year 2015. See Table A.12 in Appendix for a detailed breakdown of all income components.
Source: own calculations based on download from Nbb.Stat on 19.02.2024.

3 Data and methodology

The distributional national accounts methodology aims to distribute the total pie of net national income over the whole population. The distributional information used to divide up the cake is obtained from microdata such as household surveys and administrative tax data. The section 3.1 gives a brief description of the microdata used and explains how we have combined different microdata sets using statistical matching techniques. More details can be found in Section C in Appendix. Section 3.2 explains how we have matched these data to aggregate information from national accounts. Further details can be found in Section D in Appendix.

3.1 Distributional information from microdata

In recent decades, income surveys have become an invaluable tool for studying changing patterns of income distribution within a country. The EU Statistics on Income and Living Conditions (EU-SILC) is the most widely used dataset in Europe to assess the economic and social situation of households in Europe. For Belgium, this survey has been carried out annually since 2004 by the Belgian Statistical Office (Statbel). It asks for detailed information on a wide range of socio-economic variables, including several income variables, making it well suited for measuring income inequality, poverty and other socio-economic indicators. We use the EUROMOD arithmetic microsimulation model to simulate taxes and benefits. In addition, we extend the distributional backbone of our DINA series with two other surveys that preceded EU-SILC for monitoring inequality and poverty in Belgium: the Socio-Economic Panel (1985, 1988, 1992 and 1997) and the European Community Household Panel (ECHP) for the period 1993-2000. More detailed information on these three income surveys can be found in Section B in Appendix.

We adjust the income surveys by applying techniques that are now commonly used in the literature. A first well-known limitation of income surveys, especially in the context of inequality measurement, is the under-representation of individuals and households at the top of the income distribution (Ravallion, 2022). This is partly due to the imperfect participation of those included in the sample (i.e. unit non-response and item non-response/mis-response), especially among the richer households. On the other hand, there is the so-called 'small sample bias': for obvious practical reasons, income surveys are carried out on a sample of the population. A limited sample size usually leads to imperfect representation at narrow income intervals at the top of the income distribution. As the distribution is not dense at the top, the probability of being drawn into the sample is low. This leads to biased inequality figures, especially when using top-sensitive inequality measures such as top income shares. To increase the representativeness at the top of the income distribution, we use administrative data from personal income tax records (IPCAL) to adjust the top of the earnings distribution by applying a correction method developed by Blanchet, Flores, and Morgan (2022), hereafter called 'BFM-method'. The adjustment consists of two steps: reweighting and replacement. First, the survey weights are adjusted on the basis of the earnings distribution of the tax data. Second, above a certain threshold, earnings are replaced in order to increase the precision at the top of the distribution. More details on both steps and the determination of the threshold, the so-called merging point, can be found in subsection C.1 in Appendix.

The second limitation of the income surveys available to us is the lack of good quality information on household wealth and capital income. In EU-SILC, for example, all financial capital income (e.g. interest and dividends) is recorded in one variable. On the other hand, national accounts provide a detailed breakdown of total financial capital income into different components. In Section 2 we have shown that capital income accounts for a substantial share of net national income and that its composition has changed drastically, with potentially large distributional consequences. We have enriched the information on different capital incomes in EU-SILC by including information on households' asset portfolios from the Household Finance and Consumption Survey (HFCS), based on a predictive mean matching (PMM) imputation method. In addition, as wealth surveys - for similar reasons as income surveys - often do not cover the very wealthy, we added a top correction for the 'missing wealthy'. This adjustment, based on Disslbacher et al. (2023), is explained in detail in subsection C.2 in Appendix. Finally, we calculated asset-specific rates of return by dividing the corresponding financial income in the annual sector accounts for the household sector in year t by the corresponding level of assets in the financial accounts for the household sector in year $t-1$. For each household in the income survey, we applied these rates of return to the imputed asset portfolio to determine its capital income.

We have enriched EU-SILC with HFCS information only for those years for which an HFCS survey is available: 2009, 2013, 2016 and 2019. We therefore present results for two different DINA series: one in which we rely exclusively on the distributional information of the uncorrected income surveys from 1985 to 2022, and another - shorter - one in which we have applied the corrections described above.[17] The first series is discussed in Section 4, the second in Section 5.

3.2 Methodology to insert distributional information into the National Accounts

To construct DINA, we link each income component of net national income to a corresponding concept in the micro-data set. We then use the distribution of that concept as observed in the survey to distribute the macro counterpart. We also define a unit of observation and a reference population on which to base the distributional analysis.

Linking macro and micro data

For each component of the NNI, we have identified a (combination of) variable(s) in the micro dataset that conceptually corresponds to the macro aggregate. We document these links in detail in Section D in Appendix. [18] To discuss some of the choices made, we provide a simplified snapshot of the links in Table 8 for the main components of the first three DINA income concepts. We show the nominal value (in billion €) in the national accounts and compare it with the closest proxy in the microdata for the income year 2019. In the right-hand column, we express the correspondence in terms of a coverage rate, which we calculate as the survey total divided by the macro total.

For some income components, such as wages and salaries, the link between the macro concept and its counterpart(s) in the micro data is obvious. A coverage close to 1 is evidence of this rather straightforward correspondence. Other variables that are easily linked and for which we observe a coverage close to 1 are some benefits such as sickness and disability benefits and pensions. In this case, we rescale the income variable in the microdata in order to recalibrate its sum to the aggregate observed in the national accounts:

$$y_{c,i}^{DN} = \frac{Y_c^{NA}}{Y_c^{mi}} \cdot y_{c,i}^{mi} \tag{1}$$

where $y_{c,i}^{DN}$ is the rescaled variable of the income component c for observation i, $y_{c,i}^{mi}$ is the original value observed in the microdata, and the variables with the capital letter Y denote aggregates for the income component c (Y_c^{mi} in the microdata and Y_c^{NA} in the national accounts). [19] This proportional adjustment procedure amounts to producing a version of the survey in which the distribution of the income component c is aggregated to the level of the National Accounts. The factor by which we multiply the survey observations is the inverse of the 'coverage rate' for income component c in Table 8.

[17] The correction of the income distribution based on administrative tax data can be done for more years than the years for which we have an HFCS survey. However, for ease of presentation, in Section 5 we present results for years in which we have applied all corrections together.

[18] This linkage is mainly based on the definition of national accounts concepts and their implementation in Belgium as documented in Eurostat (2013) and National Bank Belgium (2017) and also inspired by previous comparisons between national accounts and microdata as for example in Eurostat (2018) or **jestl_linequality_2023<empty citation>**.

[19] The aggregate in the survey is calculated using the weight available to make it representative of the population.

Table 8: link between income components in national accounts and surveys (2019)[a]

DINA-element	NA code	NA bn€	Survey (SILC) code		Survey (SILC) bn€	coverage (SILC/NA)
Components of pre-tax factor income (DINA income concept 1)						
Wages & salaries[b]	$D.11_{R,S14}$	181.4	PY010G	cash	179.4	1.013
			PY020G	non-cash	4.3	
Mixed income[c]	$B.3_{n,R,S14}$	30.0	PY050G		18.2	0.607
Real & imputed rents[c]	$B.2_{n,R,S14}$	9.8	HY030G	imputed	11.8	1.623
			HY040G	actual	4.1	
Interest income hh's	$D.41_{R,S14}$	1.3	HY090G		2.7	0.148
Dividend income hh's	$D.42_{R,S14}$	17.0				
Undistributed profits	$B.5_{n,U,S11+12}$	43.4	n.a.	n.a.	n.a.	n.a.
Value added tax	$D.2_{R,S13}$	32.2	n.a.	n.a.	n.a.	n.a.
Excise duties	$D.2_{R,S13}$	11.1	n.a.	n.a.	n.a.	n.a.
Components of pre-tax post-replacement income (DINA income concept 2)						
SSC's households	$D.613_{U,S14}$	27.7	simulated		29.4	1.060
SSC's employer	$D.611, 612_{U,S14}$	63.5	simulated		45.4	0.715
Pension benef. (1st pillar)	$D.621_{R,S14}$	59.6	PY100G	old-age b.	59.9	1.005
Pension benef. (2nd pillar)	$D.622_{R,S14}$		PY110G	survivor b.		
Unemploym. Benefits	$D.621_{R,S14}$	6.4	PY090G		9.1	1.427
Sickn. & inval. benef.	$D.621_{R,S14}$	9.4	PY120G	sickness	2.2	1.146
			PY130G	invalidity	8.6	
Components of post-tax disposable income (DINA income concept 3)						
Income tax hh's	$D.51_{R,S13}$	48.8	simulated		55.8	1.142
Tax movable property hh's	$D.51_{R,S13}$	3.7	simulated		0.8	0.199
Taxes on fin. & capital transact.	$D.2_{R,S13}$	5.4	n.a.		n.a.	n.a.
Taxes on land & building	$D.2_{R,S13}$	5.8	n.a.		n.a.	n.a.
Corporate inc. tax	$D.51_{R,S13}$	17.7	n.a.		n.a.	n.a.
Child allowances	$D.623_{R,S14}$	6.9	HY050G		6.3	0.917
Social assistance benefits	$D.623_{R,S14}$	2.0	HY060G		1.4	0.709

[a] Own calculations based on download of the National Accounts from Nbb.Stat on 19.02.2024. All variables are in nominal levels. For some items the NA-code refers to the code of an aggregate, but the value in the table is obtained by selecting components from the more detailed series: '*Breakdown of paid social benefits*' and '*Received taxes and actual social contributions by kind*' in the National Accounts. The detailed codes can be found in Tables A.9 to A.12 in Section D. In the column 'code' for the survey we give the variable name when directly available in the dataset, 'simulated' when the variable is obtained by means of the microsimulation model EUROMOD, and 'n.a.' when there is no counterpart in the EU-SILC for this item in the National Accounts.

[b] Remuneration of employees including social security contributions paid by employee, but excluding the contributions paid by the employer.

[c] Both 'mixed income' and 'imputed rents' are in *net* terms for the NA-values.

Table 8 shows that despite the fact that at the conceptual level we have good counterparts in EU-SILC for 'mixed income' and for 'rental income from property', the divergence between the aggregates in the micro and in the macro data is considerable. In the case of mixed income, EU-SILC significantly underestimates income in the national accounts, while the opposite is true for actual and imputed rents.

Obviously, at the level of the first DINA concept, the distribution of income from financial assets is the most challenging. Not only are dividend income and income from fixed-interest assets (such as bonds or savings accounts) not observed separately, but also the aggregate coverage is poor. In 2019, the total financial income of €2.7 bn in EU-SILC accounts for only 14.8% of the aggregate financial income in the national accounts. Undistributed profits are an example of a component of NNI that has no direct counterpart in the household micro-data. In this case we look for a proxy variable in the microdata that is conceptually as close as possible to the NA aggregate. The manual of WIL recommends to distribute retained earnings *"in proportion to stock ownership, be they held directly or indirectly, in privately or publicly traded companies"* (Alvaredo et al., 2024, p.57). For the four years in which we used the HFCS to give each household a portfolio structure of its financial assets (our second DINA series), we follow this guideline. However, for the 1985-2022 series, where we rely only on the original information from the income surveys, we are forced to use the single variable 'financial capital income' as the closest proxy for allocating undistributed profits across the distribution.

In some cases we do not even observe a proxy (income) variable in the microdata, in which case we have no choice but to distribute the NA aggregate purely 'by assumption'. In most of these cases we choose to allocate the aggregate in such a way that the distributional assessment using scale-invariant inequality measures remains unchanged. A specific case where we apply this proportional allocation is the income component 'other income' of the first DINA income concept (the bottom row in Table 3). This 'other income' consists mainly of the net balance of primary income of the general government sector ($B.5_{n,S13}$), which in turn includes net indirect taxes and net property income of general government (i.e. interest received minus interest paid). As factor income is recorded net of indirect taxes, we inflate all factor incomes uniformly *"to line up with the national income aggregate. That way, they reflect the purchasing power of pretax income at the post-tax prices that exist in the economy. Because this is pre-tax (before any consumption decision is made), it makes the most sense to do a uniform rescaling"* (Alvaredo et al., 2024, p.59). Of course, this proportional distribution of 'other income' at the level of the first DINA concept does not prevent us from choosing, for specific applications, to distribute indirect taxes differently, e.g. based on household specific spending patterns. The second way of distributing income 'by assumption' is to allocate the NA aggregate on a per capita basis. We do this for some components of the fourth DINA income concept, government expenditure ($P.31_{U,S13}$ and $P.32_{U,S13}$, not shown in Table 8, but see Section D in Appendix for more details).[20]

[20] A detailed overview of how each income component from the national accounts is distributed using the micro-data

For the second and third DINA income concepts, the aggregates of many income components in the microdata correspond quite well to the national accounts counterpart. This is not surprising for employee social contributions or child benefits. But also pension benefits are quite well covered. Unemployment benefits, but also taxes on movable property, differ quite strongly between EU-SILC and the national accounts.

Defining the reference population and the division of household income

Part of the confusion in inequality analysis is caused by different underlying choices about how to define the population or how to share household income among family members. For this reason, WIL strives to make the choices made explicit in comparative analyses so that they can be applied consistently over time or across countries (see Alvaredo et al., 2024).

The first choice concerns the definition of a reference population. This choice is, of course, partly determined by the research question. For a distributional analysis at the level of pre-tax income, it may be useful to restrict the analysis to the sub-population of those who are (or could be) still economically active. This is often done by using an - inevitably arbitrary - age criterion, e.g. all adults aged between 20 and 65. On the other hand, it makes sense to include children and pensioners in an analysis of after-tax income. The empirical results in this paper are mainly based on the choice of a middle way: we allocate income to all ‘adults’, defined here as everyone aged 20 or over.

The second choice concerns the way in which income is distributed within the household. There are several options here, as some income components are individual in nature (e.g. earnings), while others are more collective in nature, such as income from jointly owned assets. The WIL distinguishes between the following possible scenarios:

1. The ***individual-split*** scenario (also referred to as *no split*) assigns individually earned income to the household member who earns it; income components observed at household level are divided equally between the household members. The latter, of course, overlaps with the choice of reference population in the previous step: if only the adult population is chosen for the analysis, these income components will only be split between the household members belonging to the chosen reference population.

2. The ***broad equal-split*** scenario first aggregates all income components from all household members and then distributes them equally among all household members. Again, this depends on the choice of reference population: if only the adult population is chosen for the analysis, then the total household income (of the adults) is distributed only to these members. If the whole population (including children) is chosen, then a per capita distribution among all individuals is obtained.

3. The ***narrow equal-split*** scenario is a variant of the previous one. It only divides the total income of the *spouses* equally within the couple (instead of between all household members).

is included Tables A.9 to A.12 in Section D in Appendix.

For example, in the broad equal-split scenario, the household income of a couple living with their adult child (20+) is divided equally between the three household members. In the narrow equal-split scenario, the incomes of both parents are shared equally within the couple and the income of the child, if any, is attributed in full to the child.

Note that the WIL practice differs somewhat from what is common in international statistics on income inequality in that it does not rely on *equivalised* household disposable income. The latter would be obtained by dividing household income by an equivalence scale that tries to capture economies of scale and different needs of household members by age.[21] Alvaredo et al. (2024) argue that the DINA methodology does not use equivalence scales for two reasons. The first argument is pragmatic. Dividing household income by an equivalence scale means that all incomes of individuals no longer add up to aggregate income. The second reason is conceptual, as it states that the primary concern of DINA is to measure the distribution of income and not individual 'welfare', which is approximated by the use of equivalised household income.

As a benchmark concept, WIL therefore advocates the use of the '*broad equal split*' in combination with a reference population of 20+. In this paper we follow WIL's suggestion as far as possible by using a 'per adult' distribution of non-equivalised incomes as the default scenario. We explicitly mention when a different scenario is used.

4 Income inequality in Belgium (1985-2022)

In this section we summarise the evolution of income inequality in Belgium from 1985 to 2022. Based on the combination of national accounts information and the uncorrected income surveys, we use two measures of inequality: the Gini coefficient and, in line with published figures for other countries, top income shares.[22]

We will first discuss the evolution of inequality in *post-tax disposable income*, hereafter referred to as DINA3, then the changes in inequality in *pre-tax factor income*, or DINA1, and the extent of redistributive activity as measured by the difference in inequality between those two income concepts.

[21] The OECD, Eurostat and Statbel all publish Gini coefficients of equivalised household disposable income based on the total population with incomes equivalised using the modified OECD equivalence scale. The latter assigns a weight of 1 to the first adult (14+) in the household, a weight of 0.5 to each additional adult (14+) and a weight of 0.3 to each child (under 14).

[22] It is somewhat surprising that top income shares have regained such a prominent place in the analysis of inequality in recent years, since income shares as a measure of inequality do not satisfy one of the basic axioms of inequality measurement: the transfer principle. The increasing use of income shares is probably due to their simplicity and ease of use, especially in the case of historical analyses, since they do not require detailed information on the distribution of income across the population.

4.1 Evolution of inequality in post-tax disposable income (DINA3)

As in most other countries, evidence on income inequality in Belgium is mainly based on the distribution of equivalised household disposable income as measured in income surveys. The DINA income concept that comes closest to this survey income concept is DINA3 or post-tax disposable income. Figure 3 compares the evolution of inequality of equivalised household disposable income as reported in the surveys (red line and dots) with our distributional national accounts series (blue line and dots). Both series are based on the same reference population (total population) and divide household income by an equivalence scale.[23] Differences in the level and evolution of the Gini can therefore only be due to differences between the disposable income in the surveys and the macro-aligned concept of post-tax disposable income.

For the period 1985-2022 we use distributional information from three income surveys: SEP (1985-1997), ECHP (1995-2000) and EU-SILC (2003-2022). Due to methodological differences and breaks between these three surveys, as discussed in detail in Assal et al. (2023), we are cautious about comparing levels of inequality across surveys. In Figure 3 we therefore only link Gini coefficients from the same income survey and restrict statements on the evolution of inequality to trends within the sub-periods covered by the same income survey source. The SEP survey shows a stable but slowly increasing Gini from 22.1 in 1985 to 23.3 in 1997. The period covered by the ECHP varies between a Gini of 27.2 and 29.5. According to EU-SILC, the distribution of equivalised disposable household income in Belgium has become more equal over the last twenty years. Income inequality as measured by the Gini decreased from 26.7 in 2003 to 24.5 in 2022.

