Measuring Capital-Labour Shares and Inequality

Increasing Gaps between National Accounts and Micro-data

Ignacio Flores[*†]

[*][1] INSEAD: Office 2-10, Boulevard Constance, 77305 Fontainebleau (for correspondence and/or offprints) [2] Paris School of Economics: Office R5-01, 48 Boulevard Jourdan, 75014 Paris. [3] e-mail: i.floresbeale@gmail.com. Replication codes and data are available at: `https://github.com/ignacio-flores/income-under-the-carpet`

Abstract

As a response to the recent interest for reconciling macro- and micro-economic estimates of income, I measure the relative underestimation of capital and labour incomes in distributive data with respect to the system of national accounts (SNA). I study harmonised household surveys in a group of countries but also administrative records for the US. I find a higher underestimation of capital income, due to both conceptual differences and measurement error. The gap has grown in most countries during the last 20 years. Surveys therefore tend to exaggerate the impact of changes in the labour income distribution and to undermine those in the distribution of capital income. I present a novel theoretical framework, based on accounting identities, which shows that inequality estimates are likely affected in level, trend and composition. Yet, the intensity of the phenomenon varies by country and income concept (national vs. household income). To give a sense of magnitudes, in a 19-countries balanced panel with comparable SNA, the household sector receives around half of gross capital income, as opposed to corporations. In the subset of countries with comparable surveys, these only capture around 20% of that aggregate, versus 70% of labour income. For any quantile in the distribution, a unit increase in its share of labour income translates in a 0.91 increase in its share of total income, while the effect for capital income is only 0.09. Gaps are narrower in tax data but still present.

Keywords: Capital Income, Inequality, Top income shares, Measurement error.

Classification: D31, D33, G35.

Introduction

During the last 50-60 years most developed countries recorded a substantial growth in their *capital share* of national income (IMF 2007; Arpaia, Pérez, and Pichelmann 2009; Piketty and Zucman 2014; Karabarbounis and Neiman 2014). That is, the part of macroeconomic income remunerating capital, as opposed to labour, has been growing for decades. This phenomenon occurs in parallel to the increasing overall income-concentration recorded by Atkinson, Piketty, and Saez 2011 and Alvaredo et al. 2018, which has led researchers to explore the relationship between national factor shares (i.e. capital or labour shares) and inequality. From a theoretical perspective, accounting identities depict a clear and mechanical relation between these varibles (Atkinson and Bourguignon 2000; Milanovic 2017), yet its empirical assessments are sometimes contradictory and findings seem to depend largely on underlying data, which leaves room for ambiguity (Bengtsson and Waldenström 2018; Francese and Mulas-Granados 2015). How to explain discrepancies between theoretical and empirical findings in this literature? How does the choice of data sources affect estimates of economic inequality and their factor-income composition? How does the macroeconomic distribution of income translate into microeconomic inequality?

This article aims to answer these questions by putting a special focus on discrepancies –or data gaps– on aggregate income, as measured in different sources: Namely, harmonised national accounts from United Nations and distributive data, such as harmonised household surveys from the Luxembourg Income Studies and administrative records from the American Internal Revenue Service (IRS). As standard, I consider national accounts to be the benchmark, mostly because its production is given first-order political priority, involving bigger investments, more and better resources. Indeed, not only national accountants use distributive data as an input in the process, but they also have access to a wider variety of sources, including restricted data (United Nations 2009).

Aggregate incomes from national accounts –from which factor shares are estimated– are often substantially higher than those reported in distributive data sets –from which inequality estimates are computed. This "data gap" is due to both conceptual differences and measurement errors. The most obvious conceptual difference is that the System of

National Accounts (SNA henceforth) considers all incomes in the economy, including undistributed profits that are retained in corporations, while distributive data deliberately only considers household incomes (OECD 2013). Yet, even if we focus on household income alone, the SNA still uses broader definitions and generally reports larger aggregate income than the one reported in microdata (Zwijnenburg 2016; Zwijnenburg et al. 2017). For instance, accruing to households' capital income, the SNA counts items, such as imputed income of insurance policy holders or income reinvested abroad by nationals, which are absent from distributive data. On the labour-income side, the same goes in most cases for employers' social security contributions. Another part of the discrepancy between data sets is caused by measurement errors, which correspond mainly to biases caused by sampling procedures, non-random misreporting and heterogeneous response rates, all of which likely affect the top tail of the income distribution disproportionately (Blanchet, Flores, and Morgan 2018). Törmälehto 2011; Endeweld and Alkemade 2014 expose these gaps by comparing similar income concepts in harmonised surveys from the Luxembourg Income Studies Database to those of national accounts. Both studies find that "other factor income", which includes both capital and self-employment income, are generally more underestimated than labour incomes, which are considered to be relatively well represented in surveys. They both also point out that using register data in the surveying process seems to narrow these gaps.

The main contribution of this paper is to establish stylized facts on data gaps and to assess their impact on the measurement of inequality. Relevant variables are obtained by building a simple theoretical framework, which traces the path of capital and labour incomes from the national accounting level to the distributive context. The main finding is that these gaps likely result in the underestimation of inequality. They also seem to affect the sensitivity of inequality estimates to the dynamics of the capital share. Survey estimates largely understate the influence of capital income and thus seem to follow almost exclusively the distribution of labour income alone, which explains that Francese and Mulas-Granados 2015 find capital income to contribute marginally to increasing inequality, as measured in surveys. Tax data appears to be moderately more sensitive to the capital

share and its distribution, at least in the United States, during the period 1975-2015, which could help explain that Bengtsson and Waldenström 2018 find a positive impact of capital shares on top income shares estimated from tax data. Additionally, I find evidence to support the use of register data in the construction of household income surveys as a way to reduce inconsistencies.

I distinguish two types of data gap. First, as a way to indirectly account for the amount of national income that is out of the scope of distributive data, I measure the "household share of capital income" –as opposed to the share of private and public corporations. I find a generalised and strong decrease in the share received by the household sector, which implies that an increasing part of macroeconomic income is retained in corporations and thus is ignored by most distributive data. This trend is found by analysing 43 countries which have sufficient detail in the United Nations' official statistics on National Accounts. It holds when studying the aggregate income of 19 countries forming a balanced panel during the period 1995-2015. But also at the country level, during the same period, in an unbalanced panel including all countries with requisite data. The longest available series show that trends start falling around 1990 in most cases.

Second, I measure the relative underestimation of labour and capital incomes of the household sector in distributive data. In order to account mostly for measurement error, I use the flexibility of income definitions in household surveys to get the closest possible to those used in the SNA. I use the "factor income" definition of the Luxembourg income studies database, which is defined as gross of both taxes and contributions. However, some of the items that accrue to the household sector in the SNA are not included in surveys at all, such as investment income attributed to pension fund holders or other imputed incomes. Therefore, these "factor gaps" should be considered as a product of both measurement errors and conceptual discrepancies, which I attempt to disentangle for recent years. I find that both labour and capital income appear to be undervalued in a balanced panel of 13 countries for the period 1995-2013. Capital income is, in all cases, relatively more underestimated, with only around 20% of its gross figure being recorded in surveys, against 70% for labour income. This relation remains generally stable over

the period. For the United States we compare the level of coverage with that of tax data, which is moderately higher for capital income, with close to 30% captured, versus close to 80% of labour income.

Both data gaps are formalised in this paper and their estimates are used to study marginal contributions to change in income concentration. I analyse the evolution of groups such as the top 10%, the middle 40% and the bottom 50% of the distribution, at country level. I find that distributive estimates are rather insensitive to the dynamics of capital income concentration, while they are over-sensitive to those of labour income. An example are fiscal income declarations in the United States, which report a top decile gathering close to 50% of reported capital income in 1975 and 70% of it in 2010. Despite the sizeable increase in factor income concentration during the period, it only contributed with close to 1.3 points to the more than 14 points increase in the top decile's share of total income, the rest being mostly due to labour income dynamics, which also tended to more concentration in the top decile. The increase in the capital share of national income contributed positively but modestly, with close to 0.6 points to the aggregate variation. The evolution of measurement gaps slightly attenuated the increase of the top 10% share, having a negative contribution close to -0.7 during the same period. I also find that the effect of data gaps not only applies to top income groups but also –systematically– to all other groups, yet the sign of marginal effects varies depending on the group under study and its factor income composition.

This study is organised as follows: Section 1 discusses both theoretical and empirical literature on the link between capital shares and inequality. Section 2 defines stylized facts on the distribution of capital income across institutional sectors and on the measurement gaps of factor income in surveys. Section 3 presents the theoretical framework that is used to decompose the relation between capital shares and inequality measures. Section 4 displays empirical applications of the model, aiming to understand the composition of variations in inequality estimates. The last section discusses the main findings and concludes.

1 Literature on Capital Shares and Inequality

Theoretical work by Atkinson and Bourguignon 2000 Atkinson 2009 and Milanovic 2017 shows that, although the relation between capital shares and inequality is more complex than it seems, it is reasonable to expect a positive correlation in realistic scenarios. Since individuals receive income from different sources at the same time and recipients of each type of income can be found throughout the whole distribution, these studies use both the relative concentration of factor incomes and their joint distribution to describe the relation. Building on Atkinson and Bourguignon 2000, Atkinson 2009 defines the conditions for the capital share to have a positive impact on inequality.[1] The figure can take values above zero, meaning that at some levels an increasing capital share results in decreasing inequality, yet the critical point is expected to be rather low in plausible convex scenarios. Milanovic 2017 depicts a similar framework using the Gini coefficient. The author defines three clear requirements to have a positive relation between capital shares and inequality: first, high saving out of capital income; second, high concentration of assets; third, high correlation of capital income ranking and total income ranking. In real cases, all of these requirements are easily fulfilled.

Although the theoretical findings of this literature are based on straightforward accounting identities, its empirical assessment seems more opaque. Bengtsson and Waldenström 2018 use a panel of 21 countries to assess the statistical relation between our variables of interest. They build capital shares using historical national accounts data and then regress them to income-concentration estimates. They use two of them: Top income shares from the World Inequality Database (WID)[2], which are based on administrative

[1]The capital share (π) has to satisfy the following inequality: $\pi > (1 - \lambda\rho)/(1 - \lambda^2 - 2\lambda\rho)$, where λ is the ratio of the squared coefficient of variation of capital over that of labor income and ρ is the correlation between capital and labor incomes.

[2]The authors cite the database using its name at the time they were writing: the World Top Incomes Database (WTID).

records, and Gini coefficients, which they draw from Atkinson and Morelli 2012[3]. Their estimates, including country fixed-effects, are in line with what is expected. They find a strong positive marginal effect of the capital share over both inequality measures. A contribution with somehow contrasting results is Francese and Mulas-Granados 2015. The authors use a different data source: the Luxembourg Income Study Database, which provides harmonized household surveys (`http://lisdatacenter.org`). They perform a decomposition analysis of the Gini coefficient in 43 countries during the period 1978–2010. They break down the Gini coefficient to its accounting components and then implement a similar regression than Bengtsson and Waldenström 2018, but only using the survey's Gini coefficient as a dependent variable. After analysis they conclude that the capital share plays a negligible role in the evolution of measured inequality, especially relative to the evolution of labour-income inequality, which they judge as the main driver of total inequality.

