# DATA COMBINATION FOR INEQUALITY MEASUREMENT AND PROGRESSIVITY ASSESSMENT IN A FORMALIZING ECONOMY

David Garcés Urzainqui
Jane Kanina
Josephine Mugure

AERC Working Paper GOK-AERC-UCPH-DERG-013

**Bringing Rigour and Evidence to Economic Policy Making in Africa**

CONSORTIUM POUR LA RECHERCHE ÉCONOMIQUE EN AFRIQUE

# Data Combination for Inequality Measurement and Progressivity Assessment in A Formalizing Economy

**David Garcés Urzainqui**

Corresponding author. Department of Economics, University of Copenhagen (d.garces.urza@gmail.com).

**Jane Kanina**

**Josephine Mugure**

Kenya Revenue Authority (KRA)

AERC Working Paper GOK-AERC-UCPH-DERG-013

African Economic Research Consortium,

Nairobi November 2026

**Disclaimer:** The findings, opinions and recommendations are, those of the author, and do not necessarily reflect the views of the Consortium, its individual members or the AERC Secretariat.

**Published by:** The African Economic Research Consortium

P.O. Box 62882 - City Square

Nairobi 00200, Kenya

**© 2026**, African Economic Research Consortium

## Acknowledgements:

This paper was developed as part of the Royal Danish Embassy in Nairobi (RDE)-supported collaborative research project entitled "Economic Research and Policy Making in Kenya (ERPMK)", supported by the African Economic Research Consortium (AERC). We thank Noreen Paredes for excellent research assistance.

**Abstract**

Inequality estimates that address the under coverage of top income earners in household surveys are rare in Africa. We measure interpersonal inequality in market income in Kenya between 2016 and 2022, combining household survey and tax data to address the limitations of each of these data sources. Since our tax data mainly cover individuals in formal employment, we investigate variations of data combination approaches centred on correcting the income distribution by type of employment. Both these approaches and more standard methods reveal very high levels of inequality. We further evaluate the impact of personal income taxes on inequality using combined tax and survey data. We find that the Kenyan personal income tax (PIT) is progressive but front-loaded, with the recent expansion of statutory contributions largely offsetting the effects of changes to the income tax schedule since 2019 and weakening progressivity among formal workers.

# 1 Introduction

Inequality estimates shape public discussion and influence the perceptions of citizens and policymakers. Awareness of the actual evolution of underlying inequality trends is not just of, but essential for an informed design of tax policy. The availability of reliable data on the income distribution is a necessary condition to rigorously assess the redistributive impact of policies and, for example, an essential prerequisite for studying the effective progressivity of Personal Income Tax (PIT) structures.

This is far from a trivial requirement. Tax data in Kenya only cover the relatively small share of the population that is part of the formal workforce. On the other hand, it is well established that household surveys, the backbone of conventional inequality analysis, fail to capture the top of the income distribution. The incorporation of tax data into the study of income inequality over the last few decades has revolutionized our understanding of this phenomenon (Alvaredo et al., 2013; Atkinson and Piketty, 2007). While some researchers have also incorporated National Accounts (Garbinti, Goupille-Lebret, and Piketty, 2018; Piketty, Saez, and Zucman, 2018; Zwijnenburg, 2019) and other data sources, the combination of tax and survey data has emerged as the canonical approach for addressing the shortcomings of survey data, with researchers employing a wide variety of methods, reviewed in Lustig and Vigorito (2025) to leverage the strengths of each of these data sources.

The implementation of these approaches has mainly focused on higher-income countries. However, the underlying issues leading to the 'missing rich' problem (Lustig, 2019), such as difficulties in obtaining responses from the rich or income misreporting, are unlikely to be less serious in developing countries such as Kenya. In addition, capacity and budget constraints often prevent these countries from incorporating sophisticated but costly sampling techniques to better capture the upper tail of the income distribution, or even from collecting high-quality income data at a sufficient frequency. Existing applications to regions and countries such as Latin America (De Rosa, Flores, and Morgan, 2024), the Middle East (Assouad, 2023), or India (Chancel and Piketty, 2019) often reveal very large income disparities, sometimes at odds with estimates of levels and trends exclusively based on household surveys.

Unfortunately, such income inequality estimates combining household survey and administrative tax data are still relatively uncommon in the context of Africa (Chatterjee, Czajka, and Gethin, 2023; Czajka, 2017; Jäntti, Nalukwago, and Waiswa, 2022) due to data quality and availability issues concerning both household surveys and administrative tax data.

The discourse on inequality in Africa is thus dominated by consumption inequality estimates, since the underlying data are more reliable. In addition, stronger emphasis has traditionally been placed in the region on the use of consumption expenditure to measure poverty, as they are believed to better capture differences in living standards (Ravallion, 2016).

However, income inequality estimates, which may widely differ from their consumption counterparts, convey valuable information on disparities in command over resources that have important economic, social, and political implications. PITs are defined and imposed on taxable income. Thus, it is highly relevant for policymakers to study differences in this quantity determining the ability to serve tax liabilities. Projections based on the few countries for which data combination has been possible suggest that income inequality estimates that adequately capture top income groups may lead to substantially different conclusions on the extent of inequality in the region (Chancel et al., 2023).

We aim to contribute to our understanding of economic disparities in Kenya. For that purpose, we draw on detailed administrative tax data provided by the Kenya Revenue Authority (KRA) and household surveys conducted by the Kenya National Bureau of Statistics (KNBS) between 2016 and 2022. The tax records are obtained from three different sources. First, employment income reported by employers through the Pay-As-You-Earn (PAYE) system. Second, income declared by individuals in their annual individual tax returns (ITRs). Third, other income subject to withholding and rental income taxes, mainly interest, dividends, rents and professional fees.

Household surveys also provide data on income from self-employment and informal employment. We refer to the resulting income concept as (pre-tax) individual market income: labor and capital income accruing directly to individuals, as observed in survey and tax records. Our post-tax inequality estimates reflect the impact of income taxes, but do not incorporate social and informal transfers, sources of income that are likely particularly relevant at the bottom of the distribution.

One distinct element of our data combination exercise, as compared to most applications in higher income countries, is that the coverage of individuals in tax data depends on their labour market status. While we obtain some information on non-employment income from various tax returns, only a very small share of informal employees and self-employed workers report accurately. In contrast, PAYE returns are filed by employers and as third-party reported data, they are generally perceived as accurate (Kleven et al., 2011). They thus notably enhance the

accuracy of the information on the distribution of formal-sector workers with respect to the survey, limited by statistical capacity constraints and a more pronounced focus on other aspects such as consumption. In Kenya and often in SSA, when incomes in the informal economy largely overlap with those in the formal sector, the standard method simply replaces the survey with tax data beyond a certain income level (Günther and Launov, 2012; Bartels and Metzing, 2018; Blanchet, Flores, and Morgan, 2022; Jenkins, 2017). This approach has the drawback of missing informal workers in the higher-income segment measured with tax data and may also misrepresent the distribution of formal employees in the lower-income range.

We thus propose a simple alternative data combination procedure to obtain improved inequality estimates for developing countries with a large informal economy where incomes largely overlap with those in the formal sector. We rely on administrative data to (by definition) accurately measure, fully or partially, the formal sector, while information about the informal economy is obtained from survey data. The link to combine survey and tax data is given by a survey variable capturing whether income tax is deducted from the worker's earnings. Inequality indicators can then be readily computed from combined tax and survey data after proportionally correcting sampling weights to account for the underrepresentation of formal workers. As will be discussed below, whether this procedure is more appropriate than the standard approach dealing with the overall income distribution, depends on the reasons that explain the discrepancies between the income distribution for formal workers in tax and survey data.

This approach requires a nationally representative household survey providing information on formality status and the income distribution, and administrative tax data on personal income covering high-income earners in formal employment better than the survey. Following the classification in Lustig and Vigorito (2025), it can be seen as combining replacement and reweighting elements. We believe that it might prove useful in the context of ongoing digitalization of tax collection and administration, and hence increasing availability of and improved access to tax data across Africa, where high-quality administrative tax data may often disproportionately cover income derived from formal employment.

An additional challenge in our context is to estimate the distribution of income as measured by the household survey across all the paid employees and self-employed. Unfortunately, income data are reliable only for a non-representative subset of them, skewed towards those in more stable jobs. We use that set of reliable observations to predict income based on consumption, allowing for heterogeneous saving rates, and several relevant individual

characteristics, estimating separate models for employees and the self-employed. We then use those models to impute income for observations without reliable income information.

Our data combination exercise allows us to provide estimates of interpersonal inequality in market income that address the undercoverage of top incomes in household surveys. We find high levels of inequality, with pre-tax Gini coefficients ranging between 0.62 and 0.67, and earners in the top 1% capturing over 20% of aggregate income. In all cases, data combination yields higher inequality estimates than each of the data sources regarded in isolation. Surveys particularly struggle to precisely measure top income shares. The approach suggested in this paper tends to lead to somewhat higher Gini coefficients than the standard approach of replacing the top tail with tax data.