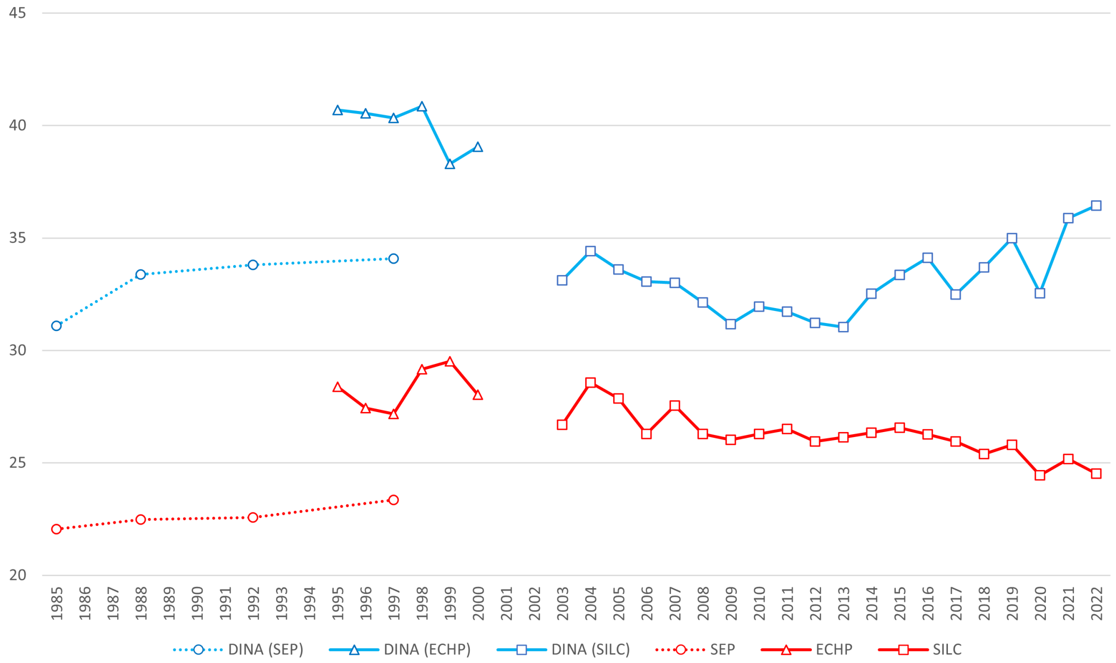
The distributional national accounts challenge these findings. Both the level and the trend of inequality are very different. In line with similar findings for many other countries, the level of income inequality is higher in each of the three sub-periods. The *evolution* of inequality follows more or less the same trend up to 2012 but is very different from that point onwards. From 2013 onwards, the alignment of distributional microdata with macro aggregates from the national accounts reveals an increase in income inequality, which differs markedly from the decreasing trend based on income surveys.

Figure 4 compares these results with the DINA series for Belgium available in the World Inequality Database. As the Gini coefficient is not available on the WID, we use the top income share (for the top 10%) as an inequality index. However, the evolution of this top share is very similar to that of the Gini coefficient. The level of the WID series is close to our series covered by distributional information from SEP and EU-SILC. This should not come as a surprise: the WID series is the one from Blanchet, Chancel, and Gethin (2022) and relies on SEP and EU-SILC (but not ECHP) for distributional information.[24] They use interpolation techniques to obtain inequality figures for the

[23] With this scenario we deviate from the WIL proposal: a per adult distribution of non-equivalised but equally split household income. The next figures in this section follow the WIL proposal.

[24] Blanchet, Chancel, and Gethin (2022) actually use LIS data from 1985 to 1997 which are based on SEP data. And although EU-SILC is available for all years from 2003 onwards, they state that they do not use EU-SILC in 2004,

Figure 3: Inequality in equivalised (post-tax) disposable income, 1985-2022 (Gini)



Note: All series use the total population as the reference population. Incomes are equivalised using the modified OECD-scale. The income concept used is disposable income in income surveys and post-tax disposable income (DINA income concept 3) for the DINA series. The red dots indicate series based on income surveys, the blue dots indicate DINA series.

years not covered by SEP or EU-SILC.[25]

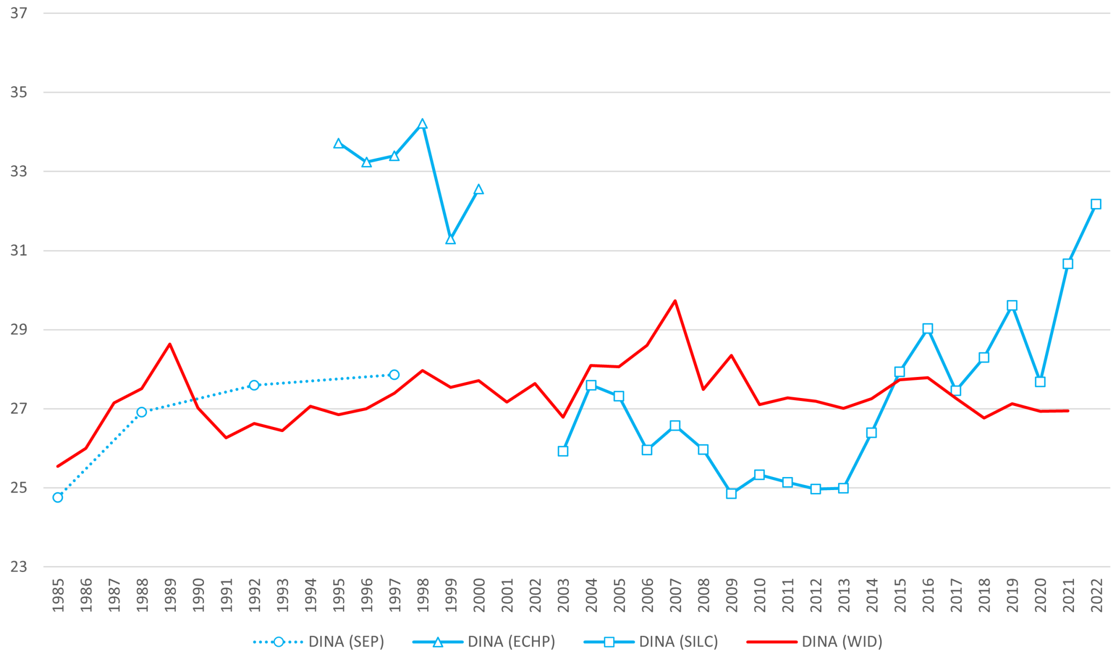
Although we use the same underlying distributional information, the recent marked increase in income inequality does not show up in the WID series. Below we will examine this remarkable increase in our series in more detail.[26]

2005 and 2006 (but do not explain why). Their series for Belgium are available up to 2017 and have been updated to more recent years by Neef and Sodano (2022).

[25]Starting from survey tabulations from different sources, they 'recover' full distributions using a generalised Pareto interpolation method developed by Blanchet, Fournier, and Piketty (2022). In addition, they use a machine learning algorithm (XGBoost) to impute the distribution of missing 'target' concepts based on observed related 'source' concepts. The algorithm is trained (across country-years) on cases where the income distribution is simultaneously observed for both the 'target' and the 'source' concept. This allows, for example, the imputation of taxable income and/or tax concepts that are not included in SEP or EU-SILC.

[26]From this point in the paper, we will no longer use the DINA series based on distributional information from ECHP. The level of inequality figures clearly deviates from the series based on SEP and EU-SILC.

Figure 4: Inequality in post-tax disposable income (DINA-concept 3), 1985-2022 (Top 10% income share)



Note: All series use the adult population (20+) as the reference population. Incomes are equally split among all adult household members (broad equal-split). The income concept used is disposable income in income surveys and post-tax disposable income (DINA income concept 3) for the DINA series. The blue lines are based on our own calculations, the red line is the DINA series available at the WID and downloaded on October 4, 2024.

4.2 Pre-tax distribution and redistribution

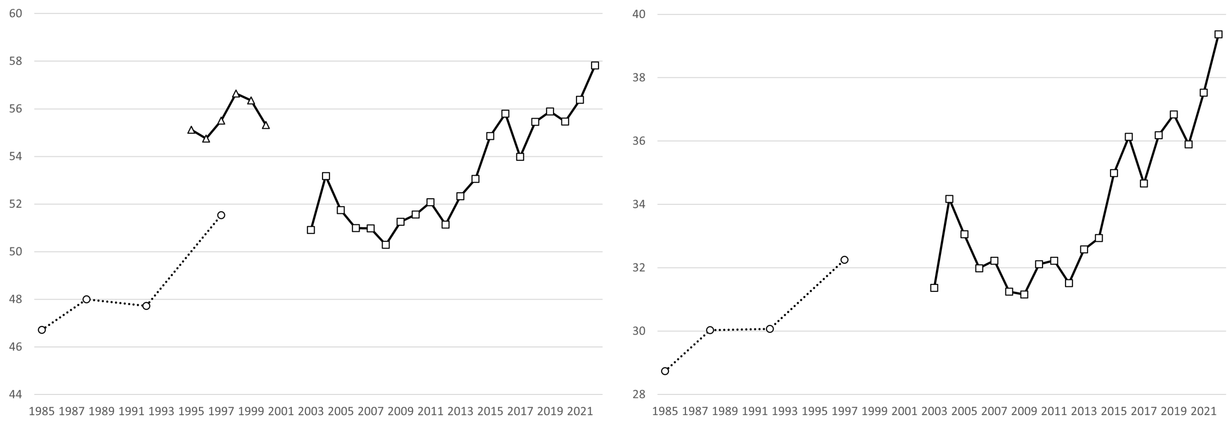
The distribution of post-tax disposable income (DINA3) results from the distribution of pre-tax factor income (DINA1) and the redistribution through social contributions, taxes and benefits. We first look at the evolution of the pre-tax factor income distribution. An increase in pre-tax income inequality may be reflected in an increase in post-tax income inequality. Secondly, we look at the evolution of redistribution from pre-tax to post-tax income.

The evolution of inequality in pre-tax factor income

In Figure 5 we show the evolution of inequality in the distribution of pre-tax factor income (DINA1), measured by the Gini coefficient (left panel) and the share of the top 10% (right panel). We observe a strong increase in inequality. In the first sub-period, the increase in inequality of pre-tax factor income is more pronounced than the increase in post-tax income inequality. And we also observe a strong increase in pre-tax income inequality since 2008: the Gini coefficient and the income share of the top 10 per cent both increase by 8 percentage points between 2008 and 2022.

To explain the evolution of pre-tax factor income inequality, we rely on a decomposition by income

Figure 5: Inequality in pre-tax factor income (dina-concept 1), 1985-2022



(a) Gini (b) Top 10% income share

Note: All series use the adult population (20+) as the reference population. Incomes are equally split among all adult household members (broad equal-split). The income concept used is pre-tax factor income (DINA income concept 1).

source. We follow what Shorrocks (1984) called the 'natural decomposition of the Gini', described in detail by Lerman and Yitzhaki (1985). The Gini G_t for period t can be decomposed into the contributions of the income sources K as follows:

$$G_t = \sum_{k=1}^{K} s_{t,k} \cdot \tilde{G}_{t,k} \tag{2}$$

where $\tilde{G}_{t,k}$ stands for the pseudo-Gini of income source k in period t, and $s_{t,k}$ is the share of income of source k in total income across the population:

$$s_{t,k} = \frac{Y_{t,k}}{Y_t} \tag{3}$$

The pseudo-Gini $\tilde{G}_{t,k}$ differs from the ordinary Gini of income y_k. The ordinary Gini ranks incomes on themselves, whereas the pseudo-Gini of income y_k ranks incomes from source k on total income y.[27] The difference in the rank order used in the pseudo-Gini compared to the ordinary Gini is reflected in the relationship between the pseudo-Gini and the Gini:

$$\tilde{G}_{t,k} = R_{t,k} \cdot G_{t,k}, \tag{4}$$

with

$$R_{t,k} = \frac{\text{cov}\left[y_{t,k}, F_t(y_t)\right]}{\text{cov}\left[y_{t,k}, F_{t,k}(y_{t,k})\right]} \tag{5}$$

the 'Gini correlation' and $F(x)$ denoting the cumulative distribution function of variable x.[28]

[27] In fact, the pseudo-Gini is based on the concentration curve of income y_k, ranked on total income y.

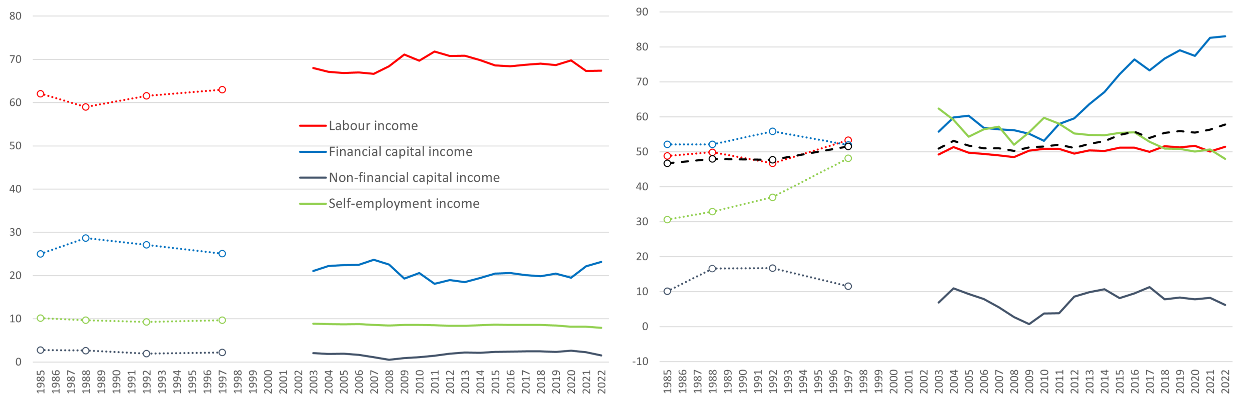
[28] The role of $R_{t,k}$ in equation (5) follows from the fact that the standard Gini can be written as a covariance

Combing equations (2) and (4) into

$$G_t = \sum_{k=1}^{K} s_{t,k} \cdot R_{t,k} \cdot G_{t,k} \tag{6}$$

shows that the evolution of total pre-tax factor income inequality, as measured by the Gini, is explained by three factors: (i) changes in the shares of the underlying income components, (ii) changes in the Ginis of these income components and (iii) changes in the Gini-correlation. We show the evolution of the first factor (shares) and the pseudo-Ginis (which capture the other two factors) in Figure 6.

Figure 6: shares and pseudo-Ginis for different income sources of pre-tax factor income, 1985-2022



(a) shares of income sources   (b) pseudo-Ginis

Note: All data series use the adult population (20+) as the reference population. Incomes are equally split among all adult household members (broad equal-split). The income concept used is pre-tax factor income (DINA income concept 1), with four main components: labour income, financial capital income, non-financial capital income (i.e. real and imputed rents) and self-employment income.

The left-hand panel shows a more or less stable trend in income shares. We see a small decline in the income share of financial capital around the financial crisis, which then recovers. The right panel shows that the main explanation lies in the distribution of income components. In particular, the distribution of financial capital income stands out. The pseudo-Gini has risen sharply since the start of the financial crisis. This increase is driven by both an increase in the Gini and the Gini correlation of financial capital income.[29]

Figure 6 also reveals some other interesting results. First, both the level and the trend of inequality in pre-tax factor income are mainly driven by labour income. With an income share of around

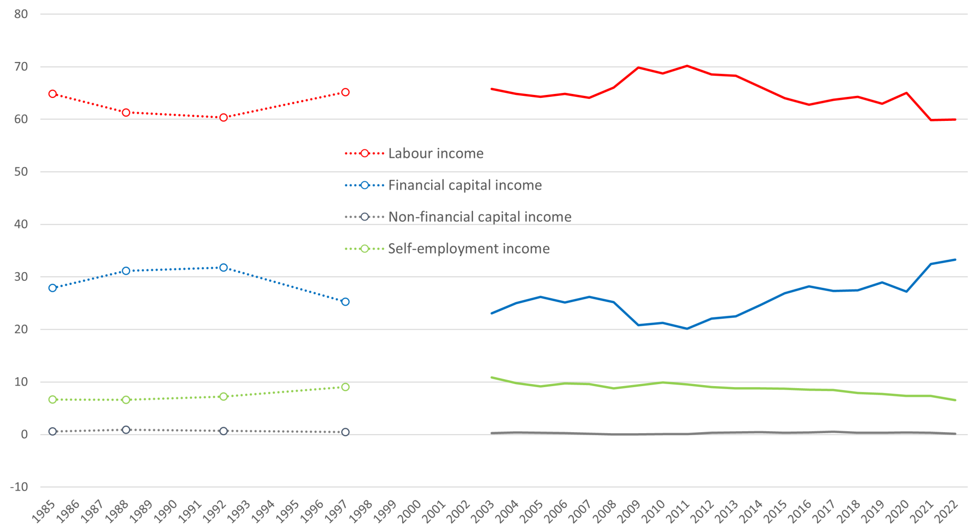
between income from source k and its rank: $G_k = (2/\mu_k) \cdot \mathrm{cov}\,[y_k, F_k(y_k)]$, with μ_k average income of source k. The pseudo-Gini is also a covariance between income from source k and a rank: $\tilde{G}_k = (2/\mu_k) \cdot \mathrm{cov}\,[y_k, F(y)]$.

[29] The Gini correlation R_t of financial capital income rises from 76.2 in 2009 to 91.9 in 2022, while the Gini of this component rises from 72.3 in 2009 to 90.4 in 2022. See Table A.15 in Section E in Appendix for all coefficients.

70%, labour income ensures that the sharp increase in capital income inequality is translated into a much smaller increase in total factor income inequality before tax. The role of income from self-employment and property is rather limited, given their small share in total income.

The decomposition into shares and pseudo-Ginis can be translated into contributions of income sources to the overall level of inequality. By dividing the RHS of equation (2) by its LHS - that is, by the Gini of pre-tax factor income in period t - we obtain the *average* contributions of income sources to the level of inequality. Figure 7 clearly shows the increasing contribution of financial capital income to the overall level of inequality as measured by the Gini. While in 2009 financial capital income contributed about 20% to the level of inequality, in ten years its importance has increased by 10 percentage points. We therefore conclude that financial capital income is the main explanation for the increase in pre-tax factor inequality since 2009. To a lesser extent, this increase is explained by a growing share of income from capital, but the distribution of financial capital income itself is more important.

Figure 7: Contribution of income sources to Gini of pre-tax factor income, 1985-2022

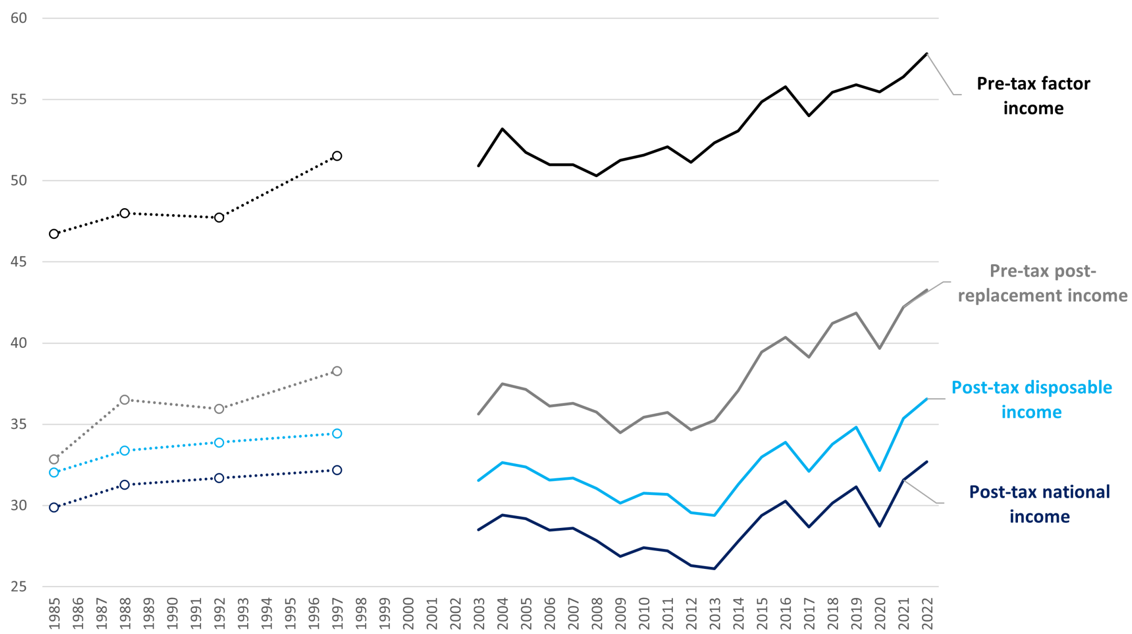


Note: All data series use the adult population (20+) as the reference population. Incomes are equally split among all adult household members (broad equal-split). Four income components of pre-tax factor income: labour income, financial capital income, non-financial capital income and self-employment income.