The issue with existing models is that they only allow for a negative –or null– relation between capital shares and inequality under very restrictive circumstances. Therefore, when negative correlations emerge empirically these models do not seem to provide a convincing description of the mechanisms at play. For instance, in Milanovic 2017, the only channels trough which it can be achieved, is by releasing at least one his requirements, which results in rather unrealistic scenarios. Additionally, when Bengtsson and Waldenström 2018 observe a negative correlation for both Argentina and Canada, they treat it as an anomaly, since their model does not provide a meaningful explanation for those particular cases.

[3]Atkinson and Morelli 2012 estimate Gini coefficients either directly from popular local household surveys or from well-known international data centers. These estimates are available for a subset of countries and a shorter span in time compared to top shares

2 Stylized Facts

This section starts by describing the concepts and data sources that are used throughout the study. It then briefly displays the capital shares that are obtained from national accounts, before proceeding to the analysis of stylized facts on both the household's share of capital income and survey measurment gaps.

2.1 Data and Concepts

National Accounts Estimates in this section are mainly built using the United Nation's 'National Accounts Official Country Data', which is publicly available at: `http://data.un.org`. This dataset distributes the whole national income to different institutional sectors. Ideally these are six: the household sector, non-profit institutions, financial and non-financial corporations, the general government, and the rest of the world. As not all countries build their accounts equally, the level of aggregation among sectors varies. For the sake of clarity and comparability, the main estimates of this paper aggregate both financial and non-financial corporations in what is referred as 'private corporations'. Although the general government, or 'public sector', is partially studied in the present paper, the evolution of its share of capital income is mostly not commented as it has little economic relevance. This is because its capital income is mainly composed of the profit of publicly held firms and payment on the interest of public debt, both of which are only a part of total public revenue and expenses. Non-profit institutions are mostly ignored in the analysis, as they always receive a negligible share of capital income. Data on the foreign sector is only used to estimate national income, as opposed to domestic income.

The guidelines of the UN's official System of National Accounts (UN-SNA henceforth) have been re-edited five times since its first version in 1953. Every revision included substantial methodological modifications, which often render different series hardly comparable. For that reason, both the aggregate and country-level estimates that are presented here do not mix information from different UN-SNA series. The series that are included in the balanced panel of 19 countries all correspond to the latest existing UN-SNA, which is the 2008 version. However, for long-run analysis and for the inclusion of less developed

countries in an *unbalanced* panel, we also use series based on the 1993 UN-SNA guidelines.

Household Surveys I use survey micro-data from the Luxembourg Income Study Database (LIS), which contains detailed harmonized datasets from a group of countries, for two empirical applications. First, it is used to assess the measurement gaps of surveys in section 2.4. Second, it provides an empirical application of the theoretical model presented in section 4. I take advantage of the flexibility of income definitions in surveys to make them more comparable to both tax data and national factor incomes. The income definition that is used corresponds to the LIS variable named 'factor income' and the population that is considered for comparison are all adults aged twenty or older. In practice, the definition includes gross yearly income (pre-tax), combining monetary and in-kind revenues. However, table 1 maps survey incomes to national accounts concepts. It displays the main problematic items, some of which are never included in the survey's definition, as imputed rents, imputed incomes to insurance policy holders, reinvested income abroad and investment income on pension funds.

Tax data In the case of the United States, we analyze data for a wider time span, that is the period 1975-2015. Piketty and Saez 2003 and relevant updates are used as estimates of total capital and labour income declared to American tax authorities. Furthermore, DINA estimates from Piketty, Saez, and Zucman 2018 are also used to make empirical comparisons. These estimates combine both survey and tax statistics to distribute the whole national income to the personal income distribution.

Income Concepts It should be noted that all the empirical estimates of inequality that are provided in this article are limited to the study of pre-tax income. Although this is standard in the top incomes literature, it is not in the broader an more traditional literature on income inequality that generally focuses on disposable income, which is better suited to study households' economic well-being. Both concepts are relevant for different reasons, while pre-tax income is the direct result of market outcomes that are closely related to the structure of production functions, disposable income depends more on the

direct effect of redistributive public policy, which is the result of historical, political and institutional decisions. This study does not discuss the role of redistribution through taxes and transfers, which are considered to be out of its scope mainly due to constraints that are given by the nature of underlying data.

Sectoral Income In national accounts, each institutional sector ($i = 1, \ldots, n$) receives a capital income KI_i, which is defined as the sum of the sector's Operating Surplus (OS_i) and Net Property Income (NPI_i).[4] The only exception is the Household sector, which also receives Mixed Income (MI). A part MI^K of this aggregate is also assumed to remunerate capital. Naturally, the sum of the capital income of every sector KI_i is equal to total national capital income. We thus can define the total capital income of the economy KI in two ways:

$$KI = \sum_{i=1}^{n} KI_i = MI^K + \sum_{i=1}^{n} (OS_i + NPI_i) \tag{1}$$

In the following subsection we mainly study each sector's share of capital income (KI_i/KI). Furthermore, labour income is by definition only affected to natural persons, which are either national or foreign households. Since the share received by foreign households is always negligible, I chose to ignore it when presenting stylized facts and results, even though it is subtracted in calculations. Labour income can be simply defined as the addition of the Compensation to Employees (CE) and MI^L, the share of Mixed Income remunerating labour.

2.2 Capital Income and Capital Shares

Capital Income Globally, the literature on factor shares defines capital income as the sum of the total Operating Surplus and Net Property Income, plus the share of Mixed Income that is assumed to remunerate capital. While the first two terms of the addition are universally accepted, the latter is often considered to be problematic. Indeed, Mixed Income broadly corresponds to the income of the self-employed, who usually combine both

[4] In the definition of Net Property Income, the word 'Net' refers to: income received less income paid.

labour and capital to produce goods and services. However, the partition of this aggregate between factor incomes always relies on a priori assumptions. The literature has developed basically three ways to deal with this issue.

The first method –which is the most data intensive– is to estimate wages of workers with given characteristics using surveys, to then assume that independent workers with similar characteristics pay themselves similar wages. In that situation, the capital share of their income is estimated as a residual. A second approach is to estimate the capital share of the private sector, which is the ratio of its Operating Surplus over total Value Added. This procedure depends largely on the level of detail in the institutional sector accounts, which is far from being homogeneous across countries. The third and simplest approach is the one that consists in assuming that a fixed share of mixed income, which is close to 2/3 or 70% remunerates capital. The first method is used in ILO 2019, while both the second and third approaches are used by Bengtsson and Waldenström 2018 and Piketty and Zucman 2014. In these studies, the second approach is used for countries and years where data is sufficiently detailed (i.e., relatively recent years), while the third method is only employed for historical scenarios with limited data availability. Since the main goal of the present paper is not to establish precise levels of capital shares, I use exclusively the third method to both maximise eligible countries and to ensure transparency in the estimates.[5]

Capital Shares Figure 1 displays both the gross and net estimate of national capital income shares in a balanced panel of 19 countries. While the gross estimate increases around 2 percentage points through the period, remaining near 40%, the net figure is lower and stable around 25%. At Country level, trends are relatively more dynamic. Gross

[5]This strategy assumes implicitly that independent workers' remuneration has the same composition in both developing and developed countries, which is most probably incorrect. Indeed, we could expect the produce of independent workers in developing countries to take a bigger share of macroeconomic income and to be more labour intensive than their developed counterparts. However, developing countries are mostly not included in the analysis and their estimates are not subject to international comparisons but are rather studied as time series.

capital shares range between 45%-30%[6] and net shares are generally between 20-30% (figure C.1).

As a way to show that the 70% assumption is not a strong one, especially for the developed countries composing the balanced panel, appendix A compares own estimates to alternatives found in the literature. In most cases, estimated gross capital shares are lower than those produced by ILO 2019, yet they both closely follow the same trends (figure A.1). Their method is probably assuming a more capital intensive production than the fixed 30% assumption that is made in this paper. Bengtsson and Waldenström 2018 estimates cover less countries than those of ILO 2019; they are generally lower than those used in this study and they also follow similar trends, except in the case of Denmark. Estimates by Piketty and Zucman 2014 are less comparable to those used in this paper since they are all net of capital depreciation (figure A.2). However, for countries with requisite data, our own net estimates are lower than theirs yet relatively close in trend. Although the use of alternative estimates would not change main results in this paper, it would reduce the number of countries with requisite data.

Fixed Capital Consumption The main estimates from national accounts presented in this paper are gross of fixed capital consumption. Since capital depreciation has increased as a share of GDP in most countries during the last decades (figure B.1), it is important to check whether the broad conclusions of this paper hold when using net definitions. Appendix B shows that they generally do. However, as the decomposition of capital depreciation by institutional sector is quite uncommon, the number of countries with sufficient data is substantially reduced if we only focus on net estimates.

2.3 The Household Share of Capital Income

This section provides evidence on the generalised decrease of the household share of capital income. I start by presenting the distribution of capital income among institutional sectors, in a balanced panel of 19 developed countries during the period 1995-2015. The period

[6]The only exception is the extreme case of Norway, which records a near 50% gross capital share

was selected in such a way that maximises the quantity of countries in the panel. Indeed, it is from 1995 that most countries provide detailed-enough series of national accounts. An unbalanced panel with 43 countries is also available in the same period. We therefore also analyse its evolution at the country level. In order to explore the historical dimension of the observed phenomenon, we end the section with the study of the 9 countries which report data before 1990.

Balanced Panel Figure 2 shows that both the Public Sector and Private Corporations increase their share of gross national capital income between 1995 and 2015. The Public Sector starts with a negative value near -2.8% and ends the period with a low but positive share of 5%. This finding does not have high economic relevance in itself as it does not take into account the full income or expenses of the Government. It only accounts for its Operating Surplus and Net Property Income, which are mostly composed of the profit of publicly held companies and the payment of interest on public debt, respectively.[7] From an accounting point of view, however, it can be interesting to understand that this phenomenon has an impact on the relationship between what we define as the capital share and measured inequality. This idea is further developed in section 3.

More relevant in figure 2 is the trend described by Private Corporations, which actually shows that retained profits represent an increasing part of capital income through the 20 years with data. That is, a bigger part of private profits are held inside corporations instead of being distributed to natural persons. Subfigure 2a displays a modest increase of their share by near 3 percentage points. Yet, economically, it makes more sense to study the evolution of this sector by excluding the government, as displayed in subfigure 2b. Indeed, value added is generated in corporations. After paying taxes, it is either distributed to natural persons in the form of wages (e.g. compensation of employees) or as distributed profits (e.g. dividends), the rest is retained inside private corporations.

Figure 2b describes a clear and relatively constant decrease close to 7 percentage points

[7]Figure C.2 shows that the trend is mainly driven by the reduction of the expenses related to negative Net Property Income. Again, without taking income from taxes into account.

in the household share of gross capital income. It is only interrupted by an ephemeral increase between 2007 and 2009, which is likely to be driven by corporate losses during the financial crisis. This trend corresponds to the aggregate capital income produced in 19 developed countries which form the balanced panel for the period. Individual countries generally follow the same trends, which are displayed in figure C.3. The weight of each country on this trend is in fact rather unequally distributed, as the first 5 contributors account for near 70% of total capital income in the panel (Table C.1). However, the same general conclusions can be achieved using other income definitions, for instance: net estimates from a sub-panel of 12 countries with available data (figures B.2 and B.3). Additionally, figures B.4 and C.4 display un-weighted averages of net and gross estimates, respectively.