On this basis, we study the distributional impact of PITs. Our data combination approach is well suited for that purpose because it preserves the formal-informal distinction when combining the data, as required to credibly ascertain the effects of PIT structure on inequality in a context characterized by high informality. Despite the well-known difficulties of household surveys in capturing the top of the income distribution, existing literature has typically evaluated the progressivity of tax systems or reforms in developing countries using household surveys (Lastunen et al., 2024; Inchauste and Lustig, 2017). We show how combining both data sources can substantially alter the conclusions on the impact of PITs on inequality. Survey data alone understate the inequality-reducing effect of the PIT, because they miss top taxpayers, while tax data alone misrepresent it, because they ignore informal earners outside the tax net.

We then rely on structural progressivity measures and the effective progressivity analysis of this combined distribution to assess the Kenyan PIT and its recent reforms. The progressivity of the system comes largely from the personal relief, which shields lower-paid workers from income tax. Beyond that point, the schedule offers little additional progression. The reforms since 2019 have increased the effective progressivity of the PIT, at the price of concentrating steep increases in average tax rates within a narrower income range containing the first effective taxpayers. However, the parallel expansion of statutory contributions has offset these gains among formal workers. Finally, we show that income types subject to lower tax rates, such as interest and dividends, are, as observed in other contexts, concentrated at the top. They are also large enough in comparison to personal income to undermine the progressivity of the tax system at very high-income levels.

These results may be of broader interest, since many of the elements of the Kenyan PIT schedule are common in the African context. These include the existence of substantial personal relief effectively exempting a large portion of the labor force, the limited progression of marginal tax rates thereafter, and recent attempts to increase the tax rates faced by top income earners.

This paper is structured as follows. Section 2 describes the data underlying the analysis. Section 3 discusses the methodology used to derive the distribution of income across all income earners from tax and survey data. Section 4 presents the inequality measures resulting from this exercise between 2016 and 2022, also covers the analysis of the structural and effective progressivity of the Kenyan PIT structure. Section 5 discusses income taxed outside the PIT schedule. Finally, in Section 6 we derive conclusions and policy recommendations from this analysis.

# 2 Data

## 2.1 Household survey data

This study draws on several rounds of nationally representative household survey data collected by the KNBS. We use the 2015/16 Kenya Integrated Household Budget Survey (KIHBS). The KIHBS is a large survey that collects information over 12 months on a wide array of demographic, welfare and economic indicators, such as education, health, employment, income, consumption expenditure, housing, agricultural production, etc. The survey informs various official statistics (CPI, National Accounts). It is designed to be representative for urban and rural areas at the county level. The 2015/16 round covers 92,846 individuals in 21,773 households.

The KIHBS is collected once every ten years. Therefore, the KNBS introduced the Kenyan Continuous Household Survey (KCHS) to obtain timely information on some socio-economic indicators. The KCHS uses a nationally representative sample of households, selected through a multi-stage stratified sampling process, and employs a rolling sample design, where data are collected continuously throughout the year. While this survey collects substantially less extensive and detailed information than the KIHBS, it includes an employment module that is used to compile official labor force statistics aligned with the International Labor Organization (ILO) standards. The questionnaire in this module is designed to correspond to that in the KIHBS (Pape and Kariuki, 2019). We use four KCHS rounds corresponding to 2019, 2020,

2021 and 2022. Their sample sizes range between 17,042 and 24,638 households including 68,277 to 86,647 individuals.

Income data in the KIHBS are substantially more comprehensive, as the survey includes data on transfers from various sources (social networks, NGOs, public, remittances), as well as pension, rental, interest, and investment income. However, both surveys are broadly comparable in their coverage of income derived from employment, to which we will therefore restrict our attention in survey data. Both surveys use the same reference period of 7 days for economic activity and collect income data that allow deriving monthly earnings based on a similar questionnaire. Income data are collected on the primary and secondary economic activity of the individual. The KCHS also collects information on income from any other jobs, but amounts are comparatively quite small.

Income data collection is similar, but not identical, across surveys. On the one hand, the KCHS uses different questions for individuals, depending on whether their primary economic activity is self-employment (own-account workers and working employers) or paid employment outside the household. On the other hand, the KIHBS collects that information with a single question common to all economically active respondents. In addition, it includes a question on all allowances received. This information is also collected by the KCHS for employees through separate questions on medical, housing, transportation, and other allowances. Both surveys take a different approach for casual workers, who provide information on the average daily rate and the number of days worked in the last month. All surveys collect income data for paid employees and the self-employed, including those in agriculture, but coverage of contributing family workers and other smaller categories varies, as the KCHS 2021 and 2022 do not elicit information on their income. We therefore restrict our attention to paid employees and the self-employed in our main analysis, who make up 65% to 75% of the economically active population.

Data preparation mainly involves dealing with two issues. First, there is a certain lack of clarity regarding the frequency of recorded payments. We deal with this issue following a conservative approach, where recorded duration is respected except in flagrant cases. Second, there are some missing values. For individuals who provide a range of income instead of an exact value, we assign them an income randomly drawn from the empirical distribution within that range. We deal with the remaining missing and unreliable values by imputing incomes for employees and the self-employed, as explained in further detail in Section 3.1. For this purpose,

we use data on consumption expenditure, household demographics, education, job characteristics and housing characteristics available from these surveys.

## 2.2 Administrative tax data

We combine household survey data with detailed administrative tax data provided by the KRA. We use the universe of PAYE returns, ITRs, withholding tax returns, and monthly rental income (MRI) returns filed between 2016 and 2024. Since the last available survey round is KCHS 2022, our inequality estimates cover 2016–2022, while the 2024 tax data are used for the progressivity analysis in Section 5.

PAYE is a wage withholding tax system in which companies deduct a portion of employees' pay every month and send that tax to the KRA. It accounts for the majority of PIT revenue (over 95% on average from FY 2017/18 to FY 2021/22). The two main sources of PAYE tax revenue are the public sector and large firms, respectively contributing 39% and 35% over that period. We observe employment income at a yearly frequency, which consists of basic pay and wages, taxable allowances in cash or in kind (housing, transport, etc.), and bonuses, commissions, and overtime pay.

We complement PAYE data with data from PIT returns, which include data on employment and non-employment income. Individuals are expected to file these returns annually to reconcile the tax withheld with their overall tax liability. Income sources reported in the ITR include business,[1] farming, rental, interest, commission, and other income. ITRs also provide data on income derived from partnerships and on income from trusts and estates.

We further obtain data from MRI returns on rental income subject to the Residential Rental Income Tax. Since 2016, this is the default regime for resident landlords receiving between KSh 288,000 and 15 million per year in rental income (between KSh 144,000 and 10 million before 2021). Under this regime, landlords file monthly returns and pay 7.5% (10% before 2024) of gross rental income as a rental tax. Residential rental income outside that range and all commercial property rental income fall under the ordinary annual income regime (PIT for individuals, CIT for companies), which is applied to net rental income. When rent is paid by an appointed withholding agent (in practice, government bodies, large companies and property

[1] Sole proprietors with gross turnover within a certain range (between KSh 1 and 25 million at the time of writing) can choose whether to pay turnover tax (TOT) on their gross turnover. Sole proprietors below or above that range, or opting out of the TOT regime, pay PIT on their net profits. We do not use TOT returns because the tax base there is self-reported gross turnover data.

managers), 10% of gross rent on commercial property is withheld as an advance tax, which the landlords of residence credit against their liability in the annual tax return.

Rental income from commercial property is thus one of the types of income subject to a withholding tax (WHT). Withholding taxes are collected at source by the payer, typically banks, companies, and other large institutions, which deduct a share of the gross payment and remit it to KRA. This collection mechanism is aimed at fostering compliance through third-party reporting and simplified tax payment. We use data on the universe of WHT returns during our study period. An important consideration for our data combination exercise is whether WHT is final. For resident individuals, this is the case for interest and dividends. For other types of income, such as professional, commission or royalty income, WHT is an advance payment credited against the annual income tax liability. The corresponding income should thus be reported in the individual return and included in the PIT base. We distinguish between incomes subject to these two tax treatments, referring to them as final and non-final withholding income, respectively. The WHT rate for residents depends on the type of payment, ranging from 5% to 15% for most categories (see Table A.1).

We combine the different tax returns using anonymized taxpayer numbers. Since income observed in WHT and MRI returns could correspond to individuals or companies, we only retain those observations that can be matched to individuals in PAYE or ITR records during the study period. We compute employment income from a combination of PAYE returns and ITRs. Individuals in the PAYE system are often missing from ITR data (17% of cases) because they simply do not file their individual tax returns. However, for 76% of observations, taxpayer numbers can be linked across ITR and PAYE, and employment incomes exactly match across sources in most (78%) of those cases. Discrepancies, which mostly amount to less than 5% of the observed income, are resolved by taking the largest employment income as valid, as there is no incentive for workers or employers to report inflated earnings numbers leading to larger tax liabilities.

Besides employment income, our income variable includes non-employment income. This consists of the sum of non-employment income from ITRs, final withholding income, and MRI rental income. We take a conservative approach and only add excess non-final withholding tax income when there is unequivocal evidence that this (gross) income was not reported in the ITR. We observe any non-employment income for 8.4% of individuals (5.2% in ITRs), any MRI income for 2.5%, and any WHT income, which results in additional income in about half of the cases, for 4.8% of them. For a small share of observations (4.5%), that mainly consists

of people fully deriving their income from sources other than employment, ITRs but no PAYE returns are available. Finally, for 1.7% of observations (ITR nil-filers or non-filers in that year), only WHT or MRI returns provide data on income.