The transition from pre-tax factor income to post-tax income

A first explanation for the increase in post-tax income inequality is thus to be found in the development of pre-tax income inequality, which in turn is driven by the increase in capital income inequality after the financial crisis. Figure 8 indeed shows that the evolution of inequality is quite similar for all four DINA income concepts. However, it is interesting to note that the turning point at which inequality starts to rise differs between the four income concepts. While the turning point for pre-tax factor income is in 2008, inequality does not start to rise until 2013 for post-tax income. This implies that redistribution through the tax and benefit system has postponed the increase in post-tax income inequality.

Figure 8: Inequality (Gini) for different DINA income concepts, 1985-2022



Note: All data series use the adult population (20+) as the reference population. Incomes are equally split among all adult household members (broad equal-split).

Inspired by the index proposed by Reynolds and Smolensky, we define the metric of redistributive activity as the difference in the Gini between pre- and post-tax income:[30]

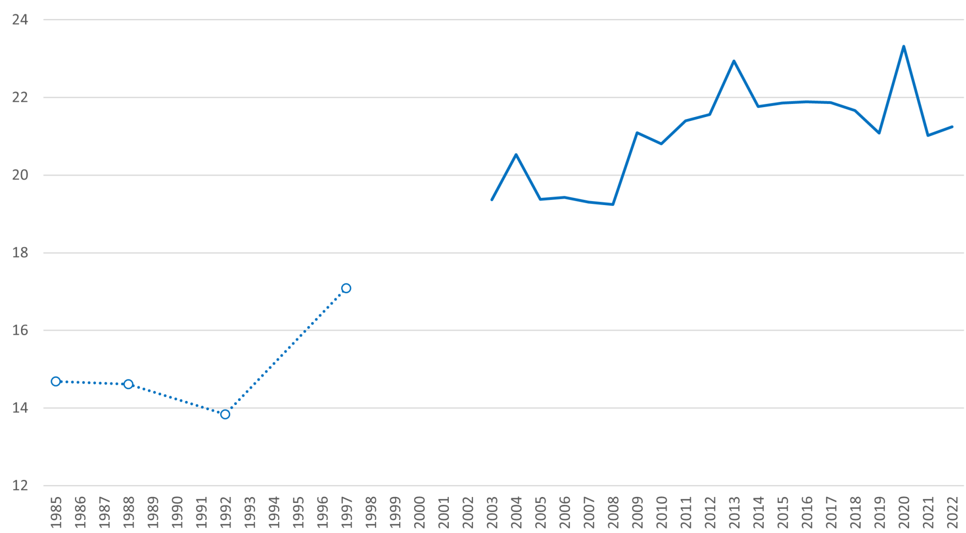
$$\Pi^{RE} = G_{Pre-tax} - G_{Post-tax} \tag{7}$$

The Gini of pre-tax factor income and of post-tax disposable income provides a measure of the redistributive impact of social contributions, income taxes and cash benefits. Figure 9 shows the

[30]The Reynolds-Smolensky index defines the metric of redistribution as the difference between the Gini of pre-tax income and the concentration index of post-tax disposable income, where the ranking is determined by pre-tax income. The greater the possible differences in rank between the pre-tax and post-tax rankings, the greater the difference between the Reynolds-Smolensky index and the definition of redistributive activity in Equation 7.

redistribution index for the period 1985-2022. The index falls between 1985 and 1992, which is not surprising. This subperiod was marked by a reform of the personal income tax system, e.g. a reduction of the tax rate for the highest income bracket from 72% in 1985 to 55% in 1992. The subsequent period (1992-1999) is characterised by fiscal consolidation to meet the Maastricht convergence criteria. Taxes were increased and benefits were reformed, with a negative impact on the disposable income of households, but with an increasing degree of redistribution.

Figure 9: Evolution of redistribution, 1985-2022



Note: The data series use the adult population (20+) as the reference population. Incomes are equally split among all adult household members (broad equal-split).

Tax cuts implemented in the subsequent policy period have benefited low-income households the most. An earned income tax credit was introduced in 2001 to reduce the tax burden on working people with low incomes. This further increased the degree of redistribution. From 2003 to 2008 we observe a rather stable redistribution index. With pre-tax income inequality decreasing until the financial crisis, one could afford not to increase the degree of redistribution of the tax and transfer system and still end up with a decrease in disposable income inequality. But once pre-tax income inequality started to rise after the financial crisis, one had to make the tax and transfer system work harder to achieve a fall in after-tax disposable income. Figure 9 shows that from 2013 onwards, rising pre-tax income inequality was no longer offset by rising redistribution. This then inevitably led to the increase in inequality in post-tax income as observed in Figure 8. In other words, the recent increase in the level of inequality in post-tax disposable income is not due to a decreasing level of government redistribution, but to an increase in pre-tax income inequality which is no longer offset by an increase in the level of redistribution. We only observe an increase in redistribution in 2020 due to the introduction of a social safety net with billions spent on benefits for the temporarily unemployed during the covid-19 pandemic.

5 Digging deeper into rising income inequality since 2009 using tax data and wealth surveys

The important finding of capital income as a driver of rising income inequality is also the weakness of the DINA series constructed above. It relies heavily on the information on financial capital income contained in a single variable in EU-SILC (HY090G), which reports *interest, dividends and profit from capital investments in an unincorporated company.* The poor coverage in EU-SILC of income from financial capital recorded in the national accounts has of course been corrected by rescaling to the level of the national accounts. However, it is fair to say that the underlying distributional information is (too) coarse. In this section we test whether our finding of rising inequality is robust to the use of richer information on the distribution and composition of capital income, as described in subsection 3.1.[31] We construct a second DINA series in which we have introduced two refinements to EU-SILC that allow for a more granular distribution of net national income, especially its capital income components.

The first refinement uses administrative tax data to improve the representation at the top of the earnings distribution. The second links the detailed information on household balance sheets in the Household Finance and Consumption Survey (HFCS) to the EU-SILC data. Compared to the DINA series presented in Section 4, this enrichment allows us to disaggregate financial capital income into three separate components: interest, dividends and other investment income. We also now have information on share ownership, which we use to allocate undistributed profits to households. Finally, HFCS also provides more detail on the distribution of real and imputed rental income. As HFCS waves are only available for income years 2009, 2013, 2016 and 2019, we have constructed this second DINA series for these four years. However, as this is precisely the period of interest in which factor income inequality before tax started to increase, this shorter period does not seem to devalue the additional insights too much.

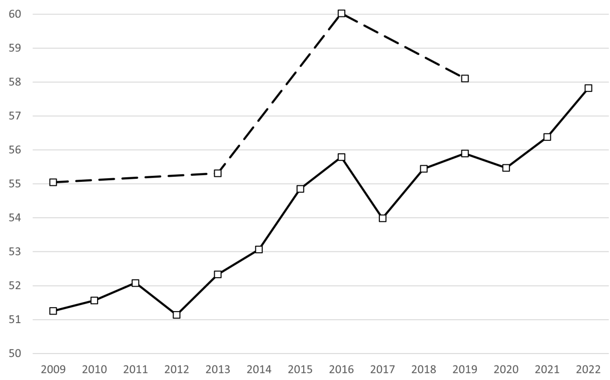
In subsection 5.1 we present the impact of this richer information on the conclusions about inequality trends in Belgium by comparing the inequality figures between our two DINA series. The more detailed information on the distribution of capital income components allows a closer look at the distribution of capital income in subsection 5.2.

5.1 Rising income inequality confirmed

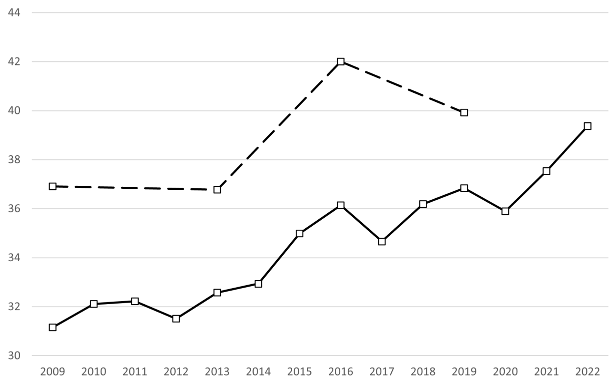
Figure 10 compares the refined DINA series with those presented in Section 4. The figures based on the extended DINA series for 2009, 2013, 2016 and 2019 are connected by dotted lines. The full lines are those shown above in Section 4 (Figure 5 for pre-tax factor income and Figure 3 and Figure 4 for post-tax disposable income). The upper part of the figure shows the evolution of inequality of pre-tax factor income (DINA1) and the lower part shows the evolution of inequality of post-tax disposable income inequality (DINA3).

[31] More details can be found in Appendix Section C.

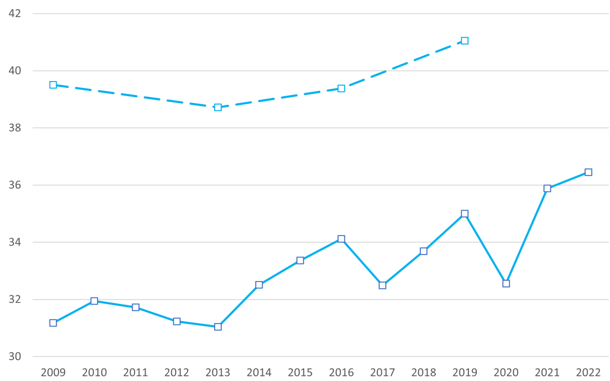
Figure 10: Inequality in pre-tax factor and post-tax disposable income, 2009-2022



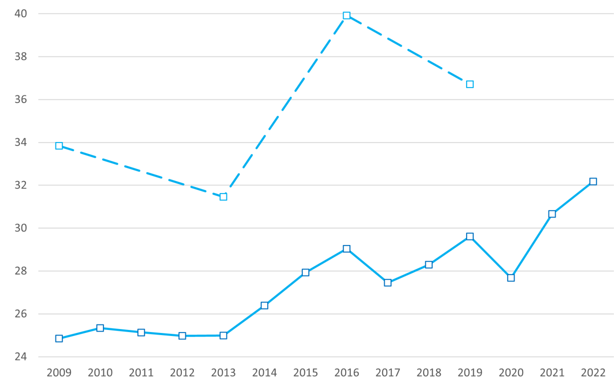
(a) Gini - Pre-tax factor income



(b) Top 10% - Pre-tax factor income



(c) Gini - Post-tax disposable income



(d) Top 10% - Post-tax disposable income

Note: The data series use as reference population the adult population (20+) (Panel A, B and D), and the total population (Panel C). Incomes are equally split among all adult household members (Panel A, B and D) and equivalised (Panel C) . The full lines correspond with the figures shown in Figure 5a (panel A), Figure 5b (panel B), Figure 3 (panel C) and Figure 4 (panel D).

The extended DINA series in Figure 10 confirms the rising inequality trend, but at a much higher level of inequality. The higher level is not surprising as we have introduced two 'top corrections': an adjustment for the distribution of earnings and a top correction for the missing rich. However, detailed analysis shows that the top earnings adjustment has only a limited impact on the level and no impact on the trend of earnings inequality (labour and mixed income).[32]

From the decomposition of Lerman and Yitzhaki (1985), expressed in Equation 2 above, we know that income components affect the overall level of inequality through two channels: their distribution, as expressed in the pseudo-Gini (or concentration coefficient), and their share in total income. By construction, the latter, i.e. the *share* of labour and capital income in total net national income, does not differ in our second DINA series from those used in Section 4. Thus, if the corrections leave

[32]See Table A.5 in Appendix subsection C.1 for the impact on the Gini coefficient and top income shares of the income distribution in 2019. The increase in the Gini is limited to about 1 Gini point.

the evolution of income inequality more or less unaffected, it must be the distribution of capital income itself that explains the rising trend. Indeed, the Gini of capital income - both financial and non-financial - rose from 77.5 in 2009 to 85.3 in 2019.

5.2 The uneven inequality of the different components of capital income

Based on the information from HFCS, we have substantially improved the distributional information of capital income. We can now distinguish different components of capital income. This allows us to apply the decomposition of Lerman and Yitzhaki (1985) to the Gini of capital income itself. The decomposition into contributions of five sources of capital income is shown in Table 9.

TABLE 9: DECOMPOSITION OF CAPITAL INCOME INEQUALITY

	2009	2013	2016	2019
Shares as % of total capital income				
Non-financial capital income	17.3	16.8	14.2	13.5
Interests received	17.9	6.7	2.6	1.6
Dividends	19.5	21.3	19.7	20.5
Undistributed profits	28.8	38.3	50.3	52.3
Other financial capital income	16.5	16.9	13.2	12.2
Pseudo-Ginis				
Non-financial capital income	44.7	40.8	44.0	42.0
Interests received	69.5	56.1	58.1	51.8
Dividends	96.8	97.3	97.3	96.9
Undistributed profits	96.8	97.3	97.3	96.9
Other financial capital income	63.7	64.3	65.0	67.5
Average contribution to Gini of capital income (in %)				
Non-financial capital income	10.0	8.6	7.4	6.6
Interests received	16.0	4.8	1.8	0.9
Dividends	24.4	26.0	22.7	23.3
Undistributed profits	36.0	46.8	58.0	59.5
Other financial capital income	13.6	13.7	10.2	9.6
Gini of capital income	77.5	79.4	84.5	85.3

Note: The bottom line displays the Gini of capital income. The upper two panels show the share $s_{t,k}$ and the Pseudo-Gini $\tilde{G}_{t,k}$ from Equation 2, with the ordering of individuals based on total capital income. The bottom panel gives the product $s_{t,k}.\tilde{G}_{t,k}$, but expressed as a percentage of the Gini in the bottom line.

The top panel of Table 9 shows how drastically the composition of capital income has changed over the last ten years, a fact already noted in Figure 2 above. The share of non-financial capital income, including the net operating surplus of the household sector (i.e. real and imputed rents) minus interests paid, falls from 17.3% in 2009 to 13.5% in 2019. But it is within the share of income from financial assets that we find the biggest swings: a collapse in fixed interest income (e.g. from savings deposits) and a large increase in the share of undistributed profits. The share of interests received melted away from 17.9% in 2009 to 1.6% ten years later. The share of distributed profits remained stable at around 20%, while the share of undistributed profits rose sharply from less than 30% in 2009 to more than 50% in 2019. The remaining share of other financial income (mainly capital income) decreases slightly.

This changing composition interacts with the pseudo-Ginis of the second panel of Table 9. It is clear that the components that have lost ground are those with the lowest pseudo-Gini. Firstly, non-financial capital income with a pseudo-Gini of 44.7 in 2009, which fell to 42.0 in 2019. Second, interest received, with a pseudo-Gini of 69.5 in 2009, which falls sharply to 51.8 in 2019. Also the other financial incomes seem to have a lower pseudo-Gini compared to the high concentration indices for dividends and undistributed profits of around 97.[33]

This interaction between shares and pseudo-Ginis leads to a change over time in the contribution of the components to the overall level of capital income inequality. The bottom panel in Table 9 shows the resulting relative contributions of the different components to the Gini of capital income. The increase in the Gini from 77.5 in 2009 to 85.3 in 2019 is mainly explained by the near doubling of the contribution of undistributed profits: from 36.0% in 2009 to 59.5% in 2019. The role of income from property and interest received in total capital income inequality falls from 26.0% in 2009 to 7.5% in 2019. These significant changes in the contribution of the different components of capital income are mainly driven by the change in shares and only to a limited extent by changes in the distribution of the components.

6 Conclusion

In this paper, we join recent country-specific attempts by, among others, Jestl and List (2023) for Austria, Bach et al. (2023) for Germany, Guzzardi et al. (2023) for Italy and Bruil (2023) for the Netherlands to construct more granular distributional national accounts for Belgium. First, we presented a consistent DINA series from 1985 to 2022, based solely on distributional information from income surveys and simulated taxes and benefits from a microsimulation model. Similar to the country-specific DINA research mentioned above, our inequality figures deviate from those produced by Blanchet, Chancel, and Gethin (2022), which are publicly available in the World Inequality Database. This discrepancy is most pronounced in the period following the financial crisis. While the publicly available series show a stable trend in income inequality in Belgium, the

[33]This very high value of the concentration index reflects the fact that the top 10% of the capital income distribution receives, for example, 98% of the undistributed profits in 2019, while the bottom 90% receives the remaining 2%.

figures presented in this paper show an increase in income inequality after the financial crisis. As labour income inequality is rather stable over the whole period, we find that capital income plays a key role in the recent increase in inequality.

This finding is confirmed by a more detailed DINA series covering the period 2009-2019. For this second DINA series, the distributional information comes from a combination of income surveys (EU-SILC), wealth surveys (HFCS) and administrative tax records (IPCAL). The additional information on household balance sheets from the wealth surveys allowed for a more detailed distribution of national net income, particularly capital income. Personal income tax records were used to correct the income distribution of the income surveys for the 'missing rich'.

This extension with additional data sources beyond the standard income surveys provided deeper insights into how capital income drives this evolution of income inequality. Using a decomposition of the Gini coefficient by source of income, we find that while the share of capital income in total net national income remained fairly stable, its distribution became more unequal in the aftermath of the financial crisis. Decomposing capital income into its different components shows that a changing composition of the (stable) capital income share leads to a more unequal distribution. Fixed interest income, which is more widely distributed across the population, fell sharply, while the share of more unequally distributed components (such as dividends or retained earnings) remained stable or even increased over the period 2009-2019.

These findings are not only relevant for the Belgian case. Other countries experienced similar developments in the composition of capital income. Italy, for example, experienced a comparable increase in the share of undistributed profits (+150% between 2009 and 2019). Not surprisingly, Guzzardi et al. (2023) points to capital income as the main factor explaining the rise in income inequality in Italy. In France, on the other hand, net undistributed profits increased by only 50% over the same period and interest received by households decreased less. This fits with the finding of Bozio et al. (2024) that inequality did not increase in France after the financial crisis.

The poor coverage of financial capital income in income surveys, combined with a stable trend in labour income inequality, explains why this recent increase in income inequality in Belgium has remained under the radar. Most of the empirical evidence was based on income surveys and showed a rather stable trend. The DINA approach revealed this previously undetected increase in income inequality by taking into account all capital income.

References

Alvaredo, F., Assouad, L., & Piketty, T. (2019). Measuring lnequality in the Middle East 1990–2016: The World's Most Unequal Region? [Publisher: John Wiley & Sons, Ltd]. *Review of Income and Wealth*, *65*(4), 685–711. https://doi.org/10.1111/roiw.12385

Alvaredo, F., Atkinson, A., Blanchet, T., Bauluz, L., Chancel, L., Blanchet, T., Fisher-Post, M., Flores, I., Garbinti, B., Goupille-Lebret, J., Martínez-Toledano, C., Morgan, M., Piketty, T., Robilliard, A.-S., Saez, E., Yang, L., & Zucman, G. (2024). Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database. *World Inequality Lab*.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2017). Global Inequality Dynamics: New Findings from WID.world. *American Economic Review*, *107*(5), 404–09. https://doi.org/10.1257/aer.p20171095

Assal, E.-M., Capéau, B., Decancq, K., Decoster, A., Kuypers, S., Vanderkelen, J., Vanheukelom, T., & Verbist, G. (2023). Les inégalités de revenu en Belgique: Ce que nous savons et ce que nous ignorons. *Revue belge de sécurité sociale*, *64*(2), 135–175.

Atkinson, A. B., Rainwater, L., & Smeeding, T. M. (1995). Income Distribution in OECD Countries. *OECD Social Policy Studies*, (18).