Unbalanced Panel The dynamics displayed in figure 2 are not exclusive to developed countries. In total, 46 countries from several continents report detailed-enough data for this period. They do not cover, however, all years in the period. We thus study in figure 3 the evolution between the first and the last year recorded by each of these countries, including those in the balanced panel.

The red countries below the bisector line in figure 3 represent close to 70% of countries which saw a decrease in the household share of gross capital income. Most countries that are present in the balanced panel appear in this side of the plot. Additionally, from the developing world, we can find several Latin American countries (e.g. Chile, Colombia, Guatemala, Mexico, Nicaragua and Peru), and an Eastern Asian one: Japan. We can also find some additional countries from Eastern Europe (e.g. Poland and Lithuania) and Southern Europe as well (e.g. Greece, Spain). The blue dots that are situated above the bisector line gather mostly at the bottom-left corner of the figure. These countries saw an increase in the household share of capital income, yet their relative position shows that they already had low levels to start with, at the beginning of the period. The country that saw the biggest increase is Netherlands Antilles, which gained around 10 percentage points during the period. This case should be noted as a special one because the country is a tax haven (Zucman 2015). When the public sector is excluded, we get a rather similar picture

(figure B.5). For estimates of net capital shares, due to increasing capital depreciation, a relatively lower majority of countries follows a decreasing trend (figure B.6).

Long-run series Some countries have enough data to estimate the household share of capital income for several decades with consistent UN-SNA definitions. We display their long-run estimates in figure 4 as a way to study the starting point of the decreasing trend observed in figures 2 and 3. Italy is the first country to start a clear decreasing trend in the early 1980s. Others, as the United States, Canada, Australia and Japan start a rather clear decreasing trend around 1990. The countries with the latest starting trend are Netherlands, at the end of the 1990s and Norway, which experiences a big drop after its fiscal reform of 2005 that introduced the permanent taxation of distributed profits (Atkinson and Aaberge 2010). France and Finland are cases where we do not observe particularly strong or sustained trends, but rather ephemeral dynamics. For instance, France experienced a drop in the household share of capital income at the end of the 1980s. This drop was then counterbalanced during the following years. Finland experienced an ephemeral jump of the estimate at the beginning of the 1980's, which was likely provoked by corporate losses during the Finnish banking crisis of 1991-1993. Excluding the public sector and/or using net estimates does not seem to affect general conclusions (figures B.7, B.8, and B.9).

The stylized decreasing trend that is observed in this subsection probably has an impact on standard measurements of inequality, since common distributive statistics (i.e. tax data and household surveys) deliberately only record the income of natural persons. They thus ignore income retained in corporations, which is not the case for factor income shares.

2.4 Gaps in Distributive Data

Standard distributive statistics generally understate the aggregate income of households compared to national accounts. In what follows, we study measurement gaps between these estimates, taking national accounts as a benchmark. Figure 5 provides empirical estimates based on both surveys and tax statistics. Harmonized household surveys are

available for a balanced panel of 13 countries with data in both the LIS database and UN's national accounts data.[8] Data from tax declarations is originally provided by the American Internal Revenue Services, which is used in Piketty and Saez 2003 and updated by the authors for more recent years.

Survey Data Figure 5a shows that both labour and capital income are underestimated in surveys relative to national accounts. The level of the underestimation, however, is not equal for both factor incomes. While labour incomes appear to be relatively well represented with near 70% of it being recorded in surveys, the figure remains around 20% for gross capital income during the period. Between 1995 and 2013, the evolution of these estimates is rather stable. At the country level, the relative underestimation also holds in every case, yet there is some variation in levels and trends (figure D.1). Table 2 reports countries using administrative records in the construction of surveys, which seem to achieve a better coverage of both factor incomes. The Netherlands is a special case, since tax records are introduced in the survey-wave of 2004 surveys, unlike the rest of countries in the table, which use register data for the whole period. In 2004, the coverage of capital incomes increases substantially in the Netherlands' survey –from close to 20% in 2000 to 40% in 2004 and 80% for the next observations. This finding holds in the long term and in both net and gross figures (figures D.1, D.2 and D.3). It should also be noted that not all countries using register data employ them equally. For instance, in the case of Austria, they are mostly used to complete variables for labour income and not necessarily for capital incomes, which explains its higher coverage for labour income but not for capital incomes. In line with Törmälehto 2011 and Endeweld and Alkemade 2014, findings in this subsection provide supporting evidence for the use of adimistrative data as a way to narrow data gaps between survey estimates and national accounts' figures.

[8]This is a sub-panel of the one presented in figure 2. It includes the following countries: Austria, Canada, Czech Republic, Denmark, Finland, Germany, Great Britain, Greece, Hungary, Italy, Netherlands, Poland and Spain.

Tax data In the case of tax data it is more difficult to find estimates for total factor income. Many studies directly use totals from national accounts to estimate top income shares (Atkinson, Piketty, and Saez 2011). And those basing their aggregate estimates on fiscal data, usually do not report its composition in terms of factor income. In the case of the United States, most of the adult population fills tax declarations as only around 20% of them do not declare income to the Internal Revenue Service. Piketty and Saez 2003 provides estimates on the decomposition of the aggregate income they use. Figure 5b shows that, although capital income remains relatively more underestimated than labour income, the gap is narrower than in survey data for most countries. Indeed, figure D.4 shows that this is also true when we compare tax and survey estimates for the US alone. It is worth noticing that both survey and tax estimates for the United States cover more than three decades. During the 1975-2011 period, survey data gets progressively worse at capturing capital income as the share it captures goes from around 30% to 20% during the period.

Decomposing Measurement Gaps Ideally, one would always distinguish between measurement error and conceptual differences to evaluate data gaps and their evolution, but it is not possible to do so using the data behind figure 5a, due to lack of detail. In order to test the main conclusions of this subsection, I use harmonised data from OECD National Accounts to match income items between surveys and national accounts. This database has the virtue of describing aggregates with more detail, yet it is available for a shorter period –starting only in 2008. Figure E.1 compares items that can be systematically matched. It shows that dividends and interests –which are part of capital incomes– are generally more underestimated than the main component of labour incomes: wages. The only exception is Denmark, where wages are covered to almost 100%, while dividends and interests are overestimated close to 20%. Countries using register data seem to be relatively better at capturing both capital and labour incomes –see table 2– yet some heterogeneity remains, likely due to the different uses that each country gives to such data before, during and after the surveying process.

Figure E.2 compares rents, which are a part of national capital income and are perceived

as problematic since they can only be partially matched using the SNA's institutional sector approach. It compares surveys' imputed and realised rents to gross operating surplus, which includes both in the same aggregate. In this case, the accuracy of estimates varies greatly across countries, with cases of over- and under-estimation. This finding is somehow expected, since imputed rents largely depend on home-ownership rates and usually accrue to substantial amounts relative to total household income, yet methodological issues render it difficult to measure.[9]

Based on table 1, figures E.1 to E.14 provide a decomposition of the household sector's income, highlighting conflicted items, which are included in the SNA but systematically excluded from distributional data. On the capital income side, we find imputed incomes on foreign investment, on pension funds and those imputed to insurance or policyholders. On the labour income side, the aggregate is basically employers' social security contributions. These figures show that in a large majority of cases, a bigger share of labour incomes is systematically excluded.

The most relevant stylized fact in terms of measurement gaps appears to be that capital income is probably more underestimated than labour income in standard inequality databases, again due to both measurement error and conceptual differences. From an intuitive perspective, we should expect this to have similar consequences to the estimates presented in subsection 2.3, since in both cases we are measuring a part of capital income that is being ignored by distributive data. The following section defines a simple theoretical framework that should help understand the implications of these stylized facts on the measure of income concentration and its relation with the capital share of national income.

[9]Although imputed rents are considered as a variable in the LIS data set, it is not included in the capital income definition used in the main analysis, since it is only available for a limited set of countries and time-coverage is patchy (see table 3).

3 Theoretical Framework

The aim of this investigation is to understand the nuances between variations in the capital share and the total income share of each quantile in the distribution –e.g., the richest 10%, the middle 40% or the bottom 50%. It should be noted that the model presented in the following paragraphs does not claim to describe causality but rather is an accounting framework that sheds light on the structure of estimates and their dynamic behaviour.

3.1 Setup

To describe the theoretical framework behind this study, we will consider the following setting. Let K and L be two non-negative real random variables whose sum is equal to 1. We will use K to represent the capital share of national income in a given economy. And we will use L to represent its counterpart: the labour share. Both variables are recorded in National Accounts, which divides income by institutional sector $i = (1, \ldots, n)$. labour income belong integrally to the household sector (h), while capital income is divided in different institutional sectors, which receive a share of total capital income $\Phi_i = (\Phi_1, \ldots, \Phi_h, \ldots, \Phi_n)$, so that $\sum_{i=1}^{n} \Phi_i = 1$. In the following subsections we will focus on the relation between K and common inequality estimates. These estimates are in practice recorded by distributive statistics which are either survey or tax data. Both data sources use a narrower definition of income than National Accounts and are also subject to mismeasurement. In consequence, we define ϵ_K and ϵ_L as two real numbers that are higher than 0. They represent, respectively, the share of national accounts' capital and labour income that is present in tax or survey data. We therefore define H, the total household income in the distributive data:

$$H = K\Phi_h\epsilon_K + L\epsilon_L \tag{2}$$

3.2 Identities

Income Shares We divide the total population in quantiles $q = (1, \ldots, m)$, such that the share of household's income received by each quantile is $S_q = (S_1, \ldots, S_m)$. In the same line, each quantile receives a share of household's capital income $S_q^K = (S_1^K, \ldots, S_m^K)$ and a share of total labour income $S_q^L = (S_1^L, \ldots, S_m^L)$. We then define the share of households' income received by each quantile q as follows:

$$S_q = \frac{K\Phi_h\epsilon_K S_q^K + L\epsilon_L S_q^L}{H} \tag{3}$$

In this expression, the share is equal to the sum of both capital and labour income held by quantile q divided by total household income (H), which is defined in equation 2. In other words, quantile q receives a given percentage S_q^K of the total capital income recorded in household data, $K\Phi_h\epsilon_K$, plus a share S_q^L of total labour income recorded in the same data, $L\epsilon_L$. Now, the same expression can be rearranged in a less intuitive yet useful way:

$$S_q = \frac{S_q^K K\gamma + LS_q^L}{K\gamma + L} \qquad \text{s.t.} \qquad \gamma = \Phi_h \times \frac{\epsilon_K}{\epsilon_L} \tag{4}$$

Equation 4 can be translated graphically into figure 6, which depicts S_q in the empirically relevant case where quantile q concentrates a relatively higher share of total capital income ($S_q^K > S_q^L$). The function is defined for all possible values of K, keeping other variables as fixed parameters, in 3 different scenarios.

The black straight line represents the situation were the household sector receives all the capital income and both capital and labour income are estimated with the same error in the distributive dataset ($\gamma = 1$). The red convex line illustrates the most realistic case, where public and private corporations have positive income and/or aggregate capital income is relatively more underestimated than labour income ($\gamma < 1$). Conversely, the blue concave line describes the situation where private and public corporations have capital losses instead of income and/or a bigger part of capital capital income is recorded in the distributive data, compared to labour income ($\gamma > 1$).