One conceptually relevant question to consider when matching survey and tax data is construct validity. Our combined tax data from various returns include certain forms of income (essentially those corresponding to property income) that are admittedly not part of the income concept measured in the employment module of the household surveys. We still focus on market income as our baseline income concept because such property income is a relevant driver of inequality. Two further arguments support this choice. First, these income sources not conceptually covered by survey data are typically very heavily concentrated at the top of the distribution, which we will measure from tax data. It is thus safe to assume that the part of the distribution captured by household surveys would not look very different if households were also asked to report capital income.[2] Second, individual tax returns contain information on non-employment income derived from a variety of sources, including investment, business and professional income. While the data do not always allow us to neatly disentangle these income sources, business and professional income account for an important part of non-employment income in tax data and conceptually correspond to income covered by the survey. We thus keep non-employment income as part of our baseline income variable and refer to our estimates as capturing market income inequality. Results for personal income, employment income and income reported in PAYE returns are reported in the appendix.

Two limitations with respect to our coverage of capital income should be mentioned. First, our information on property income is limited to that visible to tax authorities. Interest on informal loans, as well as returns paid by smaller, unregulated Savings and Credit Cooperative Societies (SACCOs), agricultural rent, or informal rentals are unlikely to be captured in these data. Second, only 1.6% of dividend and 3.3% of interest income can be attributed to individual tax filers (see Table A.2), as compared to 12.6% for professional fees or 35.4% for commission income.[3] The share of interest accruing to companies is an order of magnitude larger than that

[2] Note that imputed rents are not part of our income concept.

[3] These shares are likely to be larger when expressed in terms of capital income derived by residents. Reported shares in Table A.2 exclude payment categories that never apply to resident individuals but may otherwise include payments to non-residents. This may partly, but not fully, explain the large share of unmatched income. For example, around 53% of dividend income in the WHT data is taxed at the rate for residents and East African Community citizens (5%, see Table A.1), and only 5% can be matched to individuals or companies. This may also be related to our limitations in identifying individuals and companies only based on PAYE, individual and corporate income tax returns.

directly flowing to individuals in our data. Our analysis, which does not currently consider retained profits, is thus likely to underestimate the relevance of capital income at the top of the income distribution. We also adhere to the canonical definition of income as regular income flows, excluding capital gains, lump-sum pension payments and gambling winnings.

# 3 Methodology

## 3.1 Income distribution in household surveys

To combine household survey and administrative tax data, the first challenge is to obtain an estimate of the interpersonal income distribution across survey respondents. Unfortunately, these surveys only provide reliable data on a subset of them. We use an imputation model for income based on consumption and several relevant individual characteristics:

$$\log(y_i) = \gamma_0^e + \sum_{k=1}^{9} \rho_k^e D_{ik} \log(c_i) + \gamma^e X_i + \epsilon_i$$

where $y_i$ is income for respondent or individual $i$, $c_i$ consumption per capita, $D_{ik}$ (with $k$ = 1,…,9) a set of dummies for deciles in the consumption distribution, which allow us to capture the different saving rates at different levels of the distribution through $\rho_k^e$. Independent variables $X_i$ include sex, age, education, demographic information (household size, number of other working adults in the household), area of residence (rural or urban), region, job characteristics (occupation, industry, tenure, sector, working hours; for employees also firm size, existence of a written employment contract and union membership) and dwelling characteristics (cooking fuel, floor material and toilet type). In 2020, we add a quarter dummy. Separate models are estimated for each employment status $e$: employees and the self-employed. These models are selected to include characteristics for which partial $R^2$ is non-negligible in most years, kept constant over time, to the extent that it is feasible,[4] to avoid introducing spurious distributional differences across periods.

We estimate these income models with ordinary least squares on the set of observations $\mathcal{R}$ with reliable income data and use them to predict incomes for $\mathcal{U}$. The set including the rest of

[4] Not all characteristics are always available. There is no information on occupation and industry for 2015/16 and no consumption expenditure data for 2019, which results in only slightly lower predictive power. We use 2022 data to check that omitting consumption does not lead to markedly different income distributions.

self-employed workers and paid employees outside the household for which income information is unreliable or not directly available.[5] Our models have substantially more predictive power for paid employees than for self-employed workers, with adjusted $R^2$ ranging from 0.50 to 0.67 and from 0.18 to 0.34, respectively. Individuals with unreliable income data represent on average about 7% of the paid employees and 23% of the self-employed, with some variation across years (see Table A.3).

Assuming log-normality, we further take a draw from $N\left(0, \sigma^2_{\mathcal{R},e}\right)$, where $\sigma^2_{\mathcal{R},e}$ is the variance estimated for income residuals in $\mathcal{R}$ for employment status $e$. This last step adds random noise that is uninformative for each household, but it is necessary to account for the variance not explained by observable characteristics and thereby adequately reflect the dispersion of the income distribution.

The validity of this procedure depends on three assumptions:

(A1) $E_{\mathcal{R},e}[\log(y_i) \mid X_i, c_i] = E_{\mathcal{U},e}[\log(y_i) \mid X_i, c_i]$

(A2) $Var_{\mathcal{R},e}(\epsilon_i) = Var_{\mathcal{U},e}(\epsilon_i)$

(A3) $\epsilon_i^{\mathcal{U},e} \sim N\left(0, \sigma_\epsilon^{2\,\mathcal{R},e}\right)$

Essentially, these assumptions require that the set of individuals for which income is reliably captured in KCHS is representative of that type of workers overall both in terms of the conditional correlation between predictors and income, as well as in the dispersion of the unexplained component of income. These assumptions are still relatively strong and might not be fully satisfied, as workers with a reliable income measure differ in their observable characteristics. They are, for illustration, more likely to be paid monthly rather than daily. It is conceivable that, e.g., the association of education with income, or the extent of wage dispersion not explained by observables, might differ across these types of workers, and hence between $\mathcal{R}$ and $\mathcal{U}$.

Notwithstanding its limitations, this procedure is likely to provide a more reasonable approximation than the most straightforward alternatives, such as direct replacement by per

[5] We consider this to be the case when income information is missing or monthly income is implausibly low, i.e. below KSh 1,500, the earnings of a casual agricultural worker for 15 days of work at the lowest daily rate regularly observed in the data. Predicted incomes are bottom-coded at KSh 1,500 per month. We do not venture a prediction of incomes for contributing family workers, apprentices, etc. because zero or very small incomes are more plausible for them. In addition, KCHS 2021 and 2022 do not aim to collect income information from them, and it is likely that the relation to income of the variables used for prediction is different to that for paid employees or the self-employed.

capita consumption. It is well known that consumption distributions are more compressed than income distributions for various reasons, so that this procedure would notably underestimate income inequality and, in particular, underestimate the share of workers with incomes above the tax-exemption threshold. Another simple alternative, directly extrapolating the distribution of income within $\mathcal{R}$ to the whole workforce, does not seem convincing given the easily verifiable differences in observables between $\mathcal{R}$ and $\mathcal{U}$ mentioned above, reflected in clearly visible differences in the distribution of predicted and observed incomes within a given employment category, which our method can absorb.[6]

Sensitivity checks show that variations in defining the set of reliable observations, such as imputing very small income observations instead of treating them as daily incomes, or in the imputation model employed, such as having a joint model for employees and the self-employed with dummies for employment type, have a rather limited impact on the resulting income distribution.

## 3.2 Merging household surveys and administrative tax records

We consider several options for integrating household survey and tax data. The key distinction is whether, following canonical approaches, we only use tax data to correct the top of the distribution, or instead we use it to fully or partly replace the distribution of the formal sector in the household survey by the distribution observed in tax data.

Since we define formality as registration with tax authorities, the distribution of formal incomes provided by administrative data is correct by definition, or at least to the extent that reported incomes correspond to actual annual incomes. It is thus intuitively appealing to rely on administrative data to capture the formal sector, while survey data are considered informative about the informal economy. To combine both data sources, we characterize formal workers in the household survey as those individuals who report paid employment outside the household as their primary income and respond affirmatively to the question of whether their employer deducts income tax from their earnings. Since this information is not available in KIHBS 2015/16, this approach is limited to the KCHS survey rounds. In practice, this approach, which we refer to as Method 1A, can be implemented simply by merging tax

[6] Assumption 2 could be eventually relaxed to a conditional version by allowing for and modelling heteroskedasticity. Similarly, the normality assumption is taken for simplicity, and other distributions (including the empirical distribution of residuals) could be easily employed. We consider this procedure sufficiently good as a first-order approximation.

data for formal employees with survey data (rescaled to annual incomes) for informal employees and the self-employed, using the appropriate sampling weights.[7] Throughout the paper, we refer to these two groups jointly as informal workers, even though some self-employed individuals file tax returns.