Atkinson, A. B., & Piketty, T. (2007). *Top incomes over the twentieth century: A contrast between continental european and english-speaking countries*. Oxford University Press.

Atkinson, A. B., & Piketty, T. (2010). *Top incomes: A global perspective*. Oxford University Press.

Bach, S., Bartels, C., & Neef, T. (2023). Distributional National Accounts (DINA) for Germany, 1992-2016. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4487665

Bach, S., Bartels, C., & Neef, T. (2024). *The distribution of national income in Germany, 1992-2019* (IWH Discussion Papers No. 25/2024). Halle Institute for Economic Research (IWH). Halle (Saale). https://doi.org/10.18717/dp3wkp-zv31

Blanchet, T., Chancel, L., & Gethin, A. (2022). Why Is Europe More Equal than the United States? *American Economic Journal: Applied Economics*, *14*(4), 480–518. https://doi.org/10.1257/app.20200703

Blanchet, T., Flores, I., & Morgan, M. (2022). The weight of the rich: Improving surveys using tax data. *The Journal of Economic Inequality*, *20*(1), 119–150. https://doi.org/10.1007/s10888-021-09509-3

Blanchet, T., Fournier, J., & Piketty, T. (2022). Generalized Pareto Curves: Theory and Applications. *Review of Income and Wealth*, *68*(1), 263–288. https://doi.org/10.1111/roiw.12510

Bozio, A., Garbinti, B., Goupille-Lebret, J., Guillot, M., & Piketty, T. (2024). Predistribution versus Redistribution: Evidence from France and the United States. *American Economic Journal: Applied Economics*, *16*(2), 31–65. https://doi.org/10.1257/app.20220023

Bruil, A. (2023). Distributional national accounts for the Netherlands and a comparison with 12 other countries. *Review of Income and Wealth*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/roiw.12609

Bruil, A., Van Essen, C., Leenders, W., & Lejour, A. (2022). Inequality and Redistribution in the Netherlands. *CPB Discussion paper*.

Capéau, B., Decoster, A., De Rock, B., & Vanderkelen, J. (2024). Did Belgium withstand the storm of rising inequalities? Income inequality in Belgium, 1985–2020. *Fiscal Studies*, *45*(3), 285–308. https://doi.org/https://doi.org/10.1111/1475-5890.12383

Capéau, B., Decoster, A., Sheikh Hassan, N., & Vanderkelen, J. (2023). Distributional National Accounts for Belgium. *BE-PARADIS Working Paper wp.23.6*.

Capéau, B., Decoster, A., Vanderkelen, J., Van Houtven, S., & Wolfs, K. (2023). *A statistical matching procedure to enrich the Belgian SILC-dataset with asset information from HFCS* (Technical Report). BE-PARADIS.

Chakraborty, R., & Waltl, S. R. (2018). Missing the Wealthy in the HFCS: Micro Problems with Macro Implications. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3202094

Chancel, L., & Piketty, T. (2019). Indian Income Inequality, 1922-2015: From British Raj to Billionaire Raj? [Publisher: John Wiley & Sons, Ltd]. *Review of Income and Wealth*, *65*(S1), S33–S62. https://doi.org/10.1111/roiw.12439

Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2022). *World inequality report 2022*. Harvard University Press.

De Schrijver, A. (2020). Fiscal data in the Statistics on Income and Living Conditions (SILC) survey: A path for the future?

Disslbacher, F., Ertl, M., List, E., Mokre, P., & Schnetzer, M. (2023). *On Top of the Top: A Generalized Approach to the Estimation of Wealth Distributions* (tech. rep.). SSRN. https://doi.org/10.2139/ssrn.4499915

Eurostat. (2013). *European system of accounts – ESA 2010*. https://data.europa.eu/doi/10.2785/16644

Eurostat. (2018). *Methodological note Comparison of household income : European Union Statistics on Income and Living Conditions and National Accounts* (tech. rep.).

Fixler, D., Gindelsky, M., & Johnson, D. (2019). Improving the Measure of the Distribution of Personal Income. *AEA Papers and Proceedings*, *109*, 302–06. https://doi.org/10.1257/pandp.20191037

Flachaire, E., Lustig, N., & Vigorito, A. (2023). Underreporting of Top Incomes and Inequality: A Comparison of Correction Methods using Simulations and Linked Survey and Tax Data. *Review of Income and Wealth*, *69*(4), 1033–1059. https://doi.org/https://doi.org/10.1111/roiw.12618

Frick, J. R., Grabka, M. M., Smeeding, T. M., & Tsakloglou, P. (2010). Distributional effects of imputed rents in five European countries. *Journal of Housing Economics*, *19*(3), 167–179. https://doi.org/https://doi.org/10.1016/j.jhe.2010.06.002

Garbinti, B., Goupille-Lebret, J., & Piketty, T. (2018). Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA). *Journal of Public Economics*, *162*, 63–77. https://doi.org/10.1016/j.jpubeco.2018.01.012

Guzzardi, D., Palagi, E., Roventini, A., & Santoro, A. (2023). Reconstructing Income Inequality in Italy: New Evidence and Tax System Implications from Distributional National Accounts. *Journal of the European Economic Association*, 1–45. https://doi.org/10.1093/jeea/jvad073

Hulliger, B., & Wiegand, G. (2012). Imputation of Housing Rents for Owners Using Models With Heckman Correction. *Survey Research Methods*, *62*(2), 95–103. https://doi.org/https://doi.org/10.18148/srm/2012.v6i2.5028

Jestl, S., & List, E. (2023). Inequality, Redistribution, and the Financial Crisis: Evidence from Distributional National Accounts for Austria. *Review of Income and Wealth*, *69*(1), 195–227. https://doi.org/https://doi.org/10.1111/roiw.12589

Kuznets, S., & Jenks, E. (1953). *Shares of upper income groups in income and savings* (tech. rep.). National Bureau of Economic Research.

Lerman, R. I., & Yitzhaki, S. (1985). Income Inequality Effects by Income Source: A New Approach and Applications to the United States [Publisher: The MIT Press]. *The Review of Economics and Statistics*, *67*(1), 151–156. https://doi.org/10.2307/1928447

National Bank Belgium. (2017). *Gross National Income Methodological inventory Belgium* (tech. rep.). National Bank Belgium.

Neef, T., & Sodano, A. (2022). *2022 DINA Regional Updae for Europe* (Technical Note No. 2022/02). World Inequality Lab.

Novokmet, F., Piketty, T., & Zucman, G. (2018). From Soviets to oligarchs: Inequality and property in Russia 1905-2016. *The Journal of Economic Inequality*, *16*(2), 189–223.

OECD. (2011). *Divided We Stand*. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264119536-en

OECD. (2015). *In It Together: Why Less Inequality Benefits All*. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264235120-en

OECD. (2021). *Does Inequality Matter? : How People Perceive Economic Disparities and Social Mobility* (tech. rep.). Paris. https://doi.org/10.1787/3023ed40-en

Piketty, T., Saez, E., & Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. *The Quarterly Journal of Economics*, *133*(2), 553–609. https://doi.org/10.1093/qje/qjx043

Ravallion, M. (2022). Missing Top Income Recipients. *The Journal of Economic Inequality*, *20*(1), 205–222. https://doi.org/10.1007/s10888-022-09530-0

Shorrocks, A. F. (1984). Inequality Decomposition by Population Subgroups [Publisher: [Wiley, Econometric Society]]. *Econometrica*, *52*(6), 1369–1385. https://doi.org/10.2307/1913511

Sutherland, H., & Figari, F. (2013). EUROMOD: The European Union tax-benefit microsimulation model. *International Journal of Microsimulation*, *6*(1), 4–26. https://doi.org/10.34196/ijm.00075

Vermeulen, P. (2016). Estimating the Top Tail of the Wealth Distribution. *American Economic Review*, *106*(5), 646–50. https://doi.org/10.1257/aer.p20161021

Vermeulen, P. (2018). How Fat is the Top Tail of the Wealth Distribution? *Review of Income and Wealth*, *64*(2), 357–387. https://doi.org/10.1111/roiw.12279

Waltl, S. R., & Chakraborty, R. (2022). Missing the wealthy in the HFCS: Micro problems with macro implications. *The Journal of Economic Inequality*, *20*(1), 169–203. https://doi.org/10.1007/s10888-021-09519-1

Zwijnenburg, J., Bournot, S., Grahn, D., & Guidetti, E. (2021). Distribution of household income, consumption and saving in line with national accounts. https://doi.org/https://doi.org/https://doi.org/10.1787/615c9eec-en

Appendix

A National Accounts: a brief introduction

The National Accounts (NA) form the backbone of the DINA framework. NA consist of a sequence of accounts which aim to give a complete view of economic activity taking place in a territorially defined unit (country or region) during a given period of time (year or quarter). Belgium follows the European System of Accounts (ESA 2010) as the standard framework for compiling its national accounts. It is used by Belgium's National Bank and the Federal Planning Bureau, to compile macroeconomic indicators such as GDP and government deficit, ensuring consistency with EU reporting requirements. ESA 2010 is aligned with the international System of National Accounts 2008 (SNA 2008), developed and supervised by the United Nations. While SNA 2008 serves as a global guideline, ESA 2010 is legally binding for all European Union member states and provides more detailed classifications and methodological requirements to ensure comparability across countries. We summarize the most important elements in a simplified representation in Table A.1 and Table A.2.

The national economy consists of five institutional units, called sectors and denoted by the letter 'S': non-financial corporations (S.11), financial corporations (S.12), general government (S.13), households (S.14) and non-profit institutions serving households (NPISH) (S.15). The national economy is denoted by S.1. We represent the sectors in the columns of Table A.1 and Table A.2. The domestic institutional units have economic interactions with a sixth institutional sector, the "rest of the world" (S.2).

To depict the creation of value by sectors and the transactions and interactions between the sectors, NA use a dual accounting system: transactions that increase economic value are recorded as "resources" (the right part of Table A.1 and Table A.2) and transactions that reduce the economic value of a unit are recorded as "uses" (the left part of Table A.1 and Table A.2).

The different stages in the economic process from the creation of value added in the production process up to expenditures by households are represented in a sequence of several accounts: current accounts (production of value added, distribution and redistribution of income), accumulation accounts (changes in the assets and liabilities of each unit), balance sheet accounts (accounting for stocks of assets and liabilities). We briefly describe them in more detail below. They show up as blocks one below the other in Table A.1 and Table A.2, and are denoted with a code starting with a roman numeral.

Each transaction is recorded with a name and a code. Codes which start with a "P" stand for transactions in products (goods and services); letter "D" stands for distributive transactions where the added value is distributed to labour, capital and government; letter "B" stands for balancing items in each account. By construction, to balance resources and uses or assets and liabilities in the dual accounting system, a balancing item is defined in each account. We show it in light grey in

Table A.1 and Table A.2.

The production account (I)

The *production account* registers transactions related to production carried out on the national territory. The value of the output of production is recorded in resources as $P.1$ and intermediate consumption as $P.2$ in uses. Value added shows up as the balancing item: it is calculated before (($B.1g$ with 'g' standing for gross) or after the consumption of fixed capital ($B.1n$ with 'n' standing for net). All these items appear in each of the sectoral accounts, indicated in Table A.1 and Table A.2 by a black dot in each of the columns. At the level of the total economy (column S1), gross domestic product is calculated at market price, and therefore indirect taxes like VAT ($D.21$) less subsidies on products ($D.31$) appear at the resource side of account I. Note that in the household sector (S.14), production is composed of two components: production by self-employed persons and housing services produced by households (with real gross rents or imputed for owner-occupied persons). Intermediate consumption in these cases also consist of two components: costs incurred by self-employed persons and costs related to housing services.

The production account is followed by the *distribution and use of income accounts*. This is a sequence of four accounts: the *primary income distribution account* (account II.1), the *secondary income distribution* (account II.2), the *redistribution in kind* (account II.3) and the *use of income* (account II.4).

The primary income distribution account (II.1)

The primary income distribution account (account II.1) registers the value added from the production process and distributes it among the factors of production (capital and labour) and the general government (through taxes less subsidies). There are two sub-accounts inside account II.1. The first one is the ***generation of income account*** (II.1.1) which focuses on each institutional sector as a producer. At the resource side this account starts with the added value produced by each sector, i.e. the $B.1g$ found in the uses side of the previous account. The uses column of account (II.1.1) shows how this added value covers the use of labour inputs for production purposes, i.e. compensation of persons employed by each sector ($D.1$), and taxes ($D.2$) minus subsidies ($D.3$) in each sector. Note that only part of taxes and subsidies is allocated to the separate institutional sectors ($D.29$ and $D.39$). Taxes and subsidies on products ($D.21$ and $D.31$) are only available at the level of the national economy (S1).

The balancing item in account II.1.1 is the remuneration of capital. It is called the operating surplus ($B.2_g$) and is available for each institutional sector. It is the surplus or deficit of the production activities in the sector before interest, rents or other charges on financial assets or natural resources have been paid or received. The gross operating surplus of the household sector consist of real and imputed rents. Within the household sector (S14) there is an additional balancing item, called mixed income ($B.3_g$. It represents income of unincorporated business (like self-employed) where one cannot distinguish the remuneration of labour from the one for capital. As DINA rely on net concepts, we need to shift from gross balancing items to net ones. In the publicly available National Accounts, for the household sector there is no breakdown of the consumption of fixed capital ($P51_{c,S14}$) between that attributable to the operating surplus ($B2g_{R,S14}$) and that attributable to mixed income ($B3g_{R,S14}$). But the economic outlook of the Federal Planning Bureau has figures on the *net* operating surplus and *net* mixed income in sector S14, and we have used these to allocate the consumption of fixed capital to the two balancing items.

The second account inside account II.1 is the ***allocation of primary income account*** (II.1.2). Unlike the previous one, this account focuses on the institutional units as recipients of primary income. The resources column shows the income of each institutional sector because of its participation in the production process or because it provides financial assets or natural resources to another institutional sector. For the financial and non-financial sector (S.11 and S.12), resources only consist of the operating surplus ($B.2_{g/n}$) and incomes from property ($D.4$). For the household sector (S.14), resources at this stage mainly consist of the compensation of employees ($D.1$), the operating surplus ($B.2_{g/n}$) and mixed income ($B.3_{g/n}$) and . Finally, also the government sector (S.13) records an operating surplus ($B.2_{g/n}$), mainly from taxes minus subsidies and net property income.

Compensation of employees $D.1$ is a good example to describe the difference in perspective offered on the one hand by account II.1.1 and on the other by account II.1.2. Since firms are production units, using labour as an input, the wages paid to workers in private firms are recorded as 'uses' for sectors S11 and S12 within account II.1.1. But inside account II.1.2, wages are registered as 'resources' in S.14, because households are the beneficiaries of this remuneration of labour. Note moreover that, at the level of the whole economy, $D.1_{S1}$ in account II.1.1 is not identical to $D.1_{S1}$ in II.1.2. The first one describes the amount that the national economy spends on compensations to employees, bot the residents and non-residents, as the counterpart of production processes on the national territory. The second one reflects the amount that the national economy receives as compensation of resident employees, part of whom take part in production processes outside the territory. The second one is obtained from the first one by removing the compensations of employees received by S2 (in the resources part of S2), and adding the compensations of employees paid by S2 (in the uses part of S2).

At the uses side of account II.1.2, we find for each sector the amount they pay for using assets from others institutional sectors ($D.4$). The balancing item in account II.1.2 is the gross or net balance of primary income, better known as gross or *net national income* ($B.5_{g/n}$). It is this aggregate which is the starting point for the distributional analysis in DINA.

secondary distribution of income account (II.2)

Table A.1 contains the concepts needed to establish DINA1, that is income components before any redistribution of primary income through taxes, benefits of expenditure on public goods takes place. The three other DINA income concepts (DINA2 to DINA4, see main text) elicit this redistribution at different stages. The main elements to construct DINA2 to DIN4, starting from DINA1, are summarized in a simplified way in Table A.2

The ***secondary distribution of income account*** reveals how primary income in each sector, entering the resource side as $B.5g$, is transformed into *disposable income* (the balancing item $B.6g$) by paying taxes ($D.5$) and social contributions ($D.61$) or receiving social benefits ($D.62$).

use of disposable income account (II.4.1)

The last account we use, is the ***Use of disposable income account*** (II.4.1). It shows how disposable income (now appearing at the resource side) is allocated to individual expenditures by households, by the government and by the non-profit sector, and to public expenditures by the government. The balancing item of account II.4.1 is gross saving ($B.8g$). It is this gross saving which forms the start of the accumulation accounts (III).

Table A.1: national accounts - production, income and allocation of primary accounts

Uses								Resources							
Code	Account	S1	S11	S12	S13	S14	S15	Code	Account	S1	S11	S12	S13	S14	S15
								Production account (I)							
P.2	Intermediate consumption	•	•	•	•	•	•	P.1	Output	•	•	•	•	•	•
								D.21	Taxes on products	•					
								D.31	Subsidies on products	•					
B.1g	Gross added value - GDP	•	•	•	•	•	•								
P.51c	Fixed capital consumption	•	•	•	•	•	•								
B.1n	Net added value	•	•	•	•	•	•								
								Generation of income account (II.1.1)							
D.1	Compensation of employees	•	•	•	•	•	•	B.1g	Gross added value - GDP	•	•	•	•	•	•
D.2	Taxes on production and imports	•	•	•	•	•	•								
D.3	Subsidies on products	•	•	•	•	•	•								
B.2g	Gross operating surplus	•	•	•	•	•	•								
B.3g	Gross mixed income	•				•									
P.51c1	Fixed capital consumption (GOS part)	•	•	•	•	•	•								
P.51c2	Fixed capital consumption (Mixed inc. part)	•				•									
B.2n	Net operating surplus	•	•	•	•	•	•								
B.3n	Net mixed income	•				•									
								Allocation of primary income account (II.1.2)							
D.4	Property income	•	•	•	•	•	•	B.2g/n	Gross/net operating surplus	•	•	•	•	•	•
								B.3g/n	Gross/net mixed income	•				•	
								D.1	Compensation of employees	•				•	
								D.2	Taxes on production and imports	•			•		
								D.3	Subsidies	•			•		
								D.4	Property income	•	•	•	•	•	•
B.5g	Gross national income	•	•	•	•	•	•								
B.5n	Net national income	•	•	•	•	•	•								

Black dots (•) refers to existing National Accounts concepts.

Table A.2: national accounts - secondary distribution and use of income accounts

Uses								Resources							
Code	Account	S1	S11	S12	S13	S14	S15	Code	Account	S1	S11	S12	S13	S14	S15
Secondary distribution of income account (II.2)															
D.5	Current taxes on income, wealth, etc	•	•	•	•	•	•	B.5g	Gross national income	•	•	•	•	•	•
D.6	Social contributions and benefits	•				•		D.5	Current taxes on income, wealth, etc	•			•		
D.7	Other current transfers	•	•	•	•	•	•	D.61	Net social contributions	•	•	•	•	•	•
								D.62	Social benefits other than in kind	•				•	
								D.7	Other current transfers	•	•	•	•	•	•
B.6g	Gross disposable income	•	•	•	•	•	•								
Use of disposable income account (III.4.1)															
P.31	Individual consumption expenditure	•			•	•	•	B.6g	Gross disposable income	•	•	•	•	•	•
P.32	Collective consumption expenditure	•			•			D.8	Adjustment for pension entitlements	•				•	
D.8	Adjustment for pension entitlements	•		•	•										
B.8g	Gross saving	•	•	•	•	•	•								

Black dots (•) refers to existing National Accounts concepts.