Figure 6 shows that the γ parameter defines the linearity, concavity or convexity of the

function. While the factor income concentration variables (S_q^K and S_q^L) define the sign of the slope and both the upper and lower boundaries for each quantile's income share. The construction of the γ parameter reveals that Φ_h and ϵ_K / ϵ_L influence income shares in the same way and that they multiply each other. Indeed, both components of the γ parameter operate as filtering-out a part of the capital income from the equation. Furthermore, the γ variable has both a direct and indirect impact on quantile shares. That is, a lower γ not only results in lower S_q for a given K, but it also has an impact on the marginal effect of K over S_q.[10]

3.3 Marginal effects

Figure 6 gives relevant insight on the sensitivity of income shares to changes in the capital share. In the simpler case (black straight line) a variation ΔK always engenders the same variation ΔS_q that is proportional to the slope of the curve. However, in the convex and concave cases, the marginal effect of ΔK varies with K. In order to better understand the sensitivity of S_q to every parameter in equation 4, Table 4 displays the partial derivatives of the model in the cases with and without differences in the concept of capital income across datasets (Columns 2 and 1 respectively).

In column 1 the marginal effect of the capital share is constant: it is equal to $S_q^K - S_q^L$. Similarly, factor income concentration variables S_q^K and S_q^L also have constant marginal effects, which are equal to the value of national factor shares, K and L respectively. When differences are introduced –in column 2– one can see that marginal effects are equal to those of column 1, but multiplied by a given factor. Therefore, some of the effects will be undermined, while others will be exacerbated. In the case of both the capital share –K– and the concentration of capital income –S_q^K– in realistic scenarios, the marginal impact will be lower in column 2 compared to the corresponding value of column 1. That

[10]It is worth noticing that the relations described by figure 6 and equation 4 are based on an underlying assumption whereby individual income rankings are kept unchanged after variations in the capital share. This can be a rather strong assumption, yet if the analysis is restricted to infinitesimal variations, it should not be a problem.

is because $\gamma(K\gamma + L)^{-2}$ and $\gamma(K\gamma + L)^{-1}$ will both take values between 0 and 1. On the contrary, the impact of variations in the concentration of labour income S_q^L will be exacerbated, as it will be multiplied by $(K\gamma + L)^{-1}$, which should take values higher than 1. The marginal effect of γ is only relevant when it is different from 1. It is thus defined only in column 2, yet its interpretation is relatively less intuitive. It is worth noticing that the effect of variations in coverage rates may be underestimated if some of the income items are concentrated in a given part of the distribution.

The study of partial derivatives implies that in normal cases, we should expect the role of capital-income-related variables to be undermined and those related to labour income should be exaggerated with respect to the ideal scenario. In the following section, these derivatives will be calculated empirically for various data-points in different databases in order to study the structural drivers of motions in top income shares.

4 Applications

This section exploits the theoretical framework described in section 3 to produce empirical estimates based on the data behind the stylized facts in section 2.4. The analysis gives further insight on the driving forces of top income shares as estimated by surveys from 14 different countries in the period 1995-2013 but also from tax data and DINA estimates for the United States in the period 1975-2011. In this section, I focus on presenting results for the top 10% share. Other groups, such as the top 1% share, the middle 40% and the bottom 50% of the distribution were also analysed and corresponding estimates are available in appendix F.

4.1 General Trends

Surveys: Balanced Panel (1995-2013) Figure 7 depicts the evolution of the relevant variables in the model, using the same balanced panel presented in section 2.4 as if it was a single country. Both the capital share and the top 10% share of total income grow

during the period –figures 7a and 7b, respectively.[11] The former gains a bit less than 2 percentage points, while the latter increases close to 4 points. As is to be expected, the top 10% income share is always between the levels of labour and capital income concentration. What appears rather strikingly is that the concentration of total income follows the level of labour-income concentration extremely closely. This is also the case when analysing countries individually (figure D.5). It seems that the level of accumulation in the top 10% share is relatively insensitive to motions in capital income concentration, as it appears to depend mostly on what happens in the labour income distribution.

As is explained further, this phenomenon is related to the low values of the γ coefficient in most surveys. Through the whole period, the coefficient remains below 20% and even decreases overall. When countries are studied independently, it never takes values above 30% and also decreases in most cases – see figure D.6. The decomposition of the γ coefficient at the country level in figure D.7 shows that, although it is common to find that the household share of capital income Φ_h is higher than the mismeasurement ratio ϵ_K/ϵ_L, this is not always the case as there are substantial differences in levels and trends of both estimates across countries. Furthermore, the γ coefficient can be interpreted as the part of national gross capital income that is taken into account by the distributive data. Under these circumstances, although the capital share and the top 10% income share appear to be positively correlated, we should expect a rather low marginal effect of the former on the latter.

Comparing Datasets: United States (1975-2015) The United States is analysed separately for two reasons. First, because its aggregate income is approximately equivalent to that of all the countries in the panel put together. Thus, if we were to include it in the panel, it would monopolise trends. Second, the US is one of the few countries which have good quality data for surveys, tax data and Distributional National Accounts (DINA) at

[11] Quantiles are built ranking all households by increasing non-equalised factor income for each country separately, using survey weights in case of sample data

the same time.[12] This enables a limited but useful comparison of estimates coming from different databases and also the study of sensitivity to changes in the capital share.

Figure 8 displays estimates of income concentration in the three different datasets for the period 1975-2011. In all cases, the top 10% income share increases substantially through the period. Surveys record a near 9 points increase starting with a 27% share, while the tax data estimate increments around 15 points in the same period and starts higher, at near 32%. The DINA estimate starts at the highest level, near 34%, and shows a similar increase of about 12 points during the period. As in figure 7b, total income concentration is closer to the estimate of labour income than to the one for capital income in all databases. The distance between the red and the gray lines is different in every case. Despite reporting different levels of factor income concentration, all data sets show an increasing concentration for both factor incomes in the top 10% share.

Again, the difference in the γ coefficient, which is the product of the two estimates studied in sections 2.3 and 2.4, is likely to be crucial to understand differences in measurement. As can be seen in figure 9, the gamma coefficient is higher in tax data compared to the estimate from surveys.

4.2 Sensitivity Analysis

Now that all the variables introduced in section 3 are defined for case studies, we can produce empirical estimates of the partial derivatives appearing in table 4. Table 5, presents the average value of estimated marginal effects for each variable in the model, for every country and data source.

[12]Survey data for the US is also derived from the LIS Database. What is referred as tax data are the estimates of Piketty and Saez 2003 and subsequent updates that were made by the authors. DINA estimates come from Piketty, Saez, and Zucman 2018. They correspond to a global project that aims to combine surveys, tax data and national accounts to better study the distribution of the whole national income (Alvaredo et al. 2016).

Surveys: Balanced Panel (1995-2013) In table 5, the highest marginal effect is that of the concentration of labour income (column 2) for all the countries forming the Panel. In the aggregate scenario, an isolated increase of 1 percentage point in this variable, we should expect a systematic increase of 0.91 points in the top 10% share, which is rather close to perfect correlation. This is a high value compared to the marginal effect that we would observe if there was not any difference across distributive data sets and national accounts (if $\gamma = 1$). That is, the marginal effect would be equal to the benchmark labour share estimated using national accounts, which is close to 62% for most of the period (see figure D.6 for country estimates). Instead, here the marginal effect is equal to the labour share that is estimated by the distributive data set. Indeed, accounting identities ensure that the estimate from column 2 is actually equivalent to the labour share measured by surveys, while column 3 is equivalent to its counterpart: the capital share. The marginal effect of factor income concentration is not exclusive to top income shares, it is exactly the same for the middle 40%, the bottom 50% or the top 1% shares (see tables F.1, F.2 and F.3). The mismeasurement of the capital share by surveys therefore has an impact in measured inequality as a whole, by exacerbating the role of labour income everywhere in the distribution.

Table 5 shows that the marginal effect of variations in the capital share of national income (column 1) appears to be rather weak, as its aggregated effect in the panel is only 0.04. In the ideal scenario, this estimate is equal to the difference of concentration in factor incomes. That is, a value near 10% for most years (figure 7b), which is likely to be underestimated in surveys. Furthermore, the concentration of capital income and the gamma coefficient appear to have a relatively low effect on the survey's top 10% share as well. An isolated variation of 1 point in the former variable (column 3) is translated into only a 0.09 point increase of the top share in the aggregate scenario, while the figure for the latter variable (column 4) is 0.06. The hierarchy of estimates is the same for all countries.

In the case of other groups than the top decile, the effect of variables such as the capital share or components of γ depends on the composition of each group's income.

For instance, table F.1 shows that an increase in the capital share would have a negative impact on the group's total income share in every country. In the case of the bottom 50%, somehow surprisingly, table F.2 shows that variations in the capital share should have a positive impact on the group's share of total income, yet the effect is smaller than for the top decile. It should be noted that most capital income accruing to the bottom of the distribution is the result of private pensions being added to the factor income distribution and a higher prevalence of self-employment.[13]

Comparing Datasets: United States (1975-2015) In the case of the United States we can compare the same estimates in different data sets. All the comments made in the previous paragraphs on survey estimates also apply to US surveys. Table 5 shows that the use of tax data somehow alleviates the exacerbation of the effect of labour income concentration, as the average marginal effect (0.9) gets modestly closer than the survey estimate to the actual value of the labour share, which stays between 65% and 70% through the period (figure 9). In the same line, the estimate of capital income concentration has a higher marginal effect (0.1) relative to the one from surveys (0.08). However, although the effect of variations in the capital share is more than double in tax data compared to the survey estimate, it remains low, at 0.07. This is not the case for the DINA estimate, which exhibits a marginal effect of 0.38. Of course, this is due to the fact that DINA estimates distribute all the national income to the personal income distribution. This corresponds to the situation where there is no difference in the income definition used to estimate capital shares and inequality estimates.[14]

[13]In 2019, the Luxembourg Income Study database revised its guidelines and stopped adding private pensions to capital income, yet estimates presented here were obtained before that modification.

[14]The capital share used in Piketty, Saez, and Zucman 2018 and therefore in the DINA estimates in table 5, is slightly different from the one displayed in figure 9 because it corresponds to the authors' personal factor income definition of national income.

4.3 Estimated Contributions

We can estimate the marginal effects studied in the previous subsection in every country at every data point. This allows us to compute each variable's contribution to change in income shares from one period to the other. If we multiply each variable's yearly variation by its marginal effect, we can analyze, from the perspective of accounting identities, the structure of changes in estimated top shares. Table 6 provides estimates for both the balanced panel of survey data and the United States with its different data sources. Column 7 aggregates the total estimated contribution of variables in the model. The difference between the estimated variation of the top share and the real variation (column 8) is due exclusively to the fact that databases only report subsequent snapshots at given points in time. In fact, most of the countries in the balanced panel only report data every 2-3 years, whereas tax data is available on a yearly basis. If we had access to the continuous evolution of these variables, there would not be any error in the estimate. Again, this is because the model is based on accounting identities. In any case, when plotted together, the 'estimated' top share and the real one are indistinguishable at normal scale.