It is unclear to what extent this survey question provides a valid link between tax and survey data for lower-income workers whose earnings are below the tax-exemption threshold implicitly defined by the Kenyan PIT structure.[8] Since these workers are not supposed to pay income tax, we might expect them to respond negatively to the formality survey question, but to still feature in PAYE data, as companies are expected to report incomes for all workers. In fact, even if there is a non-negligible number of survey formal workers below the threshold,[9] a heavier left tail is visible in tax data for all survey years (see Figure 1). Other possible reasons behind this tail that would recommend disregarding tax data for these income levels are observations that do not correspond to yearly incomes (i.e., short spells of formal-sector employment) and underreporting to tax authorities. Therefore, we consider Method 1B, a variation of the approach above that only relies on tax data to capture incomes of formal workers above the implicit tax-exemption threshold.[10] Household surveys thus provide information on both informal workers and formal workers earning incomes below that threshold.

When implementing these adjustments, we exclude observations from other tax returns that cannot be matched to PAYE returns, as they do not seem to correspond to formal workers and could be redundant with the informal workers present in the survey.[11] We explore the impact of this choice below.

[7] In general, when we discuss frequencies or number of observations of a given type in one of the surveys, the use of appropriate sampling weights is implied.

[8] All Kenyan taxpayers are entitled to personal relief, a fixed amount that is deducted from their personal tax liability. This relief defines thus an income level up to which no income tax is due in practice, provided that personal relief is applied. In this paper, we refer to this level as (implicit) tax-exemption threshold.

[9] This could be explained by lack of information or a different interpretation of the question. We also observe that income tax is deducted in PAYE data in a few cases for which that would not be necessary if employers correctly applied personal relief.

[10] For the sake of comparability, we take the highest of these thresholds over the study period (KSh 288,000, corresponding to years 2021 and 2022) for all years.

[11] In principle, it would be possible to add the upper tail of these observations to the distribution of informal workers. However, note that in a high informality context our tax data are more likely to add some high-income earners rather than to represent the informal upper tail more accurately overall.

We also implement an adjustment more in line with standard methodologies (Method 2), where only the upper tail of the survey is replaced by tax data. The usual assumption is that each of the two data sources represents the income distribution accurately on either side of a merging point (Blanchet, Flores, and Morgan, 2022). Rather than looking at the crossing of densities, we choose a merging point beyond which the *number* of taxpayers (see Figure 2) exceeds (or tends to exceed) the number of individuals reporting those incomes in the survey, as this seems more indicative of the survey failing to include some high-income earners.

In all cases, a reweighting step is necessary to account for the fact that, by underestimating the mass of formal workers or top income earners, surveys assign excessive weight to other observations. Let $\phi^T$ denote the number of income earners drawn from tax data, $\phi^S$ the number of survey individuals they replace, and $\pi^S$ the total survey population, including all formal and informal workers. In Method 1, $\phi^T$ and $\phi^S$ refer to formal workers in the relevant range of the income distribution; in Method 2, to workers above the merging point. These counts are reported in Table 1. We reweight the remaining survey observations by a factor $\kappa$ chosen so that the combined population equals the survey total:

$$\kappa = 1 - \frac{\phi^T - \phi^S}{\pi^S - \phi^S}.$$

For Method 2, this is exactly the reweighting step suggested by Blanchet, Flores, and Morgan (2022); and for Method 1, it follows the same logic.

**Table 1: Workers by type, year and data source**

| | 2016 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|
| Formal workers: Tax | 2,211,066 | 2,484,347 | 2,479,464 | 2,546,133 | 2,773,579 |
| Formal workers: Survey | - | 2,286,198 | 1,477,301 | 1,708,507 | 1,691,477 |
| Taxpayers*: Tax | 1,298,636 | 1,582,333 | 1,543,119 | 1,644,323 | 1,761,546 |
| Taxpayers*: Survey | - | 1,282,271 | 890,832 | 1,213,633 | 1,166,464 |
| Top incomes: Tax | 249,189 | 738,010 | 749,610 | 290,203 | 317,718 |
| Top incomes: Survey | 180,119 | 512,402 | 405,179 | 270,165 | 205,792 |
| Informal workers | - | 11,448,768 | 11,428,429 | 11,981,254 | 9,924,035 |

Notes: This table indicates the number of individuals in various categories and data sources. Formality means tax formality and is self-reported in the survey. "Taxpayers" refers to formal workers with incomes above the implicit 2021–22 tax-exemption threshold, and "Top incomes" to those above the merging points depicted in Figure 2. Survey totals computed using sampling weights.

The implicit assumption behind all these methods is that formal-sector incomes are not accurately measured in the survey, and particularly so in upper segments of the distribution.[12]

[12] Method 2 incorporates top incomes from sources other than wage employment. In practice, an overwhelming majority of top incomes is found in PAYE.

A strong indication of this fact for the case of Kenya is that the household survey fails to capture the significant increase in the number of tax filers and taxpayers observed during the study period (see Table 1). The density plots in Figure 1 (left-hand column) show that the income distribution of formal workers differs across data sources. The frequency plots in Figure 1 (right-hand column) show that in 2019 we only find more formal workers in tax data at a given point of the income distribution, while for 2020, 2021 and 2022 tax data include more formal workers over the whole income distribution, due to the severe underestimation of the number of formal workers in the survey.

The question of which method is more appropriate can then be posed in terms of why this discrepancy between tax and survey data arises, or more precisely, how it is related to the survey income distribution for informal workers. Method 1 assumes that missing incomes in the formality variable are unrelated to the income distribution among informal workers. This distribution is thus assumed to be correctly measured in the household survey, even if the share of these workers in the population is overestimated.[13] This could be the case if the survey fails to capture these formal incomes altogether, due to, e.g., differential non-response not adequately corrected for.

Alternatively, one may suspect that the reason behind 'missing' formal employees is less related to obtaining responses from those formal employees than to their classification into that category in the household survey. If this is the case, adjustments as in Method 1 might spuriously increase the density of the distribution in segments where this mislabelling is particularly pronounced. This can be avoided through an adjustment following Method 2 that only relies on tax data at income levels where it includes more individuals than implied by the survey for all employment categories. Of course, this has the drawback of omitting any individuals not present in tax data in the higher-income segment measured with tax data (see Figure 2). It also leads to potentially misrepresenting (depending on survey accuracy) the distribution of formality in the income range covered by survey data, which is, however, not necessarily problematic if researchers are only interested in aggregate indicators for the whole population.

[13] This is the assumption behind proportional reweighting. We also assume that this excess mass is restricted to informal workers and does not affect e.g. contributing family workers.

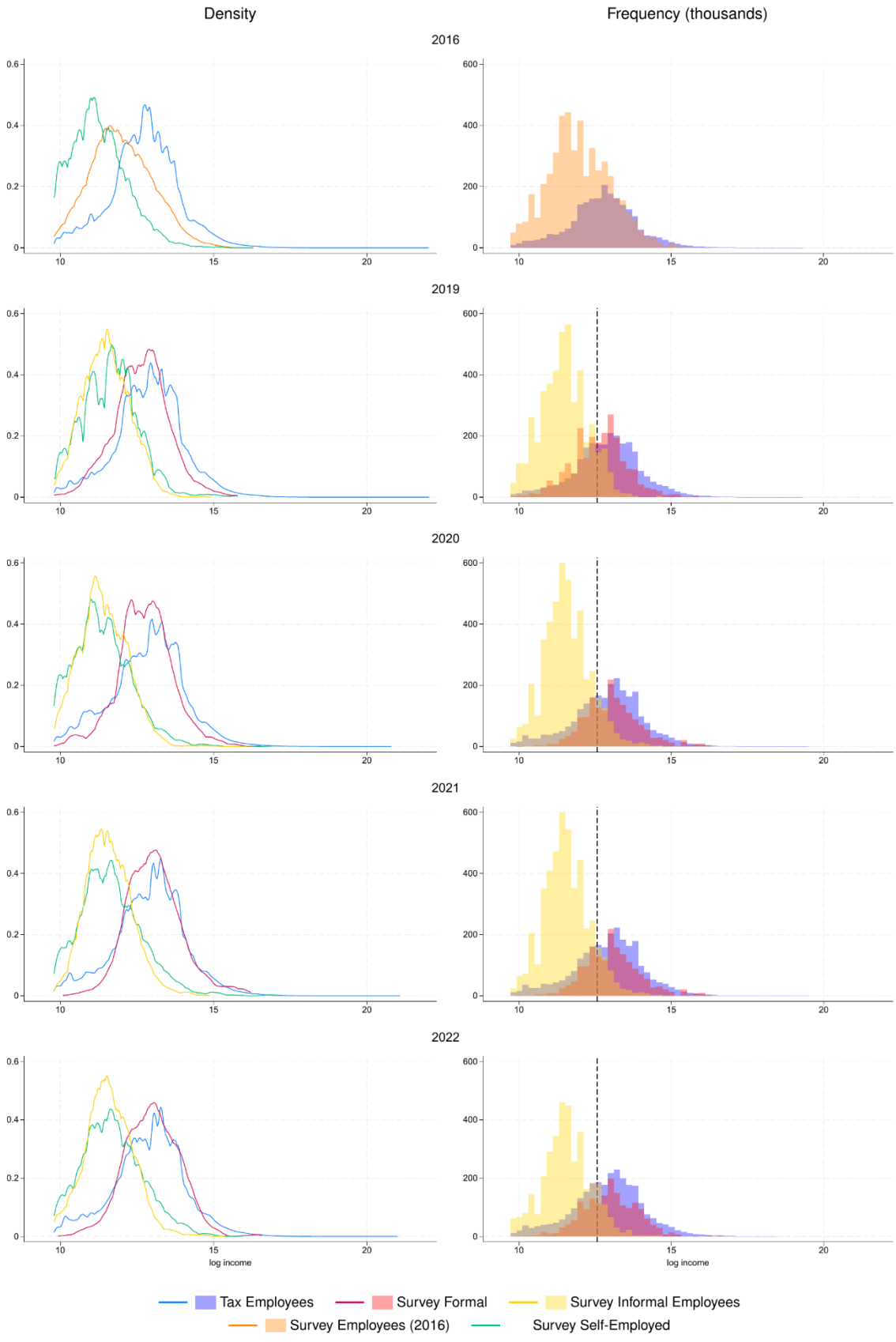



**Figure 1: Densities and frequencies by employment and data source**

Notes: Log annual income on the x-axis. All survey densities and frequencies computed using sampling weights. The dashed line in the right-hand column represents the merging point for the formal-sector distribution in Method 1B, which corresponds to the implicit tax-exemption threshold in 2021–22.