B Income surveys in Belgium

There are three, more or less consecutive, income surveys available that cover the total population of Belgian residents: the Socio-Economic Panel (SEP), the European Community Household Panel (ECHP) and the EU Statistics on Income and Living Conditions (EU-SILC). The SEP has been conducted in 1985, 1988, 1992, and 1997, the ECHP ran yearly from 1994 to 2001, and the EU-SILC started in 2004 and is still carried out yearly.

B.1 The Socio-Economic Panel (1985-1997)

The Socio-Economic Panel (SEP) is the oldest income survey for Belgium. The survey was organised by the Centre for Social Policy Herman Deleeck at the University of Antwerp. The survey has been conducted five times: in 1976, 1985, 1988, 1992 and 1997. We did not use the first survey of 1976 since it only covered the region of Flanders. The other waves were meant to be representative for the whole of Belgium. SEP differs from subsequent income surveys in two ways. First, incomes in SEP were surveyed on a monthly instead of on a yearly basis in ECHP and EU-SILC. Second, contrary to EU-SILC (but similar to ECHP), SEP only includes *net* household incomes, with the adjective *net* referring to income after personal income taxes and social contributions have been subtracted. To obtain gross income in SEP, we have used 'net-to-gross' routines from the microsimulation model MISIM. An interesting feature of SEP is that information on capital income and on total household wealth is available, be it in intervals and in different ways over the different surveys.

B.2 The European Community Household Panel (1995-2000)

ECHP is based on the Panel Study for Belgian Households (PSBH) and consists of eight yearly waves, containing socio-demographic characteristics for the period 1994-2001 and yearly disposable income and its components for the period 1993-2000. We use ECHP for all distributional information during this period but we adjust the original data in two ways.

First, similar to SEP, ECHP only includes *net* household incomes. To obtain gross income, we relied on net-to-gross trajectories which had been constructed for SEP for five different income components: gross employee cash income, self-employment income, old age and survivor pension benefits, sickness and disability benefits and unemployment benefits.

A second correction concerns imputed rents for owner-occupied houses. This variable is not available in the ECHP survey. We therefore predict imputed rents for home-owners using a hedonic regression approach as described by Frick et al. (2010) and similar to the method used by Statbel to estimate imputed rents in EU-SILC from 2006 onwards. We run a regression model with actual rents paid by tenants in private, non-subsidized markets as dependent variable and characteristics of the household and the dwelling as explanatory variables. We use the estimated coefficients of this regression to predict the imputed rents for home-owners. To account for possible selection bias between the groups of tenants and owners, we apply a Heckman selection correction as described by Hulliger and Wiegand (2012) by first estimating a probit regression to estimate the probability of being a tenant, and then incorporating the Mills-ratio in the linear regression model which explains the rents of tenants.

B.3 The EU Statistics on Income and Living Conditions (2003-2022)

EU-SILC is the yearly income survey conducted since 2004 by national statistical agencies of EU member countries, on behalf of EUROSTAT. We use the Belgian surveys from 2004 to 2023, reporting incomes for the period 2003-2022. In the 2019 survey (income year 2018) an important methodological change has been introduced in the Belgian survey. Prior to 2019, income information was exclusively survey-based, but from 2019 onwards the Belgian statistical agency Statbel introduced tax register data as primary source for most income variables.[34] A comparison of distributional results prior to 2019, with the ones of 2019 (and later) needs, therefore, to be done carefully. EU-SILC does contain information on taxes and benefits, but in a rather aggregate form. We enrich and expand the information on taxes and benefits by running the tax-benefit microsimulation model EUROMOD on gross incomes and household characteristics available in the EU-SILC data.[35]

[34]From 2019 onwards, Statbel used variables from the administrative dataset Belcotax for the following EU-SILC variables (variable-code between brackets): Employee income (PY010), contributions to individual private pension plans (PY035), pensions from individual private pension plans (PY080), unemployment benefits (PY090), pensions (PY100), survivors' pensions (PY110) and sickness and disability benefits (PY120 and PY130).

[35]Some policy years are missing in the EUROMOD-model. This is the case for the first two waves of EU-SILC (income years 2003 and 2004) and also for income year 2010. For the income years 2003 and 2004 we therefore relied on the 2005 EUROMOD policy. For income year 2010 we used the 2009 Euromod policy. For more information on the EUROMOD model, see Sutherland and Figari (2013).

C Combining income surveys with additional micro datasets

Income surveys are subject to certain limitations, which are addressed by leveraging additional datasets. Section C.1 details the methodology employed to adjust the distribution of earnings using administrative tax data to ensure representativeness at the upper end of the earnings distribution. Section C.2 presents an extension of EU-SILC with more reliable and detailed information on capital income, derived from the Household Finance and Consumption Survey (HFCS).

A total of four waves of HFCS data are currently available for analysis: the income years 2009, 2013, 2016, and 2019. We have developed inequality indicators for these adjusted data sets for all four years. However, when discussing the impact of the adjustments, our primary focus will be on the most recent year, 2019, unless there are substantial differences in the adjustment for the other survey years.

C.1 Top earnings adjustment

It is well-known that surveys often fail to capture the top of the income distribution (Flachaire et al., 2023; Ravallion, 2022). For obvious practical reasons, an income survey as the EU-SILC is only based on a sample of the total Belgian population. Due to small sample bias and imperfect participation of those included in the sample (i.e. unit non-response and item non-/mis-response), the sample might not be a perfect representation of the total Belgian population and might therefore not be adequate for accurately assessing (top sensitive) inequality measures.

The income concept

We use Belgian administrative personal income tax data (IPCAL) to assess the extent to which the earnings distribution in EU-SILC is representative for the Belgian population, and to which extent top incomes are missing. We compare taxable income earned on the labour market by active individuals (both employed and self-employed individuals) as registered in IPCAL with the distribution of the same income concept in EU-SILC. From EU-SILC, we use the sum of the employee cash income (`PY010G`), non-cash employee income (`PY020G`) and cash benefits or losses from self-employment (`PY050G`). To construct a similar income concept in IPCAL, we made a selection of variables based on De Schrijver (2020). For income year 2019, the selection of IPCAL variables is shown in Table A.3.

Representativeness

To increase the representativeness at the top of the income distribution, we use a method developed by Blanchet, Flores, and Morgan (2022), hereafter called "BFM-method". They define the representativeness at earnings level y as the survey density $f_{SILC}(y)$ to the true density $f_{IPCAL}(y)$:

$$\theta(y) = \frac{f_{SILC}(y)}{f_{IPCAL}(y)}. \tag{8}$$

Table A.3: Selection of IPCAL variables (2019)

Income from employment	v2500; v2510; v2520; v3080; v2470; v2540; v2420; v2430; v2400; v3350; v3370; v3950; v3970; v2630; v2730; v2740; v2750; v2760; v2770; v2780; v2790; v2800; v2670
Income from self-employment	v6000; v6320; v6200; v6110; v6060; v6500; v6580; v6590; v6520; v6560; v6750; v6690; v6570; v7080; v4500; v4510; v4520; v4000; v7243

A value of 1 indicates perfect representativeness. A value below (above) 1 indicates an overrepresentation (underrepresentation) of individuals in the income survey at earnings level y. The black dots in Figure A.1 indicate the representativeness $\theta(y)$ in 118 income intervals.[36] The representativeness decreases at the top of the income distribution. Some intervals lack observations in EU-SILC resulting in $\theta(y) = 0$. That is, however, not surprising given the relative small sample of EU-SILC - about 16,000 individuals in total, of which about 7,700 have non-zero earnings - and the limited interval size used at the top of the distribution. When one is interested in inequality indicators, this under-representation at the top may cause problems. Especially for indicators that are top-sensitive such as top income shares (top 1%, top 0.1%, or even smaller shares).

Choice of the merging point

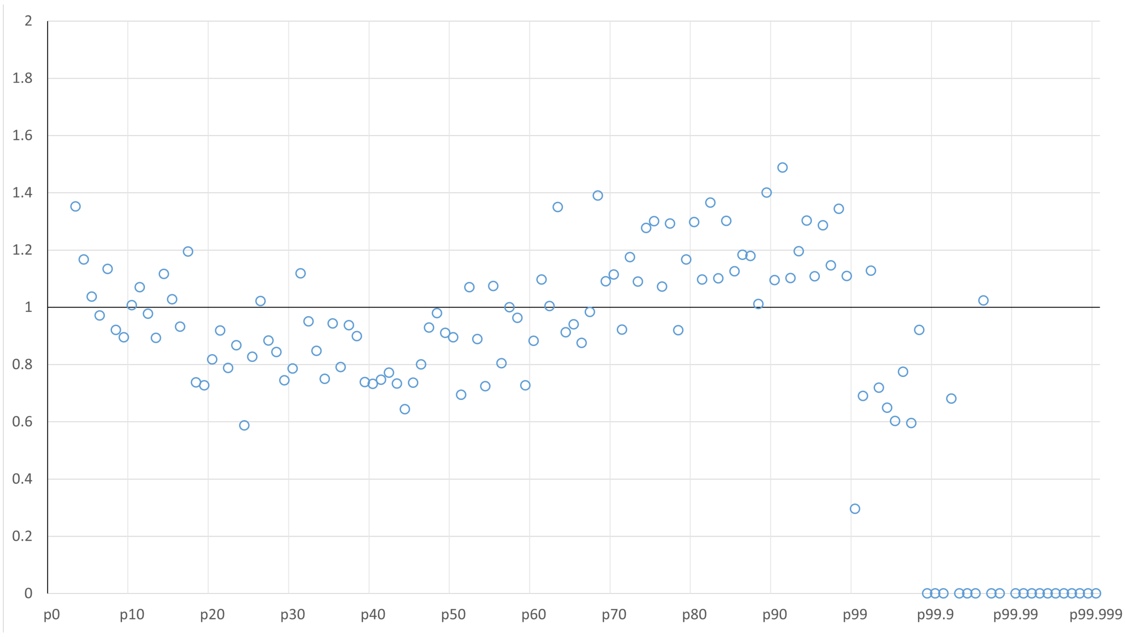
The key innovation of the BFM-method is to endogenously determine an income level, the so-called *merging point*, above which one starts to incorporate information from the tax data into the survey. After the determination of the merging point, the adjustment of the survey consists of two steps: reweighting and replacing. First, the survey weights are adjusted based on the income distribution of the tax data. This step increases the representativeness at the top of the income distribution using truncated linear calibration so that the distortion over the rest of the income distribution is minimized. Second, replacing is used above the merging point to increase the precision at the top of the income distribution.

The merging point $\tilde{y}$ is often arbitrarily set at a top income quantile (e.g., p90 of the income distribution) or the point where the two quantile functions cross. Blanchet, Flores, and Morgan (2022) formalize two objections to the latter approach and propose a data-driven way to select a merging point with desirable properties. Their selection of the merging point is based on the function $\theta(y)$ defined in Equation (8) and the cumulative rate of representativeness $\Theta(y)$, defined as:

$$\Theta(y) = \frac{F_{SILC}(y)}{F_{IPCAL}(y)} \tag{9}$$

[36]The first 99 intervals coincide with 99 percentiles based on the tax data which means that $f_{IPCAL}(y)$ is equal to 0.01 for these intervals. The last percentile p100 is split into 10 deciles (p99.0-p99.1, p99-1-p99.2, ..., p99.9-p100) with $f_{IPCAL}(y) = 0.001$ from which the top interval (p99.9-p100) is again divided in 10 smaller intervals (p99.90-p99.91, p99-91-p99.92, ... , p99.99-p100) with $f_{IPCAL}(y) = 0.0001$.

Figure A.1: Representativeness of EU-SILC (2019)



Note: Own calculations based on IPCAL and EU-SILC, income year 2019. Population of individuals with positive earnings. The dots indicate the representativeness of individuals in EU-SILC per income-interval.

The merging point $\tilde{y}$ is the highest value for which these two curves cross. The estimated empirical curves for $\Theta(y)$ and $\theta(y)$ are shown in Figure A.2.[37]

Reweighting

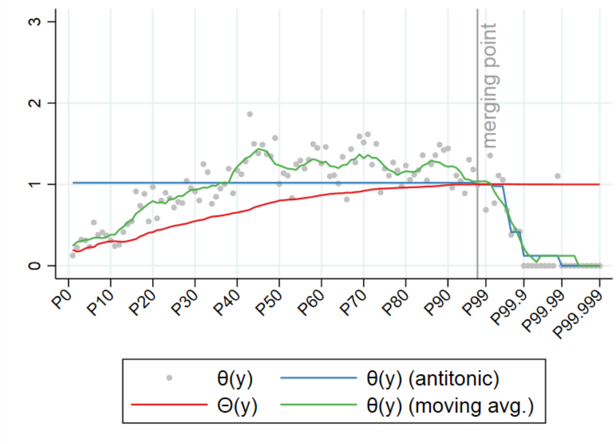
The determination of the merging point is followed by an adjustment of the sample weights. This sample weight w_i indicates how many Belgians are represented by respondent i in the survey sample. Summing up all sample weights equals the total Belgian population. Determining sample weights starts from the understanding that survey participation is determined by two random variables: the probability that individual i is included in the sample design of the survey, and the probability that individual i responds positively to the survey selection. Let S_i indicate whether an individual i is included in the sample design ($S_i = 1$) or not ($S_i = 0$) and R_i whether a selected individual participates in the survey ($R_i = 1$) or not ($R_i = 0$). The probability that a Belgian citizen is included in EU-SILC is given by the product of these two probabilities. Hence, the sample weight w_i for respondent i is the inverse of this product:

$$w_i = \frac{1}{\mathbb{P}(S_i = 1)} \times \frac{1}{\mathbb{P}(R_i = 1)}, \tag{10}$$

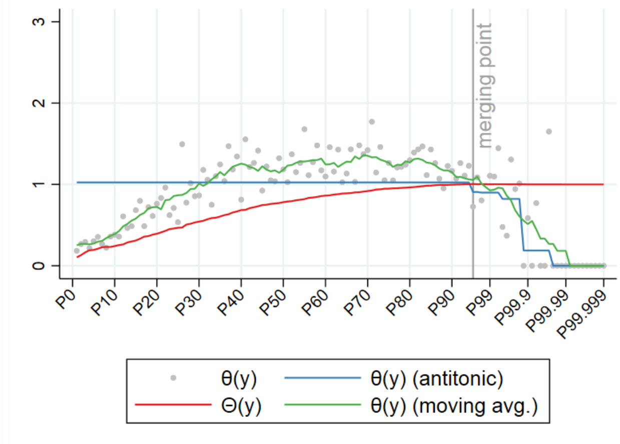
where $\mathbb{P}(S_i = 1)$ denotes the design probability (that i is included in the sample design) and $\mathbb{P}(R_i = 1)$ the response probability (that i responds to the survey).

[37]We rely on the `bfmcorr` command in Stata, developed by Blanchet, Flores, and Morgan (2022).

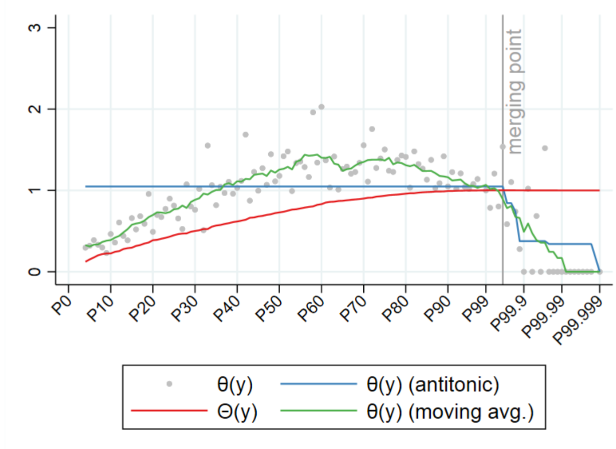
Figure A.2: Determination of the merging point



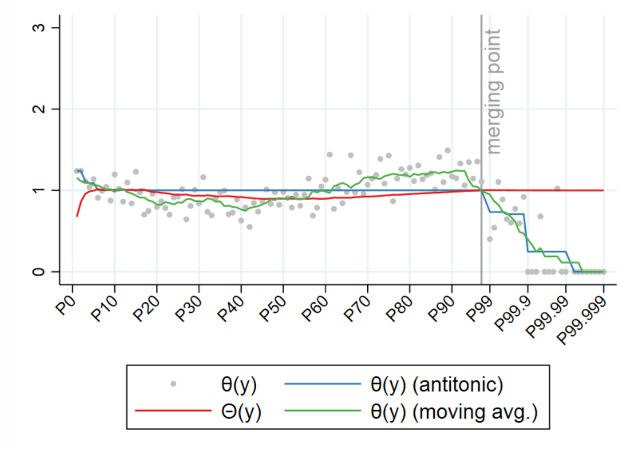
(a) 2009



(b) 2013



(c) 2016



(d) 2019

Note: Own calculations based on IPCAL and EU-SILC. Population of individuals with non-zero earnings.

The survey designers calibrate sample weights such that, for a limited number of socio-economic characteristics, like age, sex household size or region, the weighted value of these characteristics comes as close as possible to the known population value.[38] What the reweighting part of the procedure in Blanchet, Flores, and Morgan (2022) does, is to extend the number of characteristics on which representativiness is enforced, with one additional dimension: income, based on the IPCAL-dataset. Blanchet, Flores, and Morgan (2022) use truncated linear calibration to minimize deviations of the weights in the original survey data.

The income tax data are used in a tabulated form of 127 intervals (p0-p1, p1-p2, ..., p98-p99, p99.1-p99.2, ..., p99.91-p99.92, ... p99.991-p99.992, ... p99.9991-p99,9992, ... p99.999-p100). First,

[38]Since, by construction, the design probability $\mathbb{P}(S_i = 1)$ in (10) is known, this calibration amounts to an implicit modeling of the response probability $\mathbb{P}(R_i = 1)$ in function of these characteristics.

intervals below the merging point are merged to one interval. Next, the observations in the survey data are grouped to their corresponding bracket. Brackets are merged so that each interval contains at least 5 survey observations. The calibration of new weights w_i^* ensures that the shares of individuals in each bracket are in line with the population shares of each earnings bracket.

Replacing

After the reweighting step, the precision that we get at the top of the income distribution may still be insufficient for some purposes. Not only non-response but simply because the sample size is limited, we might miss high incomes in the survey. This can be problematic for inequality measurement, especially when focusing on the top of the income distribution. To account for these missing top incomes, one can not solely rely on reweighting. After reweighting, the next step consists of replacing incomes y_i of observations beyond the merging point with synthetic incomes y_i^*.

First, we replicate an individual i when their income is above the merging point ($y_i > \tilde{y}$). An individual is replicated k_i times with $k_i = r \times w_i^*$, rounded to the nearest integer. The sample weights w_i^* are adjusted accordingly by dividing them by k_i. This implies that all replicated observations have a weight equal to $1/r$. The default sampling rate is 10% ($r = 0.1$) which results in weights w_i^* close to 10 above the merging point.

Second, we rank the individuals in our dataset based on their income. We indicate the ranking of an individual i by an additional subscript j: $[y_{i_1}, y_{i_2}, ..., y_{i_j}, ..., y_{i_M}]$ with M the total number of observations (after replication) in the dataset. The following 'population share'-intervals indicate the share of the true population that is represented by each of the observations, with N the total population size (i.e., the total number of individuals with non-zero incomes in IPCAL):

$$\left[0, \frac{w_{i_1}^*}{N}\right], \left[\frac{w_{i_1}^*}{N}, \frac{w_{i_1}^* + w_{i_2}^*}{N}\right], ..., \left[\sum_{j=1}^{M} \frac{w_{i_j}^*}{N}, 1\right]. \tag{11}$$

These intervals are then used for replacing original incomes y_i by synthetic incomes y_i^* as follows. First, a continuous income distribution is estimated using the Generalized Pareto method of Blanchet, Fournier, and Piketty (2022). Second, we use the 'population share'-intervals to attribute synthetic earnings y_i^* to each observation above the merging point ($y_i > \tilde{y}$). The synthetic earnings y_i^* are set equal to the average income of their population share in the tax data.