Surveys: Balanced Panel (1995-2013) In surveys, the capital share does not appear to be a relevant driver of trends described by the top 10% income share. In the balanced panel taken as a whole, the top 10% income share increases around 3.4 percentage points during the period. But the parallel 2 points increase in the capital share only explains 0.08 points of such variation (column 1 in table 6). In fact, this contribution is completely counterbalanced by the measured the variation of the gamma γ coefficient (figure 7a), which has a negative and modest influence of -0.15 points. It is the concentration of labour income that gets the lion's share of contributions (column 2). These conclusions also apply, in general, to the country-level analysis. In surveys, the capital share and the concentration of capital income does not play a significant role in defining total income concentration in surveys. This is most likely explained by the large underestimation of capital income we observe by analyzing the trends and levels of the γ coefficient (figures 7a and D.6).

Individual Countries As an illustration, consider the case of Poland in table 6. The country experienced the biggest drop of the top 10% income share in the panel –as recorded by the survey– from close to 40% in 1995 to close to 35% in 2013. Despite an increase of 4%-5% of national income for the capital share, it only contributes marginally with 0.14 point to the evolution of the top decile, being completely offset by a considerable shift in the institutional allocation of capital income, which contributed with a negative -0.47 (see appendix G). Both components of the gamma coefficient explain the low transmission of the capital share to distributive measures, which also affects other income groups (see tables F.4 and F.5). Another example would be that of Canada, one of the countries that uses register data to construct household surveys for the whole period. It is also one of the countries that experienced a big increase in both its capital share and top 10% share. Appendix G shows that its ϵ_K/ϵ_L ratio is higher than most other countries, and even increases over the period due to an improvement in the coverage of capital incomes. In this case the increasing concentration of capital income seems to play a bigger role, explaining 0.73 out of 3.75 points in the top decile's variation. However, the impact of motions in the national capital share remains limited –contributing with only 0.06 points– likely due to a major drop in the Φ_h coefficient.

Comparing Datasets: United States (1975-2015) In US surveys, the conclusions are basically the same than with panel data, but with different levels, due to the larger extent of the period under study. The concentration of labour income also explains the largest part of variations in the local top 10% share, with an influence of near 9 points. The second largest contribution is a positive 0.38 points that is provoked by the increasing trend described by the capital share of national income. Measurement gaps have a marginal influence, contributing with negative -0.27 points, which are mainly due to the evolution of the ϵ_K/ϵ_L ratio. It indeed appears that the low level of γ has a bigger influence than its trend; this, by distorting the marginal effect of other variables.

When we analyse contributions using tax data, both the capital share and the concentration of capital income seem to be of higher relevance compared to figures in surveys, since together they contribute with close to 2 points to the increase of income concentration.

However, their aggregate contribution remains modest, adding to less than 15% of the total variation estimated by the model. This is probably due to the fact that even though tax data is better at capturing capital income income than surveys, the γ coefficient associated with tax data oscillates between 35% and 18% during the period (figure 9). That is to say, tax data still ignores around two thirds of the capital income produced by the country. It is only with DINA estimates that the capital share starts to play a substantial role in the evolution of the top 10% share. It explains near 1.8 points alone in the total increase of 14.4 points.

Discussion and Conclusion

This paper aims to reconcile national macroeconomic aggregates with distributive micro-data. To do so, the theoretical framework presented in section 3 assumes that the whole national income –including income retained in corporations– can and should be distributed to the household sector. This idea –however– is not consensual (Zwijnenburg 2019). It somehow corresponds to the view adopted by the Distributional National Accounts project (Alvaredo et al. 2016), which contrasts with that of Zwijnenburg et al. 2017 and the OECD-DNA Experts group, who aim to reconcile estimates exclusively for the household sector, following recommendations by Canberra Group 2001 and Stiglitz, Sen, and Fitoussi 2009.

The attribution of non-household incomes to the household sector makes sense for two reasons. First, corporations –or legal persons– are imaginary entities that serve natural persons owning and managing them. Therefore, following the Haig-Simons definition, which basically defines income as consumption plus changes in net wealth, the income of corporations can be considered as an increase in net worth for owners if retained earnings account for accrued capital gains (Gutiérrez, López, and Figueroa 2015). Ideally one would also add other items that generally qualify as increases in net wealth, such as changes in home equity, yet this is not done in this paper due to data constraints. Second, on a broader perspective, GDP is certainly the most widely used economic statistic and one of the most valued as well, it is even loosely used as a measure of economic progress.

Hence, independently on the current data conventions and the data limitations facing us today, it is worth at least trying to understand how such statistic is distributed across the population, and thus how economic progress is distributed to those that benefit from it.

If one were to seek macro-micro consistency exclusively for the household sector, it would translate into replacing the national capital share -K- in the model by -K_H- the capital share of the household sector in national accounts. The Φ_h variable, which accounts for the income retained in corporations, would then be absent from the model and the whole analysis would be centred around the same relative measurement gaps ϵ_K/ϵ_L. Findings in section 2.4, upon which capital incomes are systematically more underestimated in distributive data than labour income –resulting in lower transmission– would therefore remain unchanged.

Stylized facts show that the household share of gross capital income decreases in most countries reporting data during the last two decades. This finding is consistent with the study of Chen, Karabarbounis, and Neiman 2017, who find that global saving has shifted from the household to the corporate sector, without changes in the structure of investment, during the last decades. The authors cite real interest rates, the price of investment, and corporate income taxes as explanatory variables. The present paper –however– does not investigate further on the causes of such trend. For future research on this topic, a possibly relevant clue could be the generalised growth of share buybacks as a way to remunerate shareholders (as opposed to dividends). The financial literature thoroughly documents its explosive growth in the United States since the 1970's, while comparing dividends and buybacks in terms of tax efficiency and signaling, among other aspects (Bagwell and Shoven 1989; Fama and French 2001; Skinner 2008). The size of this phenomenon is remarkable, indeed, Floyd, Li, and Skinner (2015: fig. 3) show that the total amount allocated to share repurchases surpassed in the 2000's that of dividends in the United States. In the case of the European Union, Eije and Megginson 2008 also documents an increasing trend, yet with proportionally lower levels, for 15 countries since the 1990's. The relevance of this topic comes from the fact that the capital gains produced by these operations, whether they are realized or not, are generally not recorded by distributive

data. Such shifting in remuneration mechanisms would thus diminish the part of capital income that could be potentially recorded in distributive data.

The general message of this paper is that survey statistics does not capture a growing part of national income that remunerates capital. This likely understates inequality levels and trends. The size of this phenomenon renders most survey estimates almost completely insensitive to the motion of macroeconomic capital shares. Heterogeneity in income-coverage rates points to the use of register data in the survey process as a way to enforce coverage rates and thus to improve macro-micro consistency. In the same vein, transparency on the different methodologies used to estimate imputed rents, seems to be a crucial step in achieving statistical consistency. Furthermore, in the light of the evidence presented here, we can also better understand the diverging findings of Francese and Mulas-Granados 2015 and Bengtsson and Waldenström 2018.

References

Alvaredo, Facundo et al. (2016). "Distributional National Accounts (DINA) Guidelines : Concepts and Methods used in WID.world". URL: http://gabriel-zucman.eu/files/AACPSZ2017.pdf.

Alvaredo, Facundo et al. (2018). *World Inequality Report 2018*. Belknap Press of Harvard University Press.

Arpaia, Alfonso, Esther Pérez, and Karl Pichelmann (2009). "Understanding labour income share dynamics in Europe". URL: https://mpra.ub.uni-muenchen.de/15649/1/MPRA_paper_15649.pdf.

Atkinson, Anthony B. (Mar. 2009). "First". In: *Oxford Review of Economic Policy* 25.1, pp. 3–16. DOI: 10.1093/oxrep/grp007. URL: https://academic.oup.com/oxrep/article-lookup/doi/10.1093/oxrep/grp007.

Atkinson, Anthony B. and Rolf Aaberge (2010). "Top Incomes in Norway". In: *Top incomes : a global perspective*. Ed. by A. B. (Anthony Barnes) Atkinson and Thomas Piketty. Oxford University Press, p. 776. ISBN: 0191558230. URL: https://books.google.fr/books?hl=fr&lr=&id=eGaVAwAAQBAJ&oi=fnd&pg=PA448&ots=weNqqVKWAc&sig=Wx-ayJHI7MgXWMF5JCKCbB978Kw#v=onepage&q&f=false.

Atkinson, Anthony B. and François Bourguignon (2000). "Introduction: Income Distribution and Economics". In: *Handbook of Income Distribution*. Ed. by Anthony B. Atkinson and François Bourguignon. Geneva: Elsevier. Chap. Introducti, pp. 1–53.

Atkinson, Anthony B. and Salvatore Morelli (2012). "The Chartbook of Economic Inequality".

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez (2011). "Top Incomes in the Long Run of History". In: *Journal of Economic Literature* 49.1, pp. 3–71. DOI: 10.1257/jel.49.1.3.

Bagwell, Laurie Simon and John B Shoven (Aug. 1989). "Cash Distributions to Shareholders". In: *Journal of Economic Perspectives* 3.3, pp. 129–140. DOI: 10.1257/jep.3.3.129. URL: http://pubs.aeaweb.org/doi/10.1257/jep.3.3.129.

Bengtsson, Erik and Daniel Waldenström (Sept. 2018). "Capital Shares and Income Inequality: Evidence from the Long Run". In: *The Journal of Economic History* 78.3, pp. 712–743. DOI: 10.1017/S0022050718000347. URL: https://www.cambridge.org/core/product/identifier/S0022050718000347/type/journal_article.

Blanchet, Thomas, Ignacio Flores, and Marc Morgan (2018). "The Weight of the Rich: Improving Surveys with Tax Data".

Canberra Group (2001). *Expert Group on Household Income Statistics, The Canberra Group : final report and recommendations*. ISBN: 0968852408. URL: http://www.lisdatacenter.org/wp-content/uploads/canberra_report.pdf.

Chen, Peter, Loukas Karabarbounis, and Brent Neiman (2017). "The global rise of corporate saving".

Eije, Henk von and William L. Megginson (Aug. 2008). "Dividends and share repurchases in the European Union". In: *Journal of Financial Economics* 89.2, pp. 347–374. DOI: 10.1016/J.JFINECO.2007.11.002. URL: https://www.sciencedirect.com/science/article/pii/S0304405X08000949.

Endeweld, Miri and Paul Alkemade (2014). "LIS micro-data and national accounts macro-data comparison: Findings from wave i-wave viii". URL: http://www.lisdatacenter.org/wps/techwps/7.pdf.

Fama, Eugene F and Kenneth R French (Apr. 2001). "Disappearing dividends: changing firm characteristics or lower propensity to pay?" In: *Journal of Financial Economics* 60.1, pp. 3–43. ISSN: 0304-405X. DOI: 10.1016/S0304-405X(01)00038-1. URL: https://www.sciencedirect.com/science/article/pii/S0304405X01000381.

Floyd, Eric, Nan Li, and Douglas J. Skinner (Nov. 2015). "Payout policy through the financial crisis: The growth of repurchases and the resilience of dividends". In: *Journal of Financial Economics* 118.2, pp. 299–316. ISSN: 0304-405X. DOI: 10.1016/J.JFINECO.2015.08.002. URL: https://www.sciencedirect.com/science/article/pii/S0304405X15001415.

Francese, Maura and Carlos Mulas-Granados (2015). "Functional Income Distribution and Its Role in Explaining Inequality". URL: https://www.imf.org/en/Publications/

WP/Issues/2016/12/31/Functional-Income-Distribution-and-Its-Role-in-Explaining-Inequality-43415.