It is not obvious which of these two forms of mismeasurement prevails. On the one hand, the substantial gaps between the number of formal employees found in survey and tax data in the frequency plots of Figure 1 dissuade us from excessively trusting the formality indicator in the survey. On the other hand, Figure 1 also shows that most of the excess mass of formal workers beyond the tax-exemption threshold cannot be simply attributed to the formality question and, if resulting from classification errors, would arise from inaccurate measurement of the primary economic activity, which is perhaps less prone to errors as it is more salient to respondents and central to the survey.

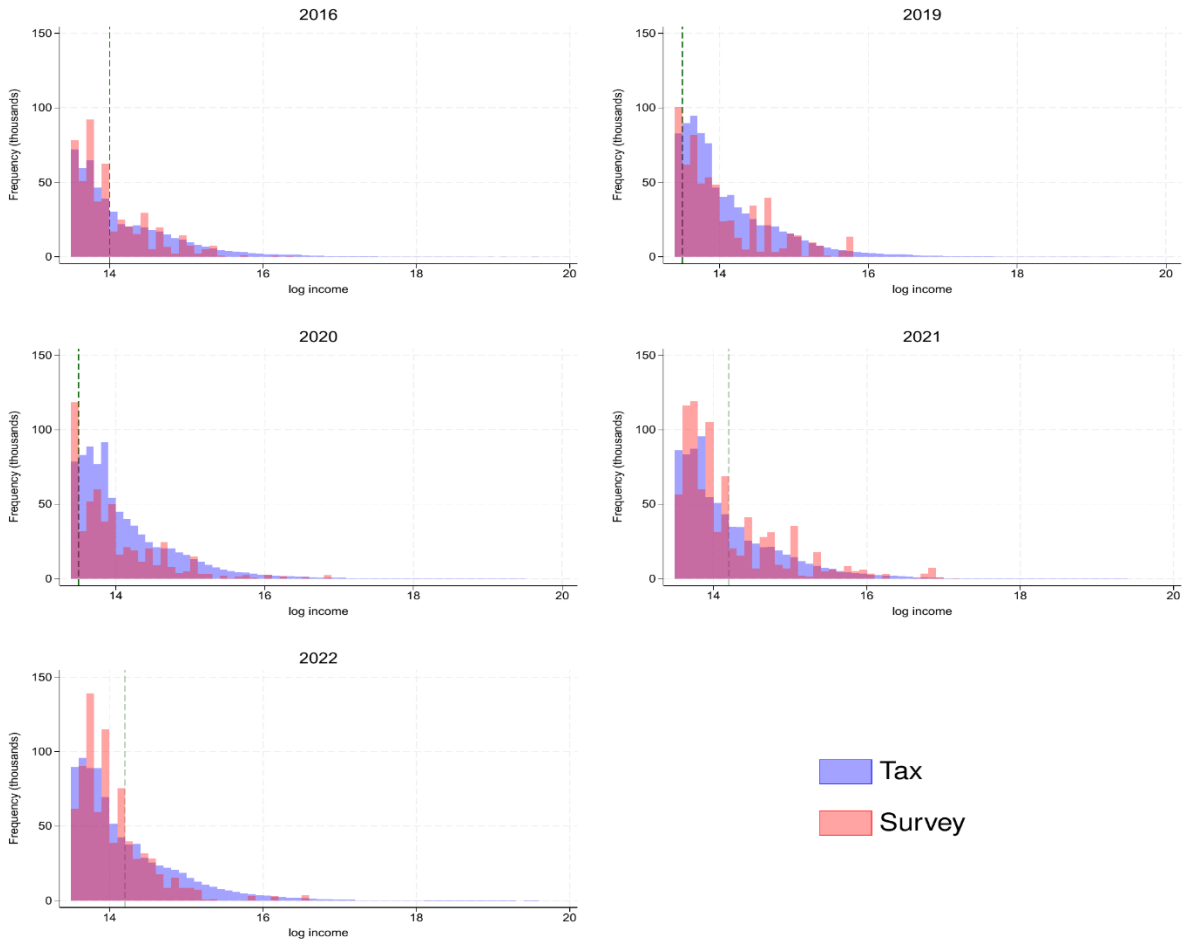



**Figure 2: Right tail of the income distribution by data source**

Notes: The figure shows the upper tail of the Kenyan income distribution in survey and tax data. Bars indicate the number of observations in the respective data sources at a given income level. The green dashed line denotes the income level selected to implement Method 2.

# 4 Results

## 4.1 Inequality levels and trends: pre-tax income

Table 2 displays the implications of data combination for the measurement of income inequality in Kenya. Gini coefficients are already remarkably high in each of the two data sources when considered separately, but still underestimated with respect to combined data, which yields values between 0.62 and 0.67. Top income shares are typically severely underestimated, except for 2021, which includes some observations with large weights and very high incomes (see Figure 2). This dependence on a few specific observations comes from the sparsity of the right tail of the income distribution and implies very noisy survey-based estimates of top income shares. Tax data on their own show higher inequality than survey data (except in 2021) but still underestimate top 1% income shares by up to about 4.6 percentage points. Using data combination methods, we estimate the top 1% and top 5% income shares at levels above 20% and 40%, respectively.

**Table 2: Pre-tax income inequality**

| Year | Survey | Tax | Method 1A | Method 1B | Method 2 |
|---|---|---|---|---|---|
| *Panel A: Gini coefficient* | | | | | |
| 2016 | 0.582 | 0.615 | - | - | 0.642 |
| 2019 | 0.553 | 0.607 | 0.634 | 0.632 | 0.616 |
| 2020 | 0.586 | 0.623 | 0.674 | 0.674 | 0.666 |
| 2021 | 0.632 | 0.613 | 0.663 | 0.663 | 0.634 |
| 2022 | 0.580 | 0.621 | 0.650 | 0.649 | 0.652 |
| *Panel B: Top 1% share* | | | | | |
| 2016 | 0.137 | 0.211 | - | - | 0.238 |
| 2019 | 0.149 | 0.199 | 0.220 | 0.220 | 0.225 |
| 2020 | 0.172 | 0.206 | 0.244 | 0.244 | 0.252 |
| 2021 | 0.219 | 0.197 | 0.242 | 0.242 | 0.225 |
| 2022 | 0.131 | 0.204 | 0.213 | 0.212 | 0.236 |
| *Panel C: Top 5% share* | | | | | |
| 2016 | 0.331 | 0.398 | - | - | 0.425 |
| 2019 | 0.323 | 0.382 | 0.420 | 0.420 | 0.407 |
| 2020 | 0.344 | 0.390 | 0.452 | 0.452 | 0.456 |
| 2021 | 0.414 | 0.383 | 0.441 | 0.442 | 0.416 |
| 2022 | 0.322 | 0.393 | 0.412 | 0.410 | 0.434 |

Notes: Authors' calculations based on KIHBS and KCHS household surveys and administrative tax data from KRA. See Section 3.2 for a description of the estimation methods. Method 1B uses KSh 288,000 (the implicit tax-exemption threshold in 2021 and 2022) as a starting point for tax data in all years.

Regarding differences across data combination methods, their levels are (as expected) most similar for measures focused at the very top. The variations of Method 1 generally lead to

higher Gini coefficients. Relying on tax data to represent formal workers also below the implicit tax-exemption threshold has no perceptible effects. For all data combination methods, inequality indicators peak in 2020, which they rarely do when data sources are separately considered.[14] The observed patterns are robust to using a consistent merging point for all years when implementing Method 2 (see Table A.4).

Different methods do yield different pairwise comparisons of years, e.g., inequality indicators computed taking the formal-sector distribution from tax data show an inverse U-shaped pattern, while income seems to have become increasingly concentrated after 2021 when only the upper tail is corrected with tax data, perhaps because Method 2 excludes the informal outliers in 2021 mentioned above.[15] This difference across methods between 2021 and 2022 persists when adjusting the informal sector in KCHS 2022 for its anomalously small size, which reduces Gini coefficients by half to one point and top 1% shares by 0.5 percentage points (see Table A.6).