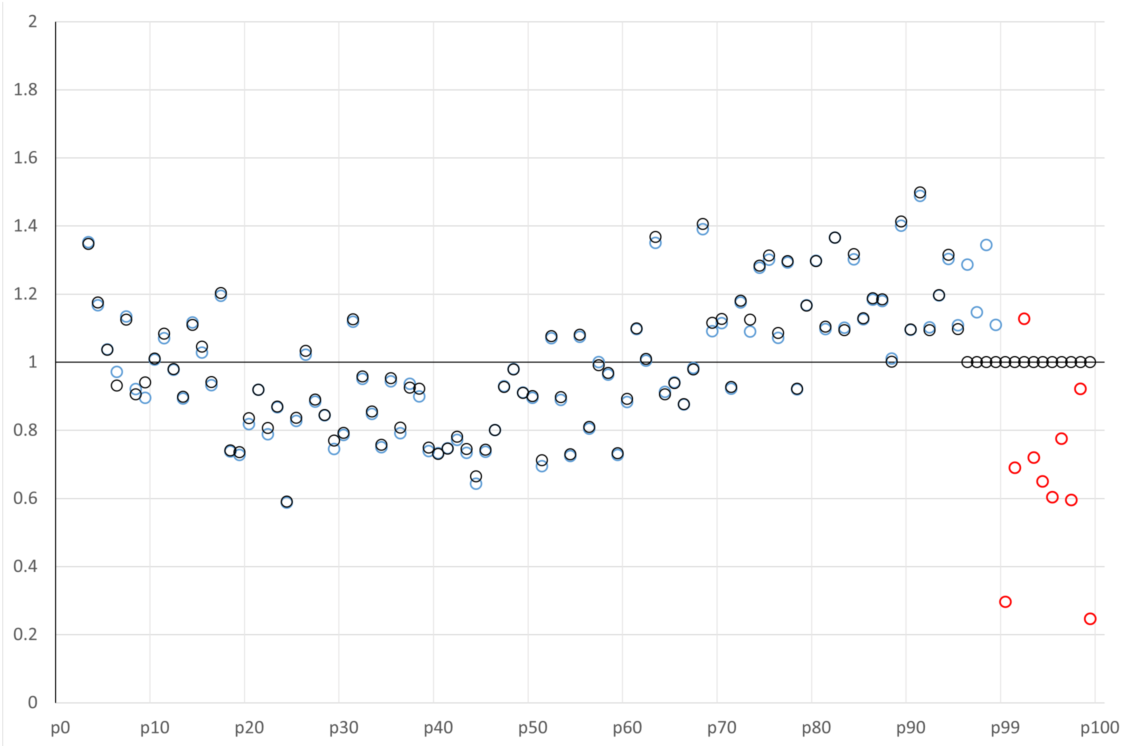
The intuition of replacing observations is well described by Blanchet, Flores, and Morgan (2022):

> *Intuitively, this process can be described as replacing the income of observations beyond the merging point with synthetic income observations with equivalent weight and rank in the tax distribution. This step ensures that we reproduce exactly the income distribution from tax data, that we preserve the survey's covariate distribution (including the household structure), and that we preserve the relationship between income and covariates from the survey data.* (Blanchet, Flores, & Morgan, 2022, p.134)

The adjusted earnings distribution

The impact of the adjustment is shown in Figure A.3. The blue dots are identical to the ones in Figure A.1, and above the merging point we have switched their colour to red. The black dots are the 'new' observations after applying the 'BFM-method'. The figure shows a clear impact on the representation for observations above the merging point.

FIGURE A.3: REPRESENTATIVENESS OF (ADJUSTED) EU-SILC (2019)



Note: Own calculations based on IPCAL and EU-SILC, income year 2019. Population of individuals with positive earnings. The black dots indicate the representativeness of individuals per income-interval in EU-SILC after the adjustment. The blue and, above the merging point, the red dots represent the representativeness before the BFM-correction.

Table A.4 shows the earnings distribution of the adjused EU-SILC (denoted by 'EU-SILC*') in comparison with IPCAL and the original EU-SILC. Table A.5 shows the impact on the Gini and top income shares.

Table A.4: Earnings distribution IPCAL vs EU-SILC (2019)

	Mean	p25	p50	p75	p90	p95	p99	p99.9	p99.99	p99.999	N
IPCAL	30662	13484	27677	40097	56805	71647	97402	223658	603361	1969923	5507828
EU-SILC	32068	14124	30485	43696	59686	73465	111435	229162	924400	924400	5698379
EU-SILC*	32685	14124	30532	43821	60098	74551	123252	295211	850317	2121044	5698379

Note: Own calculations based on IPCAL and EU-SILC income year 2016 and 2019. Population of individuals with non-zero earnings (employee and self-employment income). The mean and quantile values are expressed in euros; N indicates the total number of individuals with positive earnings (population weights used in EU-SILC).

Table A.5: Gini and top income shares - IPCAL vs EU-SILC (2019)

		Top income shares					
	Gini	10%	5%	1%	0.10%	0.01%	0.001%
IPCAL	0.4254	28.20%	17.98%	10.02%	2.49%	0.66%	0.09%
EU-SILC	0.4059	26.13%	15.90%	4.97%	0.94%	0.94%	0.94%
EU-SILC*	0.4157	27.34%	17.12%	6.16%	1.58%	0.41%	0.08%

Note: Own calculations based on IPCAL and EU-SILC income year 2019. Population of individuals with non-zero earnings (employee and self-employment income).

C.2 Income from capital

A second limitation of EU-SILC is the lack of high-quality information on household capital income. For instance, all financial capital income is captured in one variable (`HY090G`) which reports the *interests, dividends and profit from capital investments in unincorporated business*. Since capital income makes up a considerable share of net national income, it is of importance to have detailed information on this income component. We therefore make use of the Household Finance and Consumption Survey (HFCS) - which captures in-depth observations on household balance sheets - to impute households' asset portfolios into EU-SILC. This is done by pairing households in HFCS to households in EU-SILC using a Predictive Mean Matching (PMM) imputation method. A full overview of the imputation procedure and results can be found in Capéau, Decoster, Vanderkelen, et al. (2023). We apply this imputation to the top-income adjusted EU-SILC data, as described in previous subsection C.1. In addition, we perform a correction for the missing 'wealthy' observations since wealth surveys - for similar reasons as income surveys - often fail to capture the very wealthy. Finally, we compute returns on the imputed wealth of the households to replace the original information on capital income.

C.2.1 Data preparation

To impute the assets of HFCS in EU-SILC, we need to establish a set of common variables that will be used to match households in EU-SILC with households in HFCS. The following common variables are used:

1. **Gross household income.** The income concept we use, equals total gross household income (`DI2000` in HFCS and `HY010G` in EU-SILC), minus real estate income (`DI1300` in HFCS and `HY040G` in EU-SILC), financial income (`DI1400` in HFCS and `HY090G` in EU-SILC), and income stemming from other sources (`DI1800` in HFCS)[39].
2. **Number of employed household members.** This variable equals the sum of household members for which the economic status corresponds with *employee* (`PE0200` = 1 in HFCS and `PL031` = 1 or 2 in EU-SILC).
3. **Number of self-employed household members.** This variable equals the sum of household members for which the economic status corresponds with *self-employed* (`PE0200` = 2 in HFCS and `PL031` = 3 or 4 in EU-SILC).
4. **Number of unemployed household members.** This variable equals the sum of household members for which the economic status corresponds with *unemployed* (`PE0100a` = 3 in HFCS and `PL031` = 5 in EU-SILC).
5. **Number of highly educated household members.** This variable equals the sum of

[39] Only total gross household income (`DI2000`) is reported in the 2010 wave of HFCS. Therefore, instead of using our own derived gross income concept, we use total gross household income (`HY010G`) for the 2010 wave.

household members for which the highest education level completed (`PA0200` in HFCS and `PE040` in EU-SILC) is at least a bachelor or equivalent level (ISCED-level 6).

6. **Number of moderately educated household members.** This variable equals the sum of household members for which the highest education level completed (`PA0200` in HFCS and `PE040` in EU-SILC) equals ISCED-level 4 or 5.
7. **Number of household members per age class.** We define eight age classes: 0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+. These variables equal the sum of household members for which the age (`RA0300` in HFCS and `RB080` in EU-SILC) lies within the respective age class.
8. **Household type.** This is a dummy variable that equals 1 if the household consists of a single adult with or without dependent children (`DHHTYPE` = 51, 52, or 9 in HFCS and `HX060` = 5 or 9 in EU-SILC).
9. **Homeownership status.** We create three dummy variables corresponding to housing status. The first equals 1 if the household owns the dwelling without an ongoing mortgage (`HH021` = 1 in EU-SILC and `HB0300` = 1 in HFCS). The second equals 1 if a mortgage is still outstanding (`HH021` = 2 in EU-SILC and `HB0300` = 2 in HFCS). And the third dummy is equal to one if the house is rented or if any other homeownership status applies (`HH021` = 3, 4, or 5 in EU-SILC and `HB1000` = 2 in HFCS).

C.2.2 Imputation of capital assets

Table A.6 provides an overview of the asset classes in HFCS that are imputed in EU-SILC, a comparison of some descriptive statistics between the original HFCS data (denoted by *Obs.*) and the imputed assets in the top-corrected EU-SILC (denoted by *Imp.*), and – in the final two columns – a comparison of the aggregate value (over *all* Belgian households) of each asset. The results in Table A.6 indicate that, overall, the distributions of the original assets in HFCS and the imputed ones are fairly similar. Especially for assets with higher ownership rates, the imputed values closely match the observed values in HFCS.

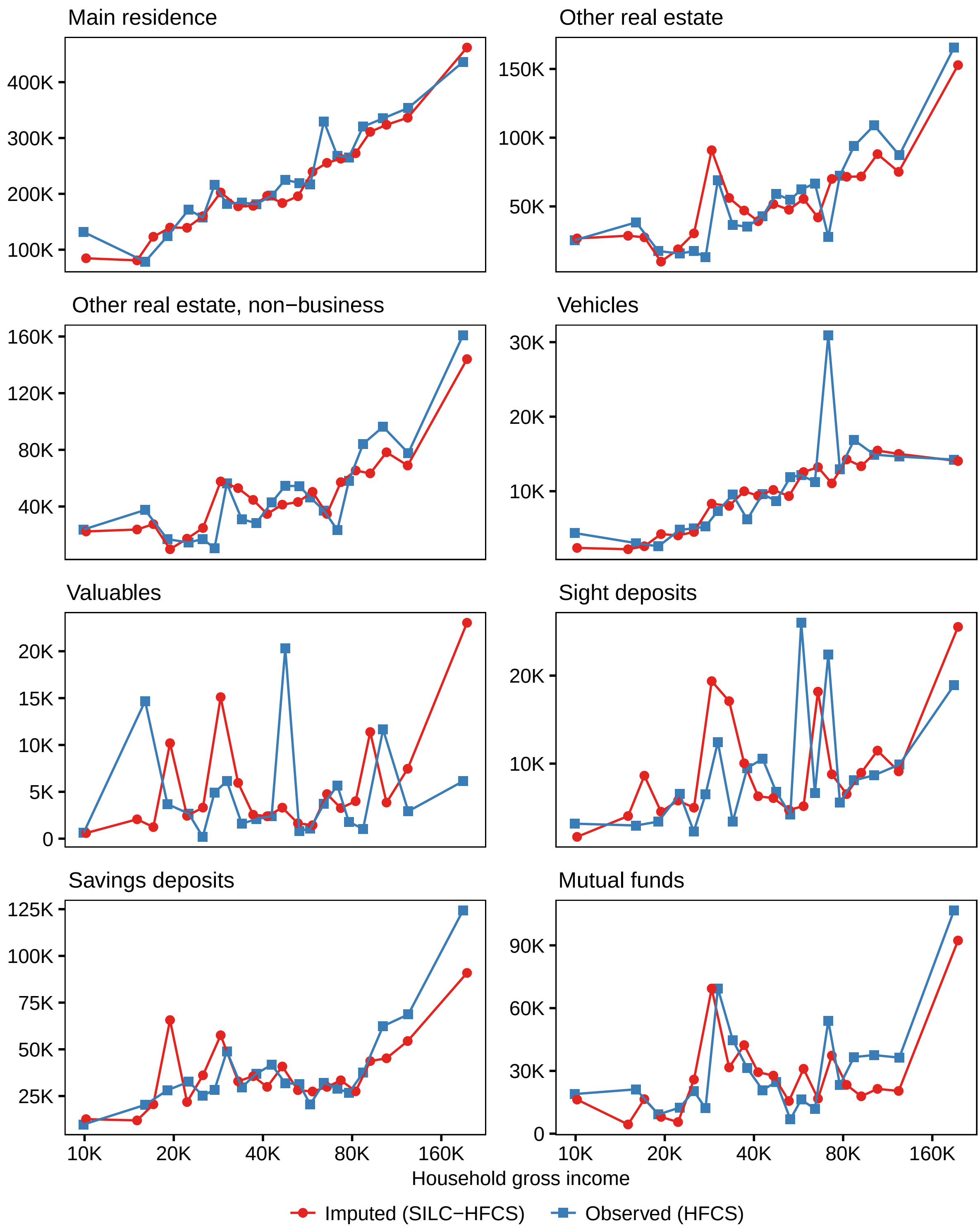
Figure A.4 presents ventile graphs of assets values of the original HFCS data and imputed assets. The imputed observations (the red line) track the observations in HFCS (the blue line) reasonably well.

Table A.6: Imputed assets from HFCS into the top-corrected EU-SILC (2019)

	% zero		Mean		Std. dev.		p25		p50		p75		Total (in bn €)	
	Obs.	Imp.	Obs.	Imp.	Obs.	Imp.	Obs.	Imp.	Obs.	Imp.	Obs.	Imp.	Obs.	Imp.
Real assets														
Main residence	27.6	33.1	317.5	323.1	194.6	186.9	200.0	202.3	299.4	300.0	384.8	400.0	1,154.9	1,060.3
Other dwellings	82.8	83.2	283.5	286.1	296.4	292.5	83.5	71.5	200.0	200.0	400.0	390.0	244.6	236.0
Vehicles and valuables	20.2	23.2	18.8	19.2	54.9	60.9	3.0	3.0	9.0	8.1	18.5	19.0	75.2	72.1
Financial assets														
Deposits	3.3	3.0	48.9	48.1	116.0	133.5	2.5	2.0	14.2	13.0	50.0	45.0	237.5	228.7
Sight deposits	4.0	3.7	9.3	9.7	51.1	53.8	0.6	0.6	2.0	2.0	6.0	5.0	44.9	45.9
Savings deposits	25.9	26.2	51.8	50.5	105.6	120.5	5.0	5.0	18.0	15.0	55.0	50.0	192.7	182.9
Mutual funds	77.2	79.6	135.4	135.8	317.7	282.1	10.0	12.0	35.6	38.0	130.0	125.0	154.9	135.6
Bonds	98.3	98.5	61.2	41.2	109.2	86.8	10.0	12.0	16.6	14.9	43.8	25.0	5.3	3.1
Private equity	93.0	92.5	331.3	375.4	927.8	1,290.7	5.0	2.3	40.9	25.0	233.2	140.0	116.7	137.6
Active role	93.9	93.6	369.7	429.2	988.0	1,382.6	5.0	2.3	50.0	30.0	300.0	150.0	113.1	135.3
Non-active role	98.8	98.7	62.1	37.3	105.7	94.9	3.0	1.2	11.8	5.0	68.5	39.0	3.6	2.3
Publicly traded shares	88.9	89.9	69.5	80.0	171.2	216.5	5.0	5.0	15.0	15.0	51.7	80.0	38.8	39.7
Managed accounts	99.2	98.3	191.7	307.5	555.9	468.9	3.5	13.5	63.0	100.0	150.5	300.0	8.0	25.5
Money owed to household	93.5	92.0	17.9	13.8	65.3	71.9	1.5	0.6	4.8	3.0	15.0	10.0	5.9	5.4
Voluntary pensions	94.3	94.2	40.3	81.6	297.3	336.6	2.9	5.0	9.8	13.0	21.9	35.4	11.6	23.0
Life insurance	62.3	65.0	25.9	23.8	50.3	40.2	5.7	6.0	14.2	15.0	26.4	27.7	49.0	40.9
Occupational pension plans	80.6	82.1	69.2	64.9	131.5	124.1	7.5	7.5	25.0	22.8	70.0	69.6	67.4	57.1
Liabilities	49.9	49.7	95.3	95.9	115.3	110.1	9.0	9.3	60.0	62.0	143.0	146.1	240.1	236.3
Net wealth (excl. pension plans)	0.7	0.7	411.1	388.0	780.7	809.6	65.5	33.0	243.6	215.9	472.5	439.5	2,051.0	1,888.4

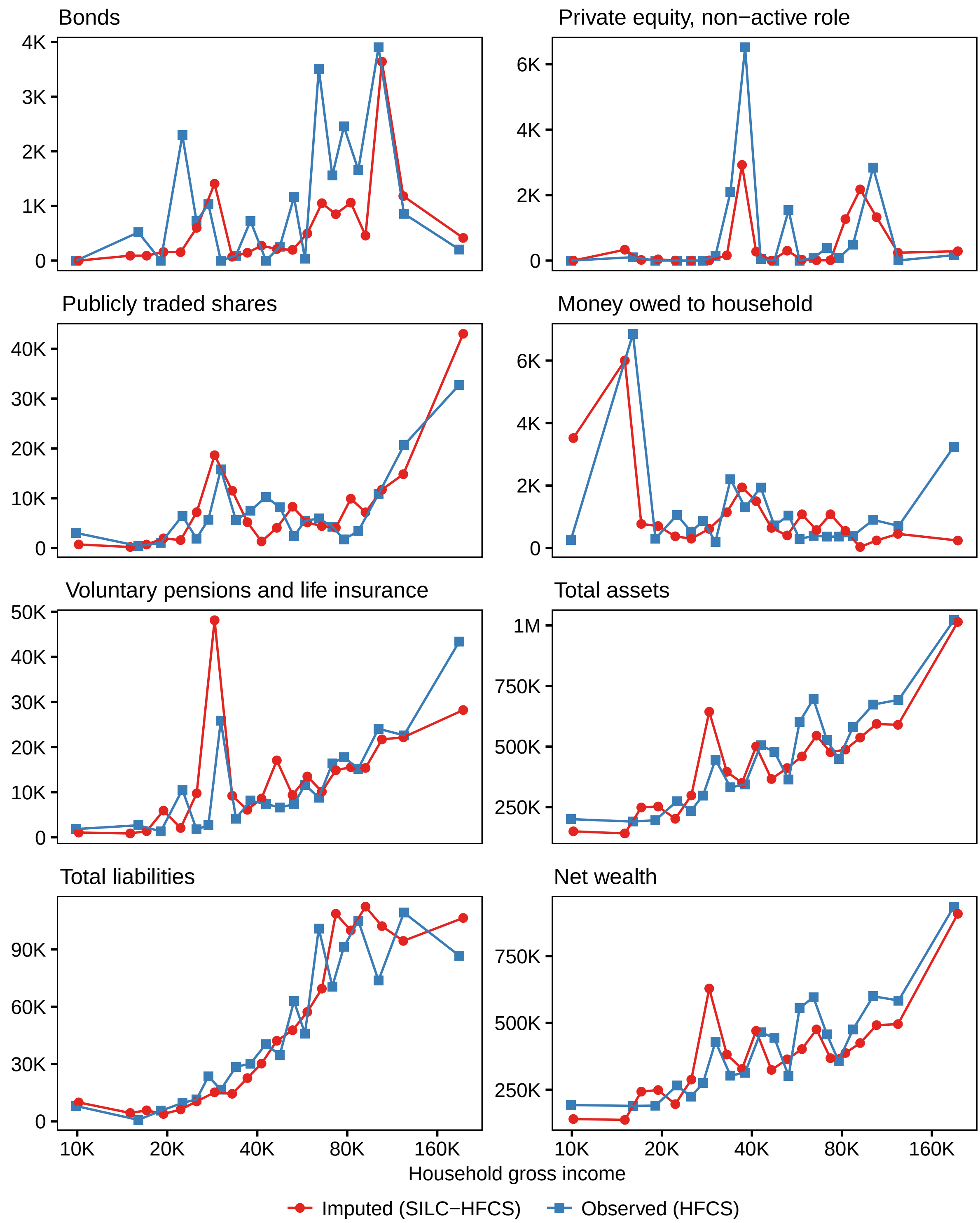
Note: All statistics are weighted using population weights and exclude households that do not own the respective asset. Obs. indicates the original observations in HFCS, Imp. denotes the imputed assets.