Gutiérrez, Pablo, Ramón López, and Eugenio Figueroa (Sept. 2015). "Top income measurement and undistributed profits". In: *Economics Letters* 134.September, pp. 138–140. ISSN: 0165-1765. DOI: 10.1016/J.ECONLET.2015.07.013. URL: https://www.sciencedirect.com/science/article/pii/S0165176515002864.

ILO (2019). "The Global Labour Income Share and Distribution". In: URL: https://www.ilo.org/global/statistics-and-databases/publications/WCMS_712232/lang--en/index.htm.

IMF (2007). "The Globalization of Labor". In: *World Economic Outlook*. Washington: International Monetary Fund (IMF). Chap. nº5, pp. 161–189. ISBN: 9781589066267. DOI: 10.2501/JAR-51-4-564-570.

Karabarbounis, Loukas and Brent Neiman (Feb. 2014). "The Global Decline of the Labor Share". In: *The Quarterly Journal of Economics* 129.1, pp. 61–103. DOI: 10.1093/qje/qjt032. URL: https://academic.oup.com/qje/article-lookup/doi/10.1093/qje/qjt032.

Milanovic, Branko (2017). "Increasing Capital Income Share and Its Effects on Personal Income Inequality". In: *After Piketty*. Ed. by H. Bouchey, J. DeLong, and M. Steinbaum. Harvard University Press. Chap. III. Dimen, pp. 235–258.

OECD (2013). *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*. OECD Publishing. ISBN: 9789264194823. DOI: 10.1787/9789264194830-en. URL: https://www.oecd-ilibrary.org/economics/framework-for-statistics-on-the-distribution-of-household-income-consumption-and-wealth_9789264194830-en.

Piketty, Thomas and Emmanuel Saez (Feb. 2003). "Income inequality in the United States, 1913-1998". In: *Quarterly Journal of Economics* 118.1, pp. 1–39. ISSN: 00335533. DOI: 10.1162/00335530360535135. URL: https://academic.oup.com/qje/article-lookup/doi/10.1162/00335530360535135.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (May 2018). "Distributional National Accounts: Methods and Estimates for the United States*". In: *The Quarterly Journal of Economics* 133.2, pp. 553–609. DOI: 10.1093/qje/qjx043. URL: https://academic.oup.com/qje/article/133/2/553/4430651.

Piketty, Thomas and Gabriel Zucman (Aug. 2014). "Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010". In: *The Quarterly Journal of Economics* 129.3, pp. 1255–1310. DOI: 10.1093/qje/qju018. URL: https://academic.oup.com/qje/article/129/3/1255/1818714.

Skinner, Douglas J. (Mar. 2008). "The evolving relation between earnings, dividends, and stock repurchases". In: *Journal of Financial Economics* 87.3, pp. 582–609. ISSN: 0304-405X. DOI: 10.1016/J.JFINECO.2007.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0304405X07002334.

Stiglitz, Joseph, Amartya K. Sen, and Jean-Paul Fitoussi (2009). *The measurement of economic performance and social progress revisited: reflections and overview*. Tech. rep. URL: https://hal-sciencespo.archives-ouvertes.fr/hal-01069384/.

Törmälehto, Veli-Matti (2011). "LIS and national accounts comparison". URL: http://www.lisdatacenter.org/wps/techwps/2.pdf.

United Nations (2009). *System of National Accounts 2008*. DOI: 10.18356/4fa11624-en.

Zucman, Gabriel (2015). *The hidden wealth of nations : the scourge of tax havens*. The University of Chicago Press, p. 129. ISBN: 022624542X. URL: https://books.google.fr/books?hl=fr&lr=&id=tpIpCwAAQBAJ&oi=fnd&pg=PR5&dq=zucman+hidden+wealth&ots=dxYkb-HeEd&sig=opV51z9h9CZ9J-gVxB_YC786OpU#v=onepage&q=zucmanhiddenwealth&f=false.

Zwijnenburg, Jorrit (2016). "Further enhancing the work on household distributional data–techniques for bridging gaps between micro and macro results and nowcasting methodologies for compiling more timely results". URL: http://www.iariw.org/dresden/zwijnenburg.pdf.

Zwijnenburg, Jorrit (May 2019). "Unequal Distributions: EG DNA versus DINA Approach". In: *AEA Papers and Proceedings* 109, pp. 296–301. DOI: 10.1257/pandp.20191036. URL: https://pubs.aeaweb.org/doi/10.1257/pandp.20191036.

Zwijnenburg, Jorrit et al. (2017). "Expert Group on Disparities in a National Accounts Framework: Results from the 2015 Exercise". URL: https://www.oecd-ilibrary.org/economics/expert-group-on-disparities-in-a-national-accounts-framework_2daa921e-en.

Tables and Figures

Table 1: Mapping survey-concepts to National Accounts

Survey Concept	System of National Accounts - 2008		
	SNA-Concept [1] = [2] + [3]	Matching [2]	Problematic [3]
Labour inc.	Comp. of Empl.	Wages & Sal.*	SSC: Employer & Employee
Self-Empl.	Mixed Inc.	Self-empl., Indep.	Rent of non-dwelling build.
Capital inc.	Op. Surplus	Rent of dwellings	Imp. rent for owner occup.
	Property Inc.**	Distrib. profits, Interest	Imp. rent to policyholders, Reinv. inc. on FDI & portfolio, Inv. inc. on pension funds

Major conceptual differences are reported in this table, for other minor differences see Table B.1 in OECD 2013. Notes: (*) The part of wages and salaries paid while an employee is on sick-leave are recorded as a part of social insurance benefits in SNA. (**) SNA does not deduct any expense when deriving property income. Source: Compiled by the author based on definitions in United Nations 2009 and OECD 2013.

Table 2: Countries using register data in surveys

Country	Period
AUT	1994 - 2016
BEL	1995 - 2000
CAN	1998 - 2013
DNK	1995 - 2013
FIN	1991 - 2016
NLD	2004 - 2013

Source: Metadata from `www.lisdatacenter.org`

Table 3: Availability of imputed rents by wave in balanced panel

Country	I	II	III	IV	V	VI	VII
AUS			x	x	x	x	
AUT			x	x	x	x	x
BEL	x						
CHE				x	x	x	
CHL	x	x	x	x	x	x	x
CZE				x	x	x	
DEU	x	x	x	x	x	x	x
DNK	x	x	x	x	x	x	
ESP			x	x	x	x	x
EST		x		x	x	x	
FIN	x	x	x	x	x	x	x
FRA		x	x		x		
ITA	x	x	x	x	x	x	
LUX		x	x	x	x	x	
NLD			x	x	x	x	
NOR	x	x	x				
SVK				x	x	x	
SVN				x	x	x	
USA	x	x	x	x	x	x	x

Source: Metadata from `www.lisdatacenter.org`

Table 4: Partial derivatives

f'_x	if γ=1	if $\gamma \neq 1$
	[1]	[2]
$S_q(K)'$	$S_q^K - S_q^L$	$(S_q^K - S_q^L) \times \gamma(K\gamma + L)^{-2}$
$S_q(S_q^L)'$	L	$L \times (K\gamma + L)^{-1}$
$S_q(S_q^K)'$	K	$K \times \gamma(K\gamma + L)^{-1}$
$S_q(\gamma)'$		$KL(S_q^K - S_L^K) \times (K\gamma + L)^{-2}$

These are the formulas of partial derivatives, for each variable in the model in equation 4. They are used to estimate empirical marginal effects. We compare cases with and without differences of income concepts across data sets ([2] and [1] respectively). The relevant difference is the multiplication by a given factor in [2].

Table 5: Empirical Estimate of Partial derivatives, Top 10% Share

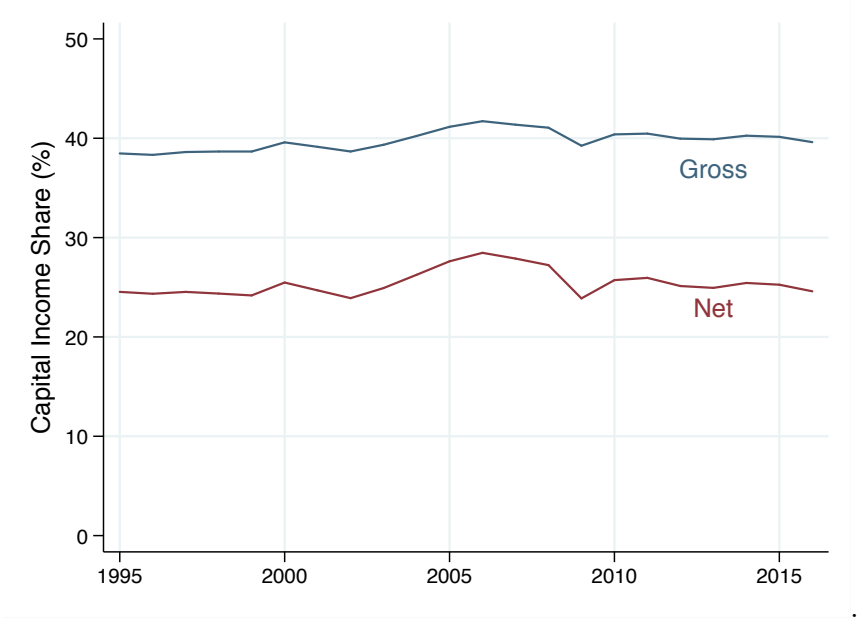
Country or Area	$S_q(K)'$ [1]	$S_q(S^L_{top10\%})'$ [2]	$S_q(S^K_{top10\%})'$ [3]	$S_q(\Phi_h)'$ [4]	$S_q(\epsilon_K/\epsilon_L)'$ [5]	$S_q(\gamma)'$ [6]
Panel (1995–2013):						
—Austria	0.03	0.92	0.08	0.01	0.03	0.05
—Canada	0.02	0.93	0.07	0.01	0.01	0.03
—Czech Republic	0.04	0.93	0.07	0.03	0.03	0.11
—Denmark	0.06	0.94	0.06	0.06	0.03	0.12
—Finland	0.09	0.90	0.10	0.07	0.04	0.14
—Germany	0.06	0.91	0.09	0.02	0.05	0.08
—Greece	0.03	0.85	0.15	0.01	0.01	0.02
—Italy	0.07	0.87	0.13	0.02	0.07	0.09
—Netherlands	0.04	0.93	0.07	0.04	0.02	0.08
—Poland	0.05	0.93	0.07	0.02	0.04	0.09
—United Kingdom	0.01	0.91	0.09	0.01	0.01	0.02
Total:	0.04	0.91	0.09	0.02	0.04	0.06
U.S. (1974–2011):						
—Survey	0.03	0.92	0.08	0.01	0.02	0.04
—Tax	0.07	0.90	0.10	0.03	0.04	0.07
—DINA	0.38	0.76	0.24			

These are empirical estimates of the marginal effects in table 4. In Austria, an isolated 1 unit variation in the capital share [1] only results in a variation of 0.03 units. But the same evolution in the concentration of labour income [2] produces a systematic increase of 0.92 units.