Table A.7 documents the sensitivity of these inequality estimates to different choices in the tax income under consideration. Panel A shows that inequality estimates for 2022 are gradually reduced as income is restricted to that under the PIT (i.e., excluding final WHT income and MRI income), wage employment income, and wage employment income reported in PAYE. By contrast, treating individuals not in PAYE data as if they were formal workers in Method 1B increases inequality estimates. Overall, estimates vary by up to 1.4 Gini points and 2 percentage points for top 1% shares. The latter is much more significant in relative terms, underscoring the disproportionate importance of non-employment income for top earners. Panel B shows that time trends are comparable across these variations.

## 4.2 Inequality levels and trends: post-tax income

An obvious advantage of our simple data combination method is that it preserves the distinction between formal and informal workers, allowing for a realistic and straightforward analysis of the impact of PITs on inequality. We conduct this analysis on the basis of income

[14] Inequality in the survey component of the data may be overstated since we observe monthly incomes reported at different points of time. We are thus likely to overestimate annual incomes of individuals observed at the height of the Covid-19 pandemic, and vice versa for those interviewed in other quarters. However, survey-based inequality in 2020 is not markedly different from other years.

[15] In fact, top income shares are higher in 2021 also under Method 2 when selecting *ad hoc* a merging point that is high enough to keep the high-income outliers in KCHS 2021 (see Table A.5).

distributions analogous to those constructed with Method 1B above.[16] Once surveys and tax data have been combined into a single distribution where the formally employed can be identified, we can assess the impact of taxes on the income distribution by simply applying statutory tax rates to individuals in formal employment.[17]

Table 3 shows our results on the impact of PITs on the income distribution, reporting post-tax interpersonal inequality and the difference to pre-tax income inequality. Combined data indicate that the personal income tax reduces the Gini coefficient by 3 to 4 points, and hence by between 4% and 6% of its pre-tax levels. The relative impact on top income shares is much larger, as they are typically reduced by over 10%. It makes sense that top-sensitive measures react more strongly to personal income taxes when considering that the upper segments of the income distribution generate the greater part of tax revenues.

Table 3 also shows that combining tax and survey data can substantially alter the conclusions on the overall impact of PIT on inequality. Survey data downplay the reduction of the Gini coefficient caused by taxes, as they underestimate high-income taxpayers, and hence also the average tax rates they face. In contrast, tax data overestimate this effect by ignoring the large share of informal income earners that are unaffected by tax rate changes. Due to their difficulties in capturing the top of the distribution, surveys produce noisy and too small estimates of the impact of taxes on top 1% shares. In tax data, this impact is more stable and larger, but still below the effects measured through data combination – in this case, probably because tax data overestimate average incomes.

[16] Here, we use the relevant personal relief for 2019 as a cut-off for using tax data. This implies that we use tax data for a somewhat larger portion of the formal-sector distribution. For simplicity, we apply the PIT structure approved in April 2020 to the full year.

[17] For simplicity and consistency with Section 4.1, in this table we directly apply tax rates to the market income variable analyzed above. This abstracts from tax benefits such as those linked to mortgage payments, insurance and pension contributions, and from statutory contributions. These are considered in the progressivity analysis below.

**Table 3: Post-tax income inequality**

| Year | Survey | | Tax | | Combined | |
|---|---|---|---|---|---|---|
| | Net | Δ | Net | Δ | Net | Δ |
| *Panel A: Gini coefficient* | | | | | | |
| 2019 | 0.530 | −0.023 | 0.563 | −0.044 | 0.595 | −0.037 |
| 2020 | 0.571 | −0.015 | 0.584 | −0.038 | 0.646 | −0.028 |
| 2021 | 0.609 | −0.023 | 0.565 | −0.047 | 0.633 | −0.030 |
| 2022 | 0.555 | −0.025 | 0.574 | −0.047 | 0.611 | −0.039 |
| *Panel B: Top 1% share* | | | | | | |
| 2019 | 0.144 | −0.005 | 0.181 | −0.018 | 0.193 | −0.027 |
| 2020 | 0.161 | −0.011 | 0.186 | −0.019 | 0.216 | −0.027 |
| 2021 | 0.210 | −0.010 | 0.178 | −0.019 | 0.221 | −0.021 |
| 2022 | 0.117 | −0.014 | 0.184 | −0.020 | 0.178 | −0.034 |
| *Panel C: Top 5% share* | | | | | | |
| 2019 | 0.275 | −0.048 | 0.350 | −0.032 | 0.374 | −0.046 |
| 2020 | 0.335 | −0.009 | 0.358 | −0.032 | 0.412 | −0.040 |
| 2021 | 0.385 | −0.029 | 0.349 | −0.034 | 0.406 | −0.035 |
| 2022 | 0.297 | −0.026 | 0.359 | −0.035 | 0.360 | −0.050 |

Notes: "Net" refers to post-tax values after applying the corresponding personal income tax structure to formal workers' pre-tax income. Δ indicates the change relative to pre-tax indicators. Combined data are based on Method 1B above (see Section 3.2), but using the contemporary tax-exemption threshold for 2019.

## 4.3 Discussion

While certain caveats may exist around the scope and quality of the household survey income data and the nature of the tax data we use to correct these surveys, our data combination exercise reveals very high levels of interpersonal market income inequality.

Remarkably, we find substantial effects of data combination on inequality estimates, and in particular top income shares, even though we essentially rely on employment income, as we observe relatively little capital income accruing directly to individuals. Capital income has been at the centre of recent debates about inequality and has usually been singled out as a key source of discrepancy between survey-based inequality estimates and those based on other sources (Burdín et al., 2022; De Rosa, Flores, and Morgan, 2024; Flores, 2021). The 'missing rich' in Kenyan household surveys seem thus to also include wage employees. A possible explanation is that a substantial share of Kenyan top earners still derives an important share of their income from employment.[18] In addition, the comparison of formal-sector distributions suggests that differential non-response extends well beyond the top centiles of the income distribution.

[18] This is based both on data inspection (see Section 5.4) and on personal communication with KRA officials specialized in the taxation of High-Net-Worth Individuals.

Our focus on the interpersonal distribution of market income entails several limitations from the point of view of inequality measurement, mostly imposed by data availability and comparability considerations. One of these limitations is ignoring the household structure, the unit across which resources are typically pooled. Another limitation is the exclusion of contributing family workers, which might understate the degree of inequality in the economically active population and may also lead to undesirable properties for inequality measures.[19] However, one can argue that these are exactly the right choices from the point of view of tax policy. First, PIT is levied at the individual level in Kenya. Second, the relevant population that authorities may ever succeed in taxing consists of those deriving their income from market activity and not within the household.

A further limitation of our exercise is that our information on property income is limited to that directly accruing to individuals and visible to tax authorities, a general concern in the inequality measurement literature (Alstadsæter, Johannesen, and Zucman, 2019) that might be particularly relevant in a high-informality context. We also neglect social and informal transfers, sources of income that are likely particularly relevant at the bottom of the distribution.

Some of the limitations of this paper might be best addressed simultaneously. It should be possible to adjust the methods explored in this paper to perform a rank-preserving replacement of incomes for the upper tail of (formal) workers in the survey with those observed in tax data, allowing for household-level inequality estimates that incorporate tax data. The impact of uncertainty around the incomes of contributing family workers might be diluted when adding those to the incomes of other household members. In addition, KIHBS 2015/16 provides household-level information on other relevant income sources, which tend to be concentrated at the extremes of the distribution and are thus important to inequality, such as transfers, that may enable us to extend this exercise to a broader income concept. Along these lines, allocating retained profits to individuals, as e.g. in Distributional National Accounts (Piketty, Saez, and Zucman, 2018), may also allow us to assess the importance of capital income more comprehensively.

[19] For instance, the incorporation of a contributing family worker to a low-paid job would be likely to increase inequality with our current measures. However, the impact of this exclusion might be limited. In 2019, when income data for most contributing family workers are collected, the Gini coefficient in the survey only increases by one point when incorporating them into the calculations.

Notwithstanding these limitations in our income inequality measures, we hope to have illustrated that data combination methods can be helpful in addressing some of the limitations suffered by individual sources in data-scarce environments. However, the compilation of reliable income inequality estimates requires an intensification of efforts to collect high-quality data on a broad definition of income with increased frequency. The very high levels of income inequality found in this study underscore that this is not an issue that can be deemed irrelevant for developing countries. We now move on to the analysis of the progressivity of the Kenyan PIT structure.

# 5 Progressivity of the PIT structure in Kenya

In this section, we draw on the data combination procedures for inequality measurement described above and standard methods in structural and effective progressivity analysis to present a rigorous and comprehensive analysis of the Kenyan PIT structure. A progressive tax system is one where the average tax rate (ATR), the total tax paid as a share of income, rises with income. Progressivity is widely held as a desirable property for PIT structures, as it ensures that wealthier individuals contribute more to public revenues. In fact, the progressive revision of bands in the PIT structure is part of Kenya's Medium Term Revenue Strategy.