FIGURE A.4: MEAN IMPUTED (EU-SILC-HFCS) AND OBSERVED (HFCS) ASSET VALUES AGAINST MEAN HOUSEHOLD GROSS INCOME PER VENTILE



Note: Ventiles are based on gross household income. Households that do not own the respective asset are also included.

Figure 4 (Cont.): Mean imputed (EU-SILC-HFCS) and observed (HFCS) asset values against mean household gross income per ventile



Note: Ventiles are based on gross household income. Households that do not own the respective asset are also included.

C.2.3 Adjustment for the missing wealthy

For similar reasons as EU-SILC fails to capture the highest incomes, HFCS fails to capture the wealthiest households (see Vermeulen, 2016; or Waltl & Chakraborty, 2022). This downward bias at the top of the wealth distribution, which the literature often refers to as the *missing wealthy*, is especially problematic since wealth is heavily concentrated at the top.

A common method to correct for the *missing wealthy* is to estimate a Pareto tail for the top of the wealth distribution. With wealth w distributed according to the Pareto-distribution from level w_0 onwards, its distribution function reads as:

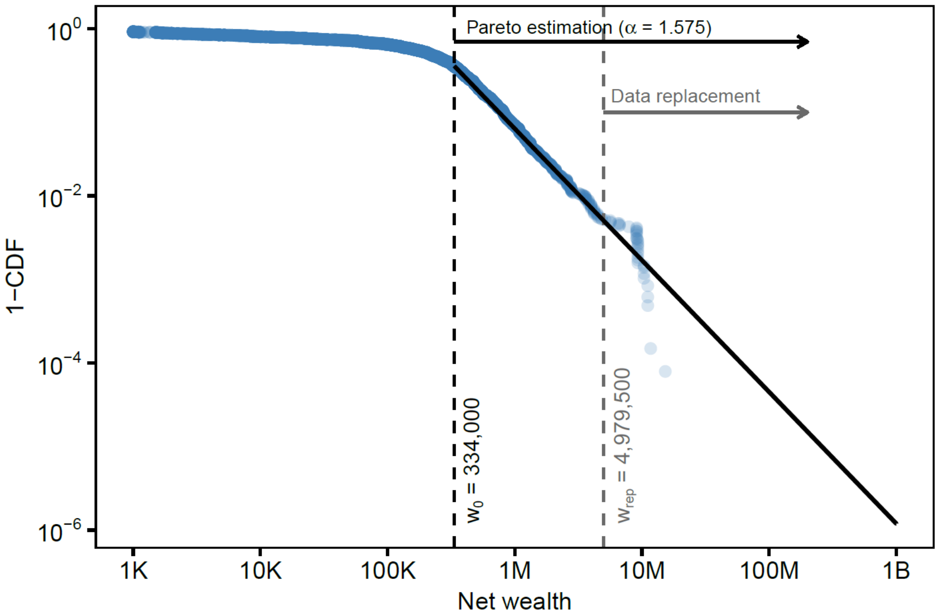
$$F(w; w_0, \alpha) = 1 - \left[\frac{w_0}{w}\right]^{\alpha} \tag{12}$$

with w_0 being the location parameter, and α de shape parameter. These two parameters can be estimated by relying solely on survey data, or by combining survey data with data from external sources (e.g., rich lists such as the Forbes World's Billionaires list or administrative tax data). Estimating the Pareto model on a combined dataset is usually preferred in the literature (e.g. Chakraborty & Waltl, 2018; Disslbacher et al., 2023; Vermeulen, 2018), as the very top of the wealth distribution is not well contained in the survey. However, we choose not to do this as we doubt the reliability of the external data (rich lists) available for Belgium. We believe that these rich lists should be treated with caution. Moreover, we find that the distributional impact of using these rich lists for Belgium is minimal.[40]

A summary of our approach to correct the top tail of the net wealth distribution is illustrated in Figure 5. First, we use the algorithm proposed by Disslbacher et al. (2023) to determine the location parameter w_0 so that the root mean squared error (RMSE) of the Pareto model is minimized. Next, we adopt a median quantile regression to estimate the shape parameter α, given our estimation of w_0 in the first step. Thereafter, we replace all survey observations with net wealth above a certain threshold – from where we judge that the survey data is not reliable any more – by net wealth observations drawn from this estimated Pareto distribution. Finally, all assets are scaled up proportionally in correspondence to the simulated net wealth. Each of these steps is explained in more detail below.

[40] The inclusion of the *Forbes World's Billionaires* list or the *Rijkste Belgen* (see https://derijkstebelgen.be/) list changes the estimated parameters hardly compared to the current method where no external data are used.

Figure 5: Complementary cumulative density function of EU-SILC-HFCS and Pareto estimation



Note: This figure shows the complementary cumulative density function of the survey data and the estimated Pareto distribution. We choose the location parameter w_0 (illustrated by the black dashed line) so that the RMSE of the Pareto model is minimized. We then estimate a Pareto model on the survey data (represented by the blue dots), of which the distribution is represented by the black line. Once the empirical distribution starts to deviate too much from the estimated Pareto distribution (i.e., starting from the gray dashed line), we replace survey observations with simulated net wealth observations drawn from the estimated Pareto distribution.

Estimation of the Pareto location and shape parameter

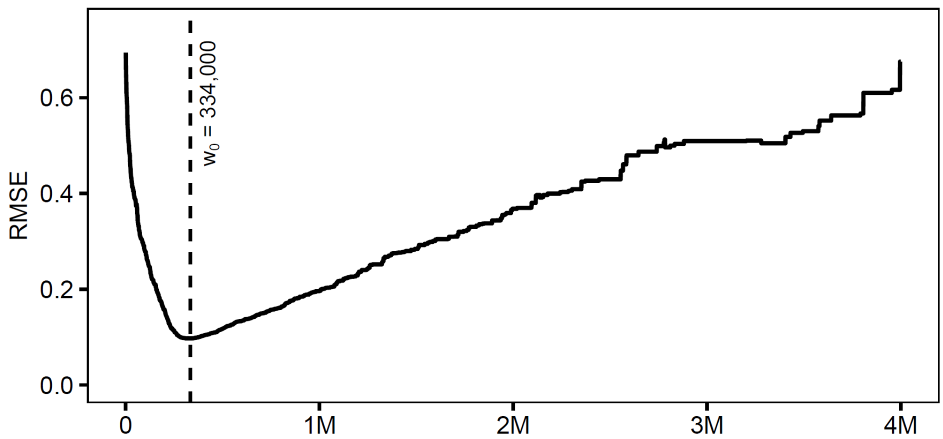
To estimate the Pareto tail, we strongly rely on the method proposed by Disslbacher et al. (2023). They estimate the shape parameter α using a median quantile regression on all survey observations with a net wealth above the location parameter w_0.[41] The regression equation is given by

$$\left(\ln(i-0.5)\cdot\frac{\overline{N_{fi}}}{\overline{N}}\right) = \underbrace{\ln\left(\frac{\overline{N}}{N}\right) + \alpha\ln(w_0)}_{\text{constant}} - \alpha\ln(w_i), \tag{13}$$

where i is a decreasing rank (i.e., $i = 1$ indicates the richest household), N is the total sum of weights, $\overline{N}$ is the average weight, and $\overline{N_{fi}}$ is the average weight of the first i observations.

We algorithmically determine the location parameter w_0 such that the root mean squared error (RMSE) of the Pareto regression is minimized. The intuition behind this is that when the RMSE is minimized, the observations above this threshold follow the most linear log-CCDF-log-wealth relationship, and hence they most closely approximate a Pareto distribution. Figure 6 illustrates the process of estimating w_0. In steps of €1,000, we estimate a Pareto model and compute its corresponding RMSE for candidate values of w_0 ranging from €0 to €4 million. This yields a location parameter of $\widehat{w_0}$ = €334,000 with a minimal RMSE of 0.0973. The estimated shape parameter given $\widehat{w_0}$ equals $\widehat{\alpha} = 1.575$.

FIGURE 6: ESTIMATION OF THE LOCATION PARAMETER w_0



Note: This figure illustrates our process of determining the location parameter w_0 of the Pareto model. In steps of €1,000, we estimate a Pareto model and compute its RMSE for candidate values of w_0 ranging from €0 to €4 million. The value of w_0 is chosen so that the RMSE of the Pareto regression is minimized.

[41] A quantile regression is expected to be more robust toward outliers compared to OLS (Waltl & Chakraborty, 2022)

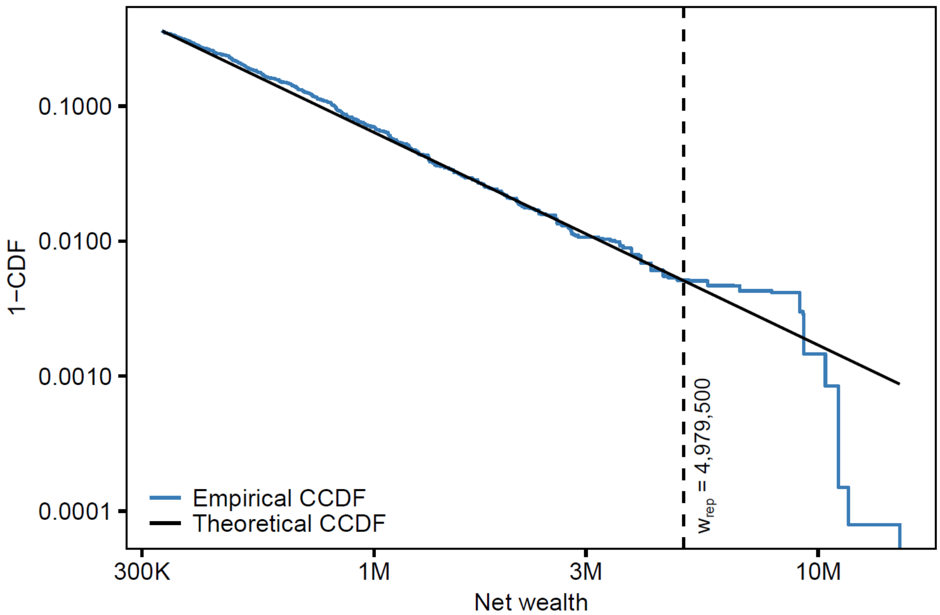
Estimation of the replacement threshold

Since we want to keep as many original survey observations in the Pareto tail as possible, we choose to only replace the net wealth observations above a certain replacement threshold $w_{rep} > w_0$. Figure 7 illustrates our procedure of determining w_{rep}. Assuming that net wealth in the tail is Pareto distributed, we determine w_{rep} as the level of net wealth from where the empirical CDF starts to deviate significantly from the theoretical CDF of the Pareto distribution. Formally, we first look for the minimum value of net wealth for which the following two conditions hold:

$$w_i > w_0,$$
$$\frac{|F_{emp}(w_i) - F_{Pareto}(w_i)|}{1 - F_{Pareto}(w_i)} > 0.5,$$

where w_i is net wealth of observation i, F_{emp} is the empirical CDF, and F_{Pareto} is the theoretical CDF. On the left of this point, we then look for the first point where the empirical and theoretical CDFs intersect. This point is our replacement threshold w_{rep}. By choosing the intersection of the two CDFs as the replacement threshold, we ensure that the observed number of households in the replacement tail is approximately equal to the number of households in the theoretical replacement tail, requiring only minimal reweighting of observations below the replacement tail to keep the total population size unchanged.

FIGURE 7: ESTIMATION OF THE REPLACEMENT THRESHOLD w_{rep}

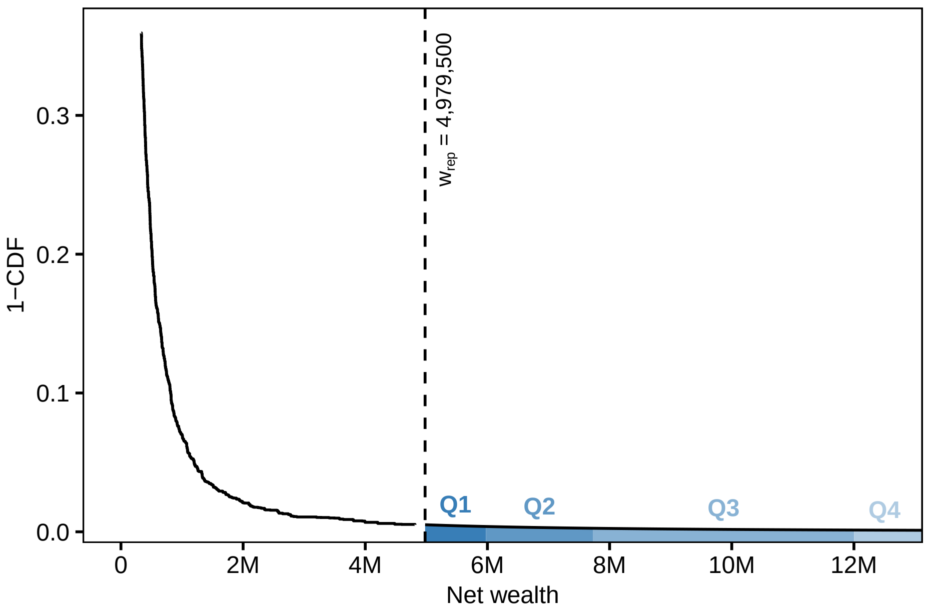


Note: This figure illustrates our process of determining the replacement threshold w_{rep}. First, we search for the point where the two CDFs already significantly deviate from each other. On the left of this point, we search for the first value of net wealth where the two CDFs intersect. This is our replacement threshold w_{rep}.

Simulation of tail observations

To construct a new micro-dataset that has been corrected for the missing wealthy, we proceed as follows. First, the survey observations above the replacement threshold are duplicated four times (with weights adjusted accordingly) to increase the granularity of the final dataset. Next, these survey observations are split into four quartiles of the observed net wealth, yielding strata Q1, Q2, Q3, and Q4. Thereafter, a random rank i is assigned to each observation within each stratum, implying that $i = 1$ will get assigned the highest value of net wealth. Per stratum, observations are sorted from the highest to the lowest rank and a new value for net wealth is drawn from the estimated Pareto distribution based on the empirical quantile of the observation. Before this, the weights of the observations in the replacement tail are adjusted so that the observed number of households in the replacement tail matches the theoretical number of households implied by the estimated Pareto distribution. The weights of observations under the replacement threshold are adjusted linearly so that the total number of households in the final dataset remains unchanged. Finally, all assets and liabilities of the households in the replacement tail are multiplied by a factor equal to the ratio of new and old net wealth.

FIGURE 8: TOP TAIL OF THE NET WEALTH DISTRIBUTION



Note: This figure shows the top tail of the net wealth distribution. Below the replacement threshold w_{rep}, we keep survey observations. Above w_{rep}, we draw new observations from the estimated Pareto distribution. Q1 to Q4 represent the quartiles of the simulated replacement tail.

C.2.4 Computation of capital incomes

Since the level of disaggregation in HFCS is larger for asset values than for capital income, we have chosen to impute asset values and not capital income. But within the framework of DINA, we are interested in capital incomes and not in the level of assets. We turned the asset values into incomes earned on this asset by using the implicit rates of return available in the national accounts. More specifically, we calculate the rate of return for asset class i in 2019 by dividing its corresponding income in the annual sector accounts for the household sector in 2019 by its corresponding level in the financial accounts for the household sector in 2018. Formally:

$$r_{i,2019} = \frac{Y_{i,2019}}{A_{i,2018}}, \tag{14}$$

where $Y_{i,2019}$ is the aggregate income of asset i in 2019 in Sector S.14 (*households*) of the national accounts, and $A_{i,2018}$ is the aggregate value of asset i at the end of 2018 in Sector S.14 of the national accounts. Note that in line with the national accounts standards, this rate of return does not include capital gains.

Income definitions between HFCS and the national accounts are not perfectly aligned. We therefore need to match the asset classes in HFCS to the income and level concepts in the national accounts. For this purpose, we use the manual of the European System of Accounts (ESA 2010) published by Eurostat (2013). An overview of the asset classes in HFCS, along with their corresponding income and value components in the national accounts, as well as the implicit rates of return for 2019, is given in Table A.7.

Table A.7: Asset classes in HFCS and their corresponding components in the national accounts with implicit rate of return for 2019

Asset class	Income in national accounts (Y_i)	Level in national accounts (A_i)	Return (r_i)
Main residence Other dwellings	Net operating surplus (B.2n)	Net value of dwellings (AN.111) Land underlying dwellings (AN.21111)	0.80%
Savings deposits Bonds	Total interest before allocation of FISIM (D.41g)	Other deposits (AF.29) Debt securities (AF.3)	0.77%
Mutual funds Managed accounts Voluntary pensions	Investment income attributable to collective investment fund shareholders (D.443)	Investment fund shares or units (AF.52)	1.49%
Private equity Publicly traded shares	Dividends (D.421)	Equity (AF.51)	4.71%
Life insurance	Investment income attributable to insurance policy holders (D.441)	Non-life insurance technical reserves (AF.61) Life insurance and annuity entitlements (AF.62)	2.62%
Occupational pensions	Investment income payable on pension entitlements (D.442)	Pension entitlements (AF.63)	2.15%

Note: Own composition based on Table 5.4 of the ESA 2010 manual (Eurostat, 2013).

D Correspondence between NA-aggregates and micro-counterparts

For the four DINA income concepts (DINA1 to DINA4), Table A.9 to Table A.12 show how we linked concepts in the micro-data (right side of the tables) with the concepts in the national accounts (left part of the tables). To assess the order of magnitude, we show the value for the national accounts concept in million €in the middle of the table, and at the left of the table we indicate whether this variable is added ('+') or subtracted ('-') to obtain an income aggregate.

Table A.13 documents the variables in HFCS (right side of the table) that are used to distribute some national account counterparts (left side of the table). These national account components are thus no longer linked to a variable from EU-SILC as depicted in the previous tables where we solely relied on distributional information from EU-SILC.