Table 6: Modeled Contribution to Variation in Top 10% Share

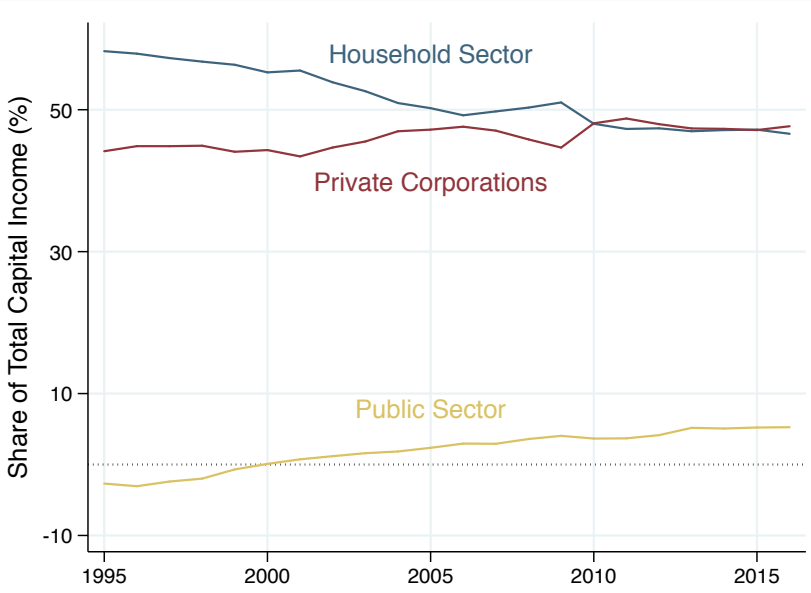
Country or Area	Estimated Contribution						Total Variation in Period (%)		
	K [1]	S_q^L [2]	S_q^K [3]	Φ_h [4]	ϵ_K/ϵ_L [5]	γ [6]=[4]+[5]	Model Δ [7] = [1] to [5]	Real Δ [8]	Error [8]-[7]
—Panel (1995–2013):									
—Austria	0.11	3.97	0.36	-0.11	0.19	0.08	4.51	4.39	0.12
—Canada	0.06	2.90	0.73	-0.07	0.12	0.05	3.75	3.81	-0.06
—Czech Republic	-0.08	2.26	-0.47	0.05	0.34	0.39	2.10	1.97	0.13
—Denmark	0.35	3.11	0.32	-0.25	-0.26	-0.51	3.26	3.29	-0.03
—Finland	-0.04	1.33	-0.33	0.25	0.00	0.25	1.20	1.86	-0.65
—Germany	0.35	4.68	-0.16	-0.12	-0.17	-0.28	4.59	4.92	-0.33
—Greece	0.15	5.94	0.70	-0.12	-0.19	-0.31	6.48	6.52	-0.04
—Italy	-0.02	1.72	0.91	-0.44	-0.59	-1.03	1.58	1.54	0.04
—Netherlands	-0.03	4.33	-0.01	-0.16	0.27	0.11	4.40	5.13	-0.73
—Poland	0.14	-2.74	-1.38	-0.47	0.10	-0.37	-4.36	-4.22	-0.14
—United Kingdom	-0.13	2.59	-0.30	-0.07	-0.05	-0.11	2.04	2.11	-0.07
Total:	0.06	3.48	0.07	-0.23	0.09	-0.15	3.46	3.54	-0.08
U.S. (1974–2011):									
—Survey	0.38	9.12	-0.10	-0.01	-0.26	-0.27	9.14	9.11	0.03
—Tax	0.64	13.17	1.33	-0.03	-0.71	-0.74	14.39	14.27	0.12
—DINA	1.84	9.50	-0.05				12.37	12.10	-0.27

The concentration of labour income appears to be the dominant factor for variations in surveys' top 1% share estimate. The role of the capital share and its distribution is largely undermined. The difference is weaker in tax data, but still present. DINA estimates are, by definition, not distorted in this sense.

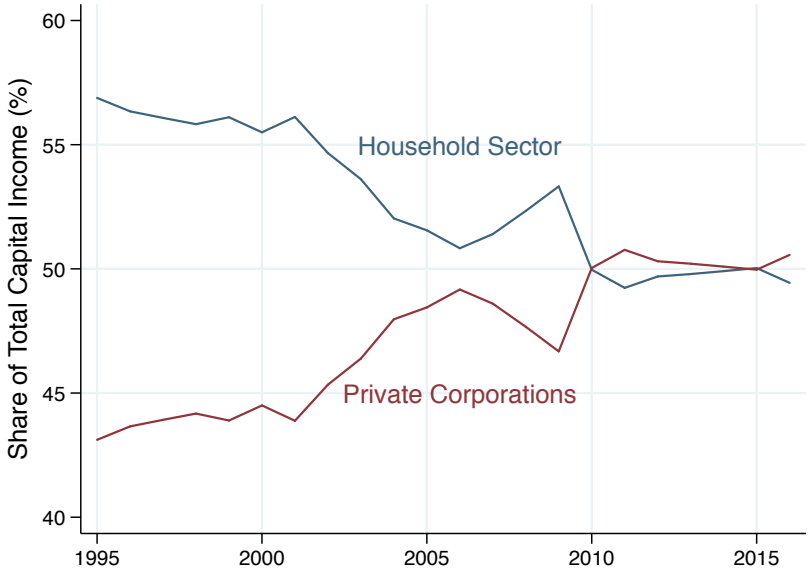


In a panel of 19 countries, the gross capital share of national income increases near 2 percentage points in 2 decades. When we subtract capital depreciation (i.e., Net capital share), the share is lower and describes a relatively more stable trend overall. Countries included (see figure 1 for country-level shares): Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovakia, Sweden, Switzerland and the United Kingdom. Income from different countries is aggregated based on yearly average Market Exchange Rates. The United States is studied separately. Own estimates from United Nations' National Accounts.

Figure 1: Capital Share of National Income, Balanced Panel (1995-2016)



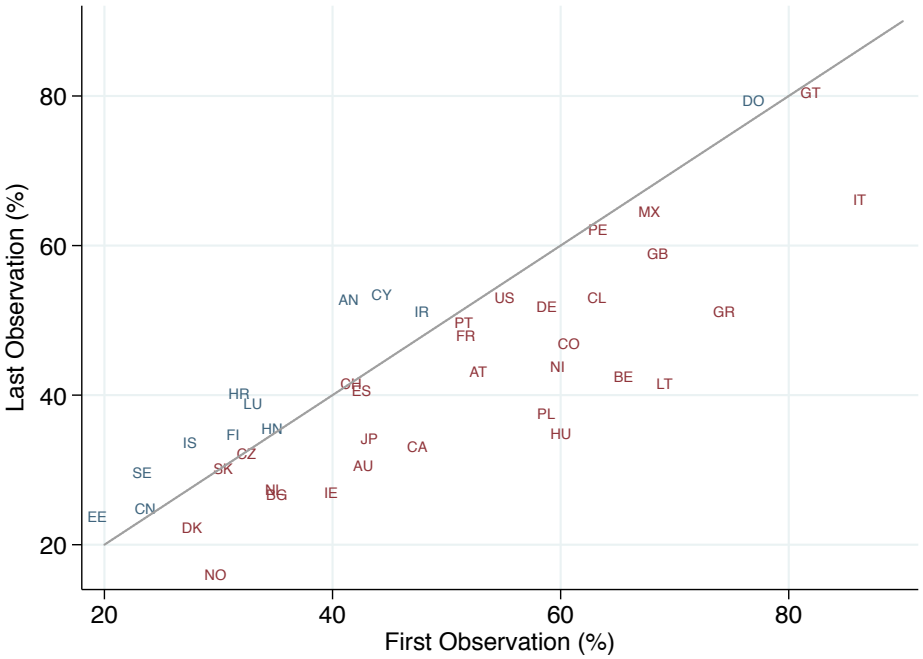
(a) Including the Public Sector



(b) Excluding the Public Sector

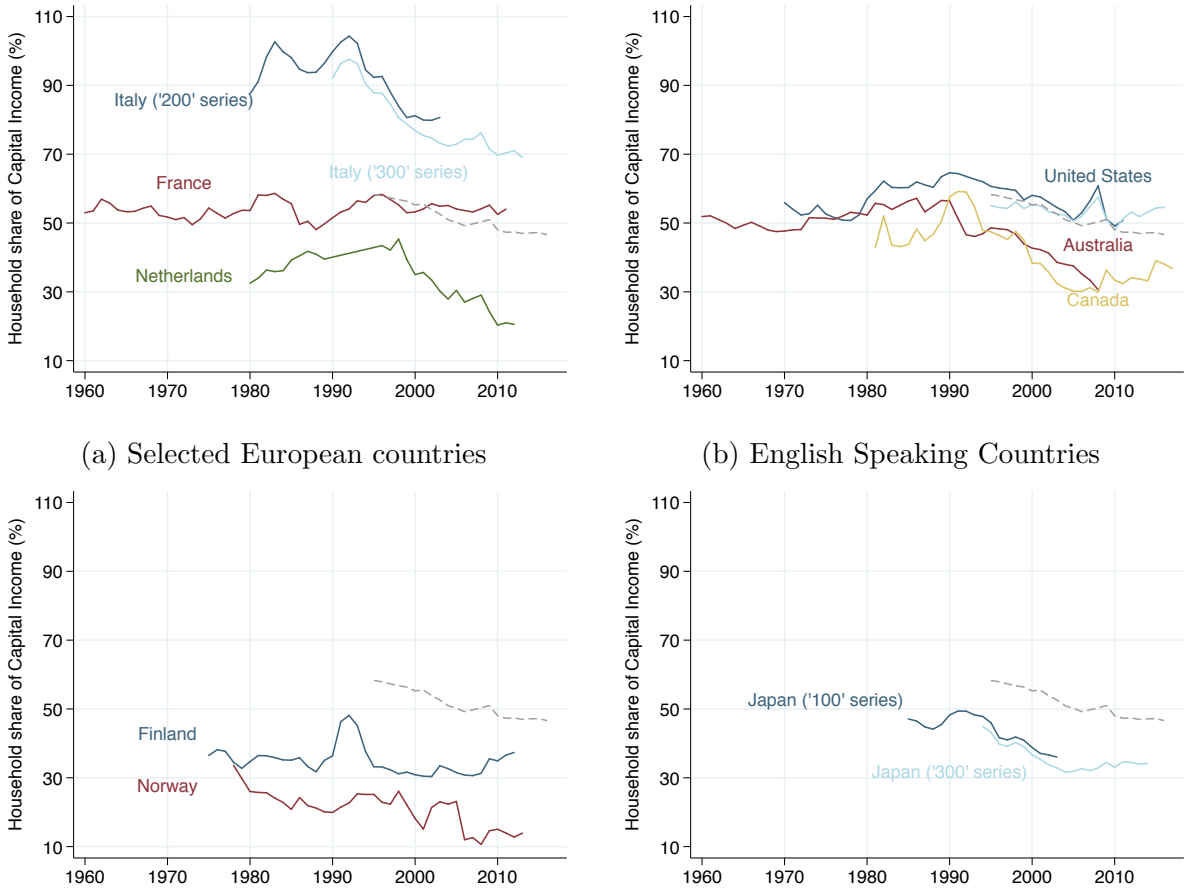
Both the Public Sector and Private Corporations increased their share of capital income, while the Household share decreases through the period. In 1995, the household sector received 57% of the capital income produced in a balanced panel of 19 developed countries (excluding the public sector). Two decades later, it receives less than 50%. Countries included: Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Sweden, Switzerland and the United Kingdom. Income from different countries is aggregated based on yearly average Market Exchange Rates. The United States is studied separately.