Structural progressivity measures, such as average rate or liability progression, or residual progressivity, evaluate the structure of statutory tax rates and can be directly computed based on the PIT schedule. In turn, effective progressivity measures evaluate the actual distribution of tax burden across different income levels and the success of a progressive tax structure in achieving one of its ultimate goals, which is reducing inequality. This depends not only on the tax structure itself but also on the underlying income distribution. For instance, a very high-top income rate will be more effective in reducing income inequality (and, potentially, raising revenue) when many citizens enjoy income levels included in that bracket. Another important consideration in the Kenyan context for determining the actual impact of a progressive tax schedule is how formal and informal workers are located along the income distribution. Therefore, the data combination exercise described above is needed to provide an accurate assessment.

In our analysis, we aim to understand the redistributive impact of PIT among formal workers, who constitute most of the PIT tax base, but also as a whole, taking into consideration the large number of predominantly low-income workers outside the tax net. Both concerns are

normatively relevant, as PIT should reduce societal inequality, but tax burdens should also be equitably distributed among those bearing them.

## 5.1 Personal income taxes and inequality

The Kenyan PIT structure (see Figure 3) is based on graduated tax bands with marginal tax rates mildly increasing in income, a common characteristic of progressive tax systems. An additional progressive characteristic is the exemption of low-income earners from paying PIT in practice through personal relief. Figure 3 below illustrates marginal tax rates between 2019 and 2023. After a long period of stability where only inflation adjustments were implemented, there were three subsequent reforms between 2019 and 2023. In 2020, policymakers increased personal relief by over 70% and generally lowered marginal tax rates at all income levels to shelter taxpayers from economic distress. Tax rates were then increased in 2021 to compensate for decreasing revenues. In July 2023, top marginal income tax rates were further increased, with annual incomes between KSh 6 and 9.6 million facing a 32.5% rate, and those above being taxed at 35%.

Beyond tax rates and income bands, mandatory contributions have been introduced in recent years. In 2019, employees paid a flat National Social Security Fund (NSSF) contribution of KSh 200 per month and a National Hospital Insurance Fund (NHIF) contribution that rose in steps with monthly pay from KSh 150 to a maximum of KSh 1,700. Under current rules in October 2026, employees pay 6% of gross pay to NSSF up to a monthly ceiling of KSh 6,480, 2.75% to the Social Health Insurance Fund (SHIF), with a minimum of KSh 300 per month, and 1.5% of gross pay as Affordable Housing Levy (AHL). All three employee contributions have been deductible from taxable income since December 2024, as opposed to only NSSF contributions in 2019. In the following analysis, we examine both the joint effects of PIT and statutory contributions on progressivity and redistribution, and those of PIT on its own, abstracting in the latter case from the deductibility of contributions.

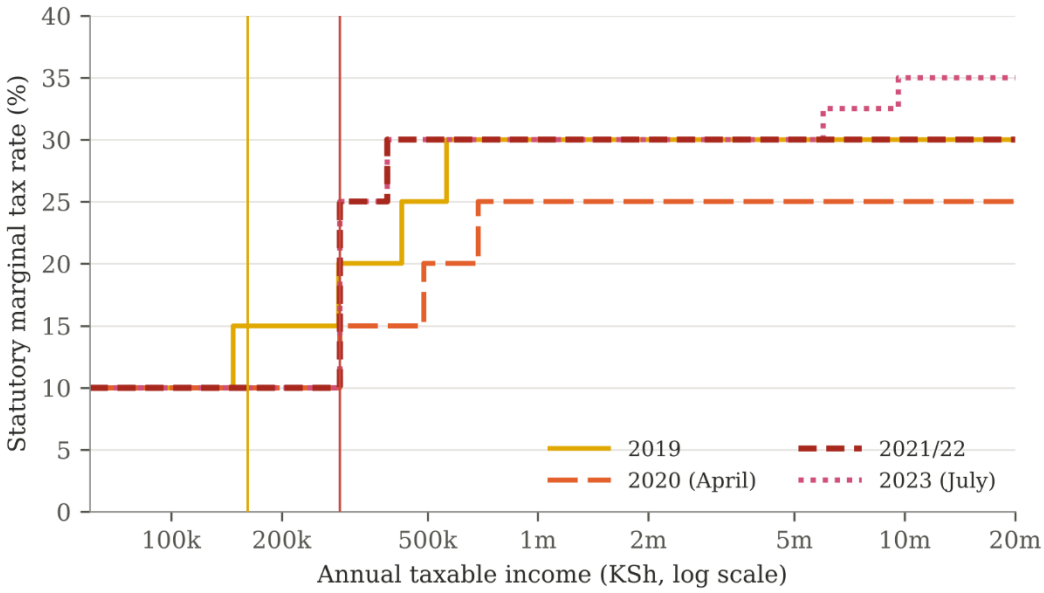



**Figure 3: Personal income tax rates in Kenya**

Notes: The figure shows the progression of marginal tax rates along nominal income between 2019 and 2023. The vertical lines represent the annual tax-exemption thresholds implicitly defined by personal relief before (gold) and following (red) the April 2020 reform.

## 5.2 Structural progressivity

This section evaluates the structural progressivity of the Kenyan PIT system using an approach consistent with the analytical framework employed in OECD tax-wedge studies (OECD, 2023). Rather than focusing solely on statutory tax rates, the analysis examines how the effective tax burden evolves as income increases. In this context, progressivity is defined as the extent to which the average tax rate rises with income. A tax system is therefore progressive when higher-income taxpayers face a larger tax burden relative to their income than lower-income taxpayers.

The structural progressivity of a PIT system is shaped by several interrelated factors. The marginal tax rate schedule determines the extent to which the tax burden rises at higher income levels, while tax-free thresholds and personal reliefs reduce the effective burden most strongly at the lower end of the distribution. In addition, statutory contributions, such as pension contributions and health insurance contributions, affect the overall tax wedge and may either strengthen or weaken progressivity depending on whether they are progressive, proportional, or capped. Progressivity, therefore, reflects the combined effect of statutory rates, relief mechanisms, and mandatory contributions rather than the tax schedule alone. For this reason, the analysis focuses on the effective ATR and the tax wedge, which provide a more comprehensive measure of the tax burden than statutory rates.

We follow a standard micro-simulation approach. To compute the tax wedge, taxable income for each gross annual income level is first derived after applying the deductions arising from statutory contributions (AHL, SHIF and NSSF).[20] The PIT liability is then calculated using the statutory tax schedule, and, after deducting personal reliefs, the effective ATR is computed as the ratio of total income tax to gross income. The tax wedge is constructed as the ratio of the combined burden of income tax and mandatory contributions to gross income. In a complementary analysis, we simply apply the statutory tax schedule directly to taxable income, thereby abstracting from mandatory contributions and their deductibility. This approach allows the analysis to distinguish the progressivity generated by the income tax system itself from the effects of statutory contributions and the associated deductions (OECD, 2023).

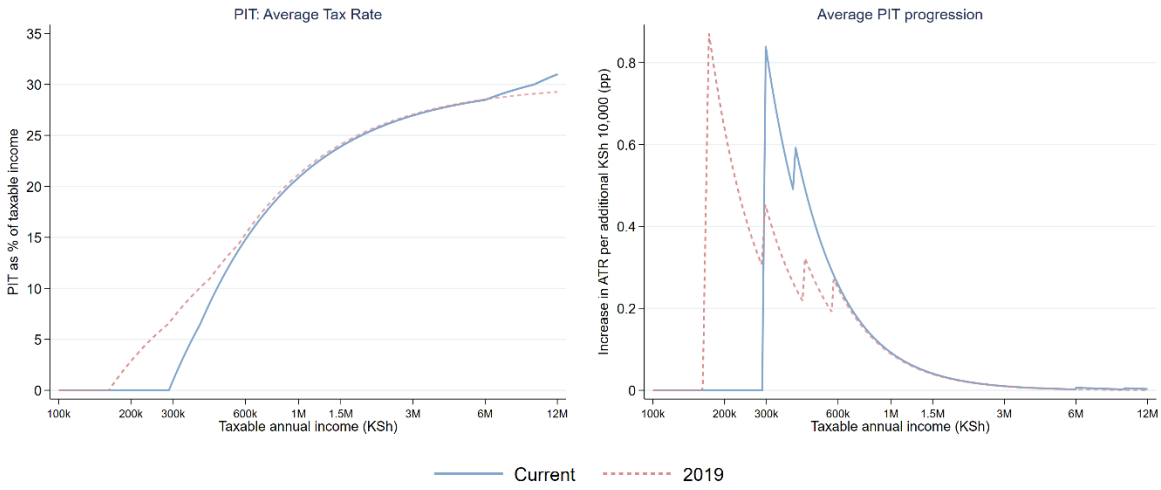



**Figure 4: Structural progressivity of PIT**

Notes: Authors' computations. See Table A.8 for the relevant PIT schedules.

The study computes the ATR and the tax wedge across a wide range of income levels and then summarizes the results at selected policy-relevant income points. This approach allows the analysis to capture both the overall shape of the tax schedule and the points at which the effective tax burden changes most significantly, while also making it possible to assess how the reforms have altered the distribution of progressivity across income groups.