Table A.9: Link between NA and microdata for Pre-tax factor income (DINA income concept 1)

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
	A. Pre-tax personal factor labor income				
+	Wages and salaries	$D.11_{R,S14}$	181429	Gross cash or near cash employee income	PY010G
				Gross non-cash employee income	PY020G
+	Employers social contributions	$D.12_{R,S14}$	60697	simulated variable (EUROMOD)	EUROMOD
	B. Pre-tax personal factor mixed income				
+	Gross mixed income	$B.3g_{R,S14}$	32573	Cash benefits or losses from self-employment	PY050G**
-	Consumption of fixed capital	$P.51c_{S14}$	2637		
	C. Pre-tax personal factor capital income				
+	Gross operating surplus	$B.2g_{R,S14}$	27211	Imputed rents	HY030G
-	Consumption of fixed capital	$P.51c_{S14}$	17365	Income from rental of a property or land	HY040G
-	Interests paid by household sector	$D.41_{U,S14}$	1477	Interests repayents on mortgage	HY100G
+	Interests received by household sector	$D.41_{R,S14}$	1307	Interest, dividends, profit from capital investments in unincorporated business	HY090G
+	Distributed income of corporations	$D.42_{R,S14}$	17038		
+	Reinvested earnings on FDI	$D.43_{R,S14}$	0		
+	Investment income attributable to insurance policyholders	$D.441_{R,S14}$	5118	Gross cash or near cash employee income	PY010G
+	Investment income payable on pension entitlements	$D.442_{R,S14}$	2126	Gross non-cash or employee income	PY020G
+	Investment income attributable to collective investment fund shareholders	$D.443_{R,S14}$	2823	Interest, dividends, profit from capital investments in unincorporated business	HY090G
+	Primary income of corporate sector (undistributed profits)	$B.5g_{R,S11+S12}$	102232	Interest, dividends, profit from capital investments in unincorporated business	HY090G
-	Consumption of fixed capital	$P.51c_{S11+12}$	58780		
+	Rents (land) received	$D.45_{R,S14}$	696	*Neutral**	
-	Rents (land) paid	$D45_{U,S14}$	340		
	D. Pre-tax government income				

Continued on next page

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
+	Gross operating surplus	$B.2g_{R,S13}$	10653		
−	Consumption of fixed capital	$P.51c_{S13}$	10582		
+	Taxes on production and imports	$D.2_{R,S13}$	64277	*Neutral**	
−	Subsidies	$D.3_{R,S13}$	17928		
+	Property income received by gov sector	$D.4_{R,S13}$	4185		
−	Property income paid by gov sector	$D.4_{U,S13}$	9762		
E. Pre-tax non-profit income					
+	Gross operating surplus	$B.2g_{R,S15}$	303		
−	Consumption of fixed capital	$P.51c_{S15}$	303	*Neutral**	
+	Property income received by NPSIH sector	$D.4_{R,S15}$	157		
−	Property income paid by NPSIH sector	$D.4_{U,S15}$	12		
= Pre-tax factor income (NNI)			393639		

Notes:

*The neutral assumption implies that the distribution of the macro-aggregate does not change the distribution of pre-tax factor income. The macro-aggregate only acts as a level-shifter.

**This micro-concept in EU-SILC is allowed to be negative (losses for the self-employment). As the NA account total is positive, we enforced negative micro-values to be zero.

Table A.10: Link between NA and microdata for Pre-tax national income (DINA income concept 2)

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
Pre-tax factor income (DINA income concept 1)			393639		
−	Net social contributions	$D.61_{U,S14}$	89086		
	Employers' actual social contributions	$D.611_{U,S14}$	47036	simulated variables (EUROMOD)	EUROMOD
	Employers' imputed social contributions	$D.612_{U,S14}$	13361		
	Households' actual social contributions	$D.613_{U,S14}$	27692		
	Households' social contributions supplements	$D.614_{U,S14}$	2357		
−	Investment income payable to pension entitlements	$D.442_{R,S14}$	2126	See Table A.9	
+	Social security benefits in cash	$D.621_{R,S14}{}^{*}$	56047		
	Payments for sickness and invalidity	$[D.62\_1]^{**}$	9350	Sickness benefits Disability benefits	PY120G PY130G
	Unemployment	$[D.62\_2]$	4747	Unemployment benefits	PY090G
	Unemployment benefits with employer top-up ('early retirements' before 2012)	$[D.62\_3]$	891	Unemployment benefits	PY090G
	Career break and time-credit	$[D.62\_4]$	739	Unemployment benefits	PY090G
	Retirement pension and survival pension (private sector)	$[D.62\_5]$	32091	Old-age benefits	PY100G
	Widow's and orphan's fund	$[D.62\_6\_1]$	1178	Survivor benefits	PY110G
	Retirement pension and survival pension (public sector)	$[D.62\_6\_2]$ to $[D.62\_6\_8]$	5724	Old-age benefits	PY100G
	Industrial accidents	$[D.62\_9]$	278	Sickness benefits Disability benefits	PY120G PY130G

Continued on next page

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
	Occupational diseases	[$D.62\_10$]	245	Sickness benefits Disability benefits	PY120G PY130G
	Fund for the Closing of Corporations	[$D.62\_15$]	167	Unemployment benefits	PY090G
	Health and social care insurance in cash	[$D.62\_16\_1$]	453	Sickness benefits Disability benefits	PY120G PY130G
+	Other social insurance benefits	$D.622_{R,S14}$	21807	Old-age benefits Survivor benefits	PY100G PY110G
+	Pension and other social insurance surplus/deficit				
	+ Pension and other social contributions + Investment income payable to pension entitlements − Social security benefits in cash − Other social insurance benefits	$D.61_{U,S14}$ $D.442_{R,S14}$ $D.621_{R,S14}$ $D.622_{R,S14}$	13354	*Neutral****	
=	Pre-tax post-replacement income (NNI)		393639		

Notes:

*Up to 2014, child allowances were included in $D.621_{R,S14}$, the social security benefits in cash. We do not treat child benefits as social security benefits and exclude them for 2014 and earlier years from $D.621_{R,S14}$.

**The codes in square brackets do not refer to ESA2010 terminology but to NA concepts available in the table *Breakdown of paid social benefits* on NBB.Stat.

***The neutral assumption implies that the distribution of the macro-aggregate does not change the distribution of pre-tax post-replacement income. The macro-aggregate only acts as a level-shifter.

Table A.11: Link between NA and microdata for Post-tax disposable income (DINA income concept 3)

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
Pre-tax post-replacement income (DINA income concept 2)			393639		
–	Taxes on production and importation	$D.2_{R,S13}$	64277	*Neutral*	
+	Subsidies on production and importation	$D.3_{R,S13}$	17928		
–	Taxes on income	$D.51_{R,S13}$	72773		
	Advance tax payment on movable property	$[D.51\_A\_1]^{*}$	3700	Simulated variables	EUROMOD
	Business' advance tax payment	$[D.51\_A\_2]$	47635		
	Advance payments	$[D.51\_A\_3]$	1586		
	Assessments	$[D.51\_A\_4]$	-391		
	Annual tax on profit sharing	$[D.51\_A\_5]$	18		
	Special social contributions	$[D.51\_A\_6]$	1267		
	Contribution large incomes	$[D.51\_A\_7]$	0		
	Tax on the worker's participation in the benefit of the capital of the company	$[D.51\_A\_8]$	9		
	Other income taxes	$[D.51\_A\_9]$	33		
	Corporate income taxes	$[D.51\_B]$	17728	Interest, dividends, profit from capital investments in unincorporated business	HY090G
	Other income taxes	$[D.51\_E]$	1187	Simulated variable	EUROMOD
–	Other current taxes	$D59_{R,S13}$	2155		

Continued on next page

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
	Current taxes on immovable capital	[D.59_A]	861		
	Poll taxes	[D.59_B]	217	Simulated variables	EUROMOD
	Household payments for receiving licenses	[D.59_D]	1270		
	Other current taxes	[D.59_F]	23		
+	Social assistance benefits in cash	$D.623_{R,S14}$	13441	Family/children-related allowances	HY050G
	Child allowance (private sector)	[D.62_7]** and [D.62_8]	6888	Education related allowances Family/children-related allowances	PY140G HY050G
	Integration income	[D.62_11]	1445	Social exclusion not elsewhere classified	HY060G
	Guaranteed income for the elderly	[D.62_12]	598	Old-age benefits Survivor benefits	PY100G PY110G
	Compensation for handicapped	[D.62_13]	2304	Gross disability benefits	PY130G
	War pensions	[D.62_14]	47	Old-age benefits Survivor benefits	PY100G PY110G
	Corona unemployment benefits	[D.62_18]	0	Unemployment benefits	PY090G
	Bankruptcy / Replacement income self-employed	[D.62_19]	5	Per capita over all self-employed	
=	Post-tax disposable income (≠ NNI)		285804		

Notes: *Codes in square brackets as breakdown of D.5 do not refer to ESA2010 terminology but to NA concepts available in the table *Received taxes and actual social contributions by kind* on NBB.Stat.
**Codes in square brackets as breakdown of D.62 do not refer to ESA2010 terminology but to NA concepts available in the table *Breakdown of paid social benefits* on NBB.Stat.

Table A.12: Link between NA and microdata for Post-tax national income (DINA income concept 4)

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
=	Post-tax disposable income (≠ NNI)		285804		
+	Individual consumption expenditure (gov)	$P.31_{U,S13}$*	72067		
	Health	[F.07]	33136	*Equally among population*	
	Recreation, culture and religion	[F.08]	1961	*Neutral: Follow the distribution of total of what precedes*	
	Education	[F.09]	27135		
	Sickness and disability	[F.10.1]	4399		
	Old age	[F.10.2]	737		
	Survivors	[F.10.3]	24		
	Family and children	[F.10.4]	2068		
	Unemployment	[F.10.5]	494		
	Housing	[F.10.6]	227		
	Social exclusion	[F.10.7]	1891		
+	Individual consumption expenditure (non-profit)	$P.31_{U,S15}$	5049	*Neutral: Follow the distribution of total of what precedes*	
+	Collective consumption expenditure	$P.32_{U,S13}$	38224		
	General public service	[F.01]	11113	Neutral: Follow the distribution of total of what precedes	
	Defence	[F.02]	3535		
	Public order and safety	[F.03]	7466		
	Economic affairs	[F.04]	11279		
	Environment protection	[F.05]	2028		
	Housing	[F.06]	479		
	Health	[F.07]	476		
	Recreation, culture and religion	[F.08]	1040		
	Education	[F.09]	385		
	R&D social protection	[F.10.8]	0		
	Social protection n.e.c.	[F.10.9]	418		

Continued on next page

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
+	Government surplus/deficit	-7504			
	+ Net taxes	$D.2_{R,S13} - D.3_{R,S13}$	46349		
	+ Taxes on income of wealth	$D.5_{R,S13}$	74927		
	− Social assistance benefits in cash	$D.623_{R,S14}$	13441	*Neutral: Follow the distribution of total of what precedes*	
	− Individual consumption (gov)	$P.31_{U,S13}$	72067		
	− Individual consumption (non-profit)	$P.31_{U,S15}$	5049		
	− Collective consumption expenditure	$P.32_{U,S13}$	38224		
=	Post-tax national income (NNI)		393639		

Notes: *Codes in square brackets do not refer to ESA2010 terminology but to NA concepts available in the table *Government spending by functions and transactions* on NBB.Stat.

Table A.13: Link between HFCS information and elements of the National Accounts)

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
Components of pre-tax factor income (capital income components)					
+	Gross operating surplus	$B.2g_{R,S14}$	27211	Value of household's main residence	da1110
−	Consumption of fixed capital	$P.51c_{S14}$	17365	Value of other real estate property	da1120
−	Interests paid by household sector	$D.41_{U,S14}$	1477	Interest payments	di1412
+	Interests received by household sector	$D.41_{R,S14}$	1307	Saving accounts	da21012
				Bonds	da2103
+	Distributed income of corporations	$D.42_{R,S14}$	17038	Value of non self-employment private business	da2104
				Publicly traded shares	da2105
+	Reinvested earnings on FDI	$D.43_{R,S14}$	0		
+	Investment income attributable to insurance policyholders	$D.441_{R,S14}$	5118	Current value of whole life insurance	pfa080 (pfa020 = 4)
+	Investment income payable on pension entitlements	$D.442R, S14$	2126	Current value of (occupational) pension plan	pfa080 (pfa020 = 2)
+	Investment income attributable to collective investment fund shareholders	$D.443_{R,S14}$	2823	Current value of (voluntary) pension plan	pfa080 (pfa020 = 3)
+	Primary income of corporate sector (undistributed profits)	$B.5g_{R,S11+S12}$	102232	Value of non self-employment private business	da2104
−	Consumption of fixed capital	$P.51c_{S11+12}$	58780	Publicly traded shares	da2105
Components of post-tax national income					
−	Employers' imputed social contributions	$D.612_{U,S14}$	13361	Current value of (occupational) pension plan	pfa080 (pfa020 = 2)
−	Benefits from pension funds and other	$D.622.P_{R,S14}$	7753	Gross income from occupational and private pension plans	pg0410
Components of post-tax disposable income (taxes)					
−	Household payments for receiving licences	$[D.59\_D]$	1270	number of cars	hb4310

Continued on next page

	Income component (National Accounts)			Distributional information (microdata)	
	Description	Code	2019 (€m)	Description	Code
–	Corporate income taxes	[$D.51\_B$]	17728	Value of non self-employment private business	da2104
				Publicly traded shares	da2105
–	Advance tax payment on property	[$D.29\_A\_1$]	3884	Simulated variable*	
–	Advance tax payment on movable property	[$D.51\_A\_1$]	3066	Simulated variable*	
–	Taxes on security accounts	[$D.59\_G$]	217	Simulated variable*	

Notes:
*We were able to simulate capital income taxes, based on the information available from the imputed HFCS-variables.

E Additional results

E.1 Decomposition of pre-tax factor income inequality

TABLE A.14: DECOMPOSITION OF PRE-TAX FACTOR INCOME INEQUALITY IN CONTRIBUTIONS OF INCOME SOURCES k, 1985-1997

	1985	1987	1992	1997
Shares s_k as % of total pre-tax factor income				
Labour income	62.1	59.0	61.6	63.0
Financial income	25.0	28.7	27.1	25.1
Non-financial capital income	2.8	2.7	2.0	2.2
Self-employment income	10.1	9.7	9.3	9.7
Gini G_k				
Labour income	58.8	61.4	57.8	61.6
Financial income	73.4	73.3	73.3	71.1
Non-financial capital income	63.1	62.2	55.6	53.8
Self-employment income	88.7	89.2	90.2	92.0
Gini correlation R_k				
Labour income	83.0	81.3	80.8	86.5
Financial income	71.0	71.0	76.2	73.1
Non-financial capital income	16.1	26.7	30.1	21.5
Self-employment income	34.6	36.9	41.0	52.4
Average contribution to the Gini $(s_k \cdot G_k \cdot R_k)/G_k$				
Labour income	64.8	61.3	60.3	65.2
Financial income	27.9	31.1	31.8	25.3
Non-financial capital income	0.6	0.9	0.7	0.5
Self-employment income	6.7	6.6	7.2	9.1
Gini of pre-tax factor income G	46.7	48.0	47.7	51.5

Note: The upper three panels show the three components of income sources k of the overall Gini G of pre-tax factor income from Equation 6: the shares s_k, Gini coefficients G_k and the Gini correlations R_k. The bottom panel includes the average contributions to the overall Gini of pre-tax factor income $(s_k \cdot G_k \cdot R_k)/G_k$ expressed as a percentage of the Gini in the bottom line.

Table A.15: Decomposition of pre-tax factor income inequality in contributions of income sources k, 2003-2022

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Shares s_k as % of total pre-tax factor income																				
Labour inc	68.0	67.1	66.9	67.0	66.7	68.4	71.1	69.7	71.8	70.8	70.9	69.9	68.6	68.4	68.8	69.1	68.7	69.7	67.4	67.4
Financial inc	21.0	22.2	22.4	22.5	23.6	22.6	19.3	20.6	18.1	18.9	18.5	19.5	20.4	20.6	20.1	19.9	20.5	19.5	22.2	23.2
Non-fin. Cap inc	2.1	1.9	2.0	1.7	1.1	0.6	0.9	1.1	1.5	1.9	2.2	2.1	2.3	2.4	2.5	2.5	2.4	2.6	2.3	1.5
Self-empl. Inc	8.9	8.8	8.7	8.8	8.6	8.5	8.6	8.6	8.5	8.4	8.4	8.5	8.6	8.6	8.6	8.6	8.5	8.2	8.2	7.9
Gini G_k																				
Labour inc	57.5	59.8	57.8	57.5	56.9	56.5	56.9	56.9	57.2	56.5	57.3	58.0	58.4	58.6	57.1	58.0	57.8	58.0	57.1	57.4
Financial inc	75.0	76.9	76.4	74.2	73.9	74.5	72.3	70.4	74.8	74.8	78.4	80.6	83.7	85.6	84.5	86.4	88.1	87.1	89.4	90.4
Non-fin. Cap inc	63.0	70.7	65.7	66.8	65.2	63.8	68.4	65.5	66.6	66.7	65.0	63.4	62.2	62.5	62.0	58.9	59.5	57.6	57.3	57.3
Self-empl. Inc	94.8	94.6	93.6	93.7	94.6	93.7	93.8	94.9	94.5	93.8	93.6	93.5	93.2	93.0	93.1	92.2	92.0	92.3	92.3	92.0
Gini correlation R_k																				
Labour inc	85.6	86.0	86.1	85.8	86.2	86.0	88.4	89.3	88.9	87.6	88.1	86.7	87.6	87.4	87.5	89.0	88.7	89.2	87.7	89.7
Financial inc	74.4	77.8	79.0	76.7	76.5	75.3	76.2	75.5	77.6	79.7	81.1	83.2	86.3	89.3	86.8	88.7	89.8	88.9	92.4	91.9
Non-fin. Cap inc	11.0	15.5	14.2	11.8	8.4	4.3	1.0	5.7	5.7	12.9	15.1	16.8	13.1	15.2	18.2	13.3	14.0	13.6	14.4	10.9
Self-empl. Inc	65.8	62.5	58.0	60.3	60.5	55.5	59.3	63.0	61.4	58.9	58.5	58.6	59.5	59.8	56.8	55.2	55.2	54.2	54.9	52.2
Average contribution to the Gini $(s_k \cdot G_k \cdot R_k)/G_k$																				
Labour inc	65.8	64.8	64.3	64.9	64.1	66.0	69.8	68.7	70.2	68.5	68.3	66.2	64.0	62.8	63.7	64.3	62.9	65.1	59.8	60.0
Financial inc	23.0	25.0	26.2	25.1	26.2	25.2	20.8	21.2	20.2	22.1	22.5	24.6	26.9	28.2	27.3	27.5	29.0	27.2	32.5	33.3
Non-fin. Cap inc	0.3	0.4	0.4	0.3	0.1	0.0	0.0	0.1	0.1	0.3	0.4	0.4	0.3	0.4	0.5	0.4	0.4	0.4	0.3	0.2
Self-empl. Inc	10.9	9.8	9.2	9.7	9.6	8.8	9.3	9.9	9.5	9.1	8.8	8.8	8.7	8.6	8.5	7.9	7.7	7.4	7.4	6.6
Gini of pre-tax factor income G																				
	50.9	53.2	51.7	51.0	51.0	50.3	51.3	51.6	52.1	51.1	52.3	53.1	54.9	55.8	54.0	55.4	55.9	55.5	56.4	57.8

Note: The upper three panels show the three components of income sources k of the overall Gini G of pre-tax factor income from Equation 6: the shares s_k, Gini coefficients G_k and the Gini correlations R_k. The bottom panel includes the average contributions to the overall Gini of pre-tax factor income $(s_k \cdot G_k \cdot R_k)/G_k$ expressed as a percentage of the Gini in the bottom line.