Figure 2: Decreasing Household Share of Gross Capital Income, Balanced Panel (1995-2016)



The share of capital income received by households, as opposed to public and private corporations, decreased in 30 out of the 43 cases that have at least 6 observations during the period (excluding the public sector). That is, it decreases in near 70% of cases. The countries that experienced an increase are those which already had relatively low shares to start with.

Figure 3: Decreasing Household Share of Capital Income, Unbalanced Panel (1995-2015)



(a) Selected European countries

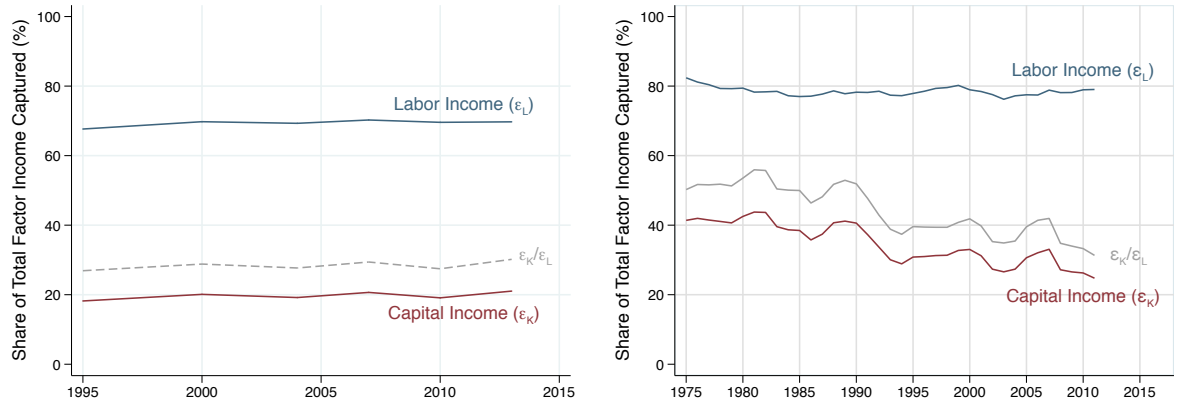
(b) English Speaking Countries

(c) Scandinavian Countries

(d) Japan

The grey dashed line represents the aggregate tendency in the balanced panel of 19 countries presented in figure 2. Most countries with long-run data exhibit a decreasing trend starting before the beginning of the panel, around 1990. Relatively more stable trends are described in previous decades.

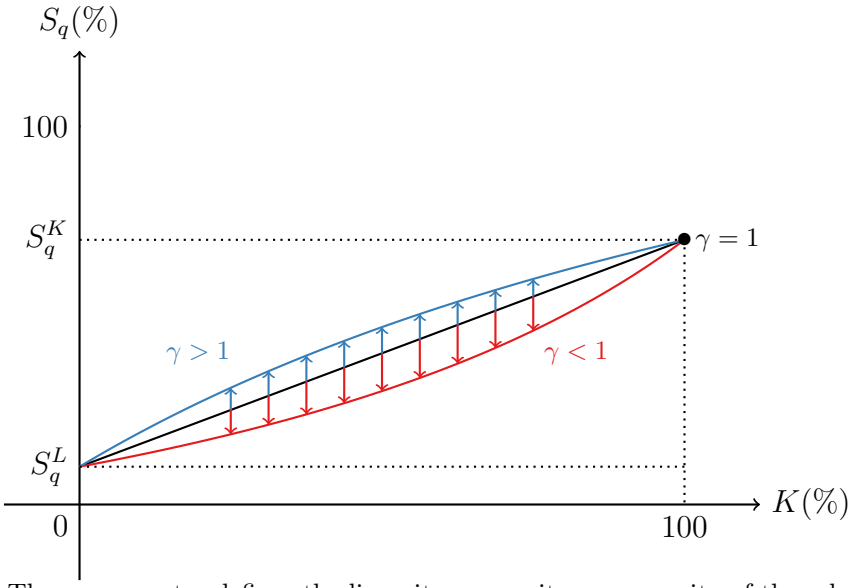
Figure 4: Decreasing Household Share of Capital Income, Long-run



(a) Survey Data: Balanced Panel (1995-2013) (b) Tax Data: United States (1975-2015)

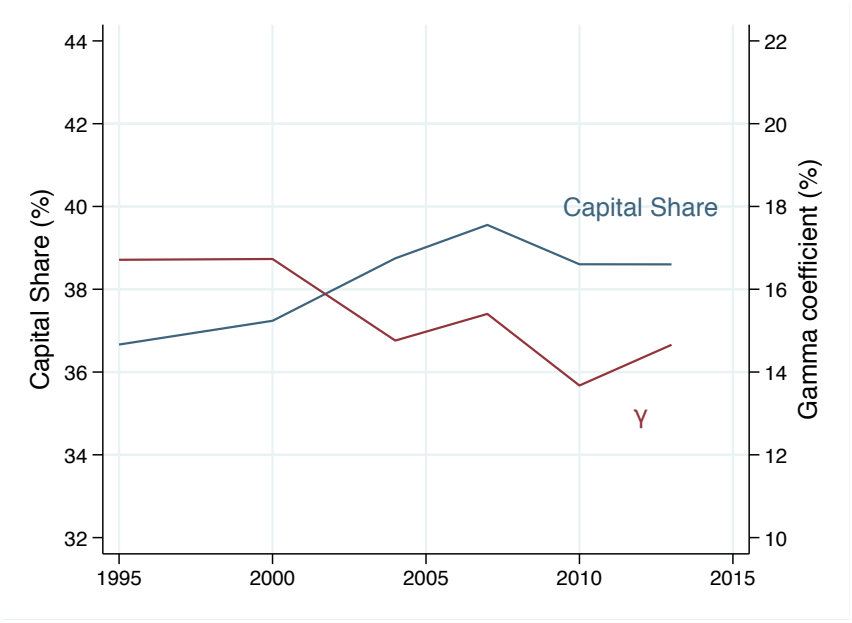
Labour income is better represented in both surveys and tax data, relative to capital income. The difference in the underestimation of each factor income, however, appears to be wider in surveys. Table D.1 displays the weight of each country in the Panel, in terms of aggregate national income. The ϵ_K/ϵ_L ratio is displayed here yet it is only commented in the following sections. Survey estimates for the United States are provided separately

Figure 5: Unequal Measurement Error in Surveys and Tax Data

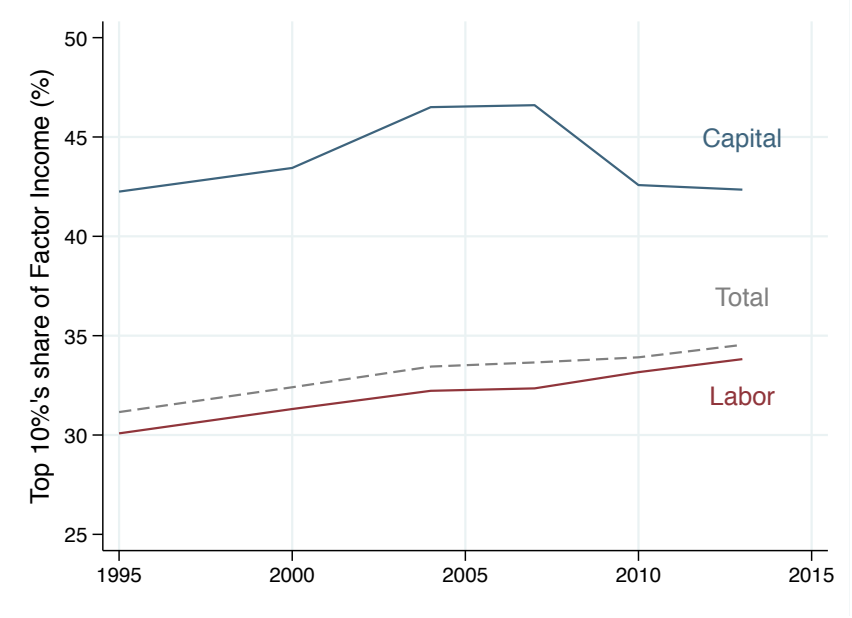


The γ parameter defines the linearity, convexity or concavity of the relation between the capital share and top income shares, while the relative concentration of factor incomes determines the slope and both upper and lower boundaries of top shares.

Figure 6: From the Capital Share to Top Income Shares



(a) Capital Share and Gamma



(b) Factor Income Concentration in Top 10%

Both the capital share and the top 10% share of total income increase during the period. The latter estimate follows the concentration of labour income extremely closely. The γ coefficient remains rather low during the period, likely filtering out the influence of both the capital share and its distribution over the top income share.

Figure 7: Evolution of relevant variables, Balanced Panel (1995-2013)

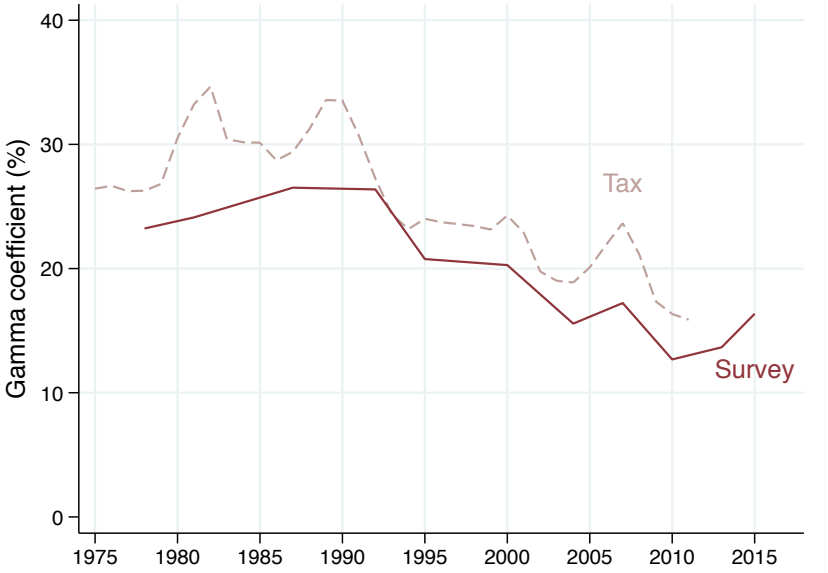
(a) Survey Data

(b) Tax Data

(c) Distributional National Accounts – DINA

Total income concentration increases substantially in all cases. It appears to follow closely the concentration of labour income especially in survey and tax data. Survey estimates are the only that show a decrease in the concetration of capital income. Tax data comes from Piketty and Saez 2003 and updates by the authors. DINA estimates correspond to the personal factor income definition in Piketty, Saez, and Zucman 2018.

Figure 8: Top 10%'s Factor Income Shares, United States (1975-2015)



The capital share increases during the period. Both surveys and tax data account for a progressively decreasing share of national capital income since the 1990's. The figure for tax data is, however, at least twice as high as the one for survey data overall.

Figure 9: The Gamma coefficient, United States (1975-2015)

Acknowledgements

I gratefully acknowledge funding from ERC (Grant 340831), Ford Foundation, INET (Grant INO14-00023) and from other partners of the World Inequality Lab. I thank Facundo Alvaredo, François Bourguignon, Thomas Blanchet, Emmanuel Flachaire, Jérôme Gautié, Amory Gethin, Elvire Guillaud, Pablo Gutiérrez, Julián Messina, Marc Morgan, Nora Lustig and Li Yang for helpful discussions of earlier versions of this paper.