[20] Employers match the NSSF and AHL contributions. In our calculation of the tax wedge, we restrict our attention to employee contributions, abstracting from these employer contributions. Including them would raise the tax wedge uniformly by 7.5% of gross pay up to the NSSF upper income limit of KSh 108,000 per month, and by a gradually declining amount eventually converging to 1.5% above it, since AHL contributions are not capped.

Figure 4 (left) presents the profile of the average PIT rate under both the current and the 2019 systems.[21] In both cases, the ATR increases monotonically with income, confirming that the Kenyan PIT remains progressive in structure, but the reforms have altered the distribution of the effective tax burden across income levels. Relative to the 2019 system, the current structure generates a lower effective tax burden at the lower end of the income distribution due to the higher effective threshold created by the revised personal relief. However, the ATR gap between both systems is essentially closed at approximately KSh 600,000 annually. This implies that, while the reforms reduced the effective tax burden for lower-income taxpayers, they also increased the rate at which the effective tax burden increases as income starts rising beyond the amount covered by personal relief. In both cases, the PIT structure generates progressivity primarily through the lower tax bands and the personal relief rather than through a sustained increase in effective tax rates at higher income levels.

This pattern becomes clearer in the right sub-figure in Figure 4, which measures structural progressivity as the change in the average tax rate (ΔATR) for each additional KSh 10,000 of income. Under the 2019 system, the increase in the ATR is more evenly distributed across the lower- and middle-income ranges, indicating a more sustained progression of effective tax rates as income rises. Under the current system, the increase in the ATR is more concentrated in the range between the first positive tax liability and the point where the gap between both tax schedules closes, at approximately KSh 600,000 annually. Beyond this level, which is below the national average wage (KSh 933,000 annually according to KNBS (2025)),[22] the increase in the ATR per additional KSh 10,000 is modest in both systems.

These results suggest that the Kenyan PIT system is structurally progressive, but that the strength of progressivity is uneven across the income distribution. Most of the increase in the effective tax burden occurs at relatively low-income levels, while the degree of progressivity weakens significantly across the middle of the formal-sector income distribution and increases only gradually at higher income levels. The current PIT structure can be described as front-loaded, since a large share of the increase in effective tax rates occurs at the lower end of the income distribution rather than being sustained across higher income levels. This makes the

[21] We present results for nominal income. Patterns are qualitatively similar when adjusting for average wage growth observed in tax data.

[22] This seems to broadly correspond to average formal-sector annual earnings in our data. It should be noted that average earnings across the whole income earner distribution, also including informal workers, are substantially lower. More generally, descriptions of income levels in this sub-section (such as “lower” or “middle” income) are with respect to the formal-sector income distribution, the relevant category for structural progressivity analysis.

effective tax schedule closer to proportional taxation at a wide range of higher incomes. This pattern indicates that the system relies heavily on reliefs and the lower tax bands to generate progressivity rather than on a consistent increase in effective tax rates across the full income distribution.

Figure 5 (left) extends the analysis to the tax wedge, comparing the average tax rate with the tax wedge across the income distribution for the current system. The tax wedge is consistently higher than the ATR at all income levels because it includes both income tax and mandatory statutory contributions. As a result, these contributions increase the total tax burden without generating a corresponding increase in progressivity. The gap between the ATR and the tax wedge therefore reflects the role of statutory contributions in raising the effective tax burden across the entire income distribution. Importantly, the difference between the tax wedge and the ATR is largest at lower- and middle-income levels, where statutory contributions represent a significant share of total deductions relative to income tax. It should be noted that statutory deductions also apply to workers whose tax liability is fully covered by personal relief. At higher income levels, the difference becomes relatively smaller in proportional terms because income tax begins to dominate the total tax burden. In addition, the cap on NSSF contributions introduces a regressive element in the system, narrowing the gap between the tax wedge and the ATR from around KSh 1.3M annually. The figure also shows that the deductibility of mandatory contributions reduces the tax wedge but does so unevenly across the income distribution.[23]

[23] The value of the deduction is increasing in the marginal tax rate and the amount of contributions deducted. As a share of income, it thus peaks at around 3% at the upper limit for NSSF contributions. Note that earners below the implicit tax-exemption threshold do not benefit from this deductibility.

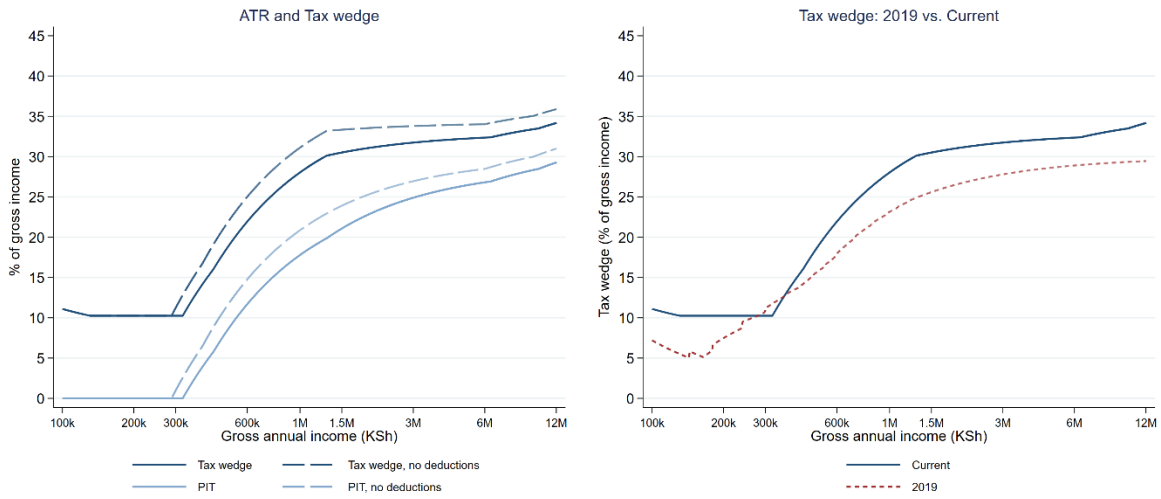



**Figure 5: ATR and tax wedge**

Notes: Authors' computations. The tax wedge includes AHL, SHIF and NSSF.

The right panel of Figure 5 shows that incorporating contributions to the analysis changes the assessment of the direction of reforms since 2019. The increase of statutory contributions dominates changes to statutory tax rates for low incomes, until a small range around KSh 300,000 for which the current tax wedge has barely changed as compared to 2019. There, the current tax wedge increases more steeply than in 2019, first due to the steeper increase in ATRs seen above, and then because proportional contributions no longer decline as a share of income until the NSSF ceiling. From that point, the current tax wedge becomes virtually flat until the levels where the top income tax rates introduced in July 2023 become applicable.

## 5.3 Effective progressivity

Section 4 illustrates how data combination methods can be helpful in addressing some of the limitations suffered by individual sources on the personal income distribution. Section 4.2 shows the usefulness of these techniques to compare the inequality of the pre-tax and the post-tax income distribution, one of the main yardsticks to evaluate the progressivity of PIT. Even though Kenya's PIT structure underwent relevant modifications in the period covered by Table 3, it would be misleading to interpret results on the inequality impact of the PIT scheme reported in that table as directly reflecting the different tax schemes in place, since differences in the estimates are also (and mainly) driven by changes in the underlying income distribution, and by sampling variance in the household surveys.

Therefore, in this section, we specifically study the impact of different tax schemes on a common income distribution to analyse how different portions of the Kenyan income tax schedule drive its redistributive impact. In Panel A of Table 4, we present the effects of different tax schedules on the distribution for all income earners, obtained using data combination Method 1B.[24] In Panel B, we show the implications of those different PIT structures on the income distribution among formal workers, directly obtained from administrative tax data for 2024.

In our simulation, we apply each tax schedule to individual data on income, insurance relief, and tax benefits, and grant the maximum allowable personal relief to all taxpayers.[25] We separately apply the current PIT structure as in October 2026 and the 2019 PIT structure to the data, and compare the effects of both tax schemes. The contrast between these two income tax schedules captures three important features of the recent evolution of PIT in Kenya, namely i) the increase of personal relief, ii) the more limited progression of tax rates for the first effective taxpayers, and iii) the moderate increase in top marginal tax rates at very high levels. We also modify these features of the current PIT structure one at a time to separately evaluate their effects (see Table A.9 for details). In a final exercise, we also quantify the redistributive impact of mandatory contributions (NSSF, SHIF, AHL) and their deductibility from taxable income. We present mechanical effects of reforms that do not incorporate behavioral reactions.

We assess the progressivity of various PIT structures, $s$, using different indicators. We complement the analysis of the Gini coefficient, which reflects developments across the whole income distribution and is often described as sensitive to changes around the middle of the distribution, with a simple top-sensitive measure, the share of income captured by the top 1% of income earners. First, we evaluate how post-tax income inequality ($I_s^{net}$) changes, reporting its level under different tax schemes ($s$). We also express the difference of those levels to post-tax inequality under the current tax structure ($I_{cur}^{net}$ )as a share of the overall inequality reduction achieved by the current tax structure:

[24] Since survey data for 2024 are not available, we use KCHS 2022. To prevent the anomalously small number of informal workers in that survey (see Table 1) from distorting our analysis, we proportionally reweight informal workers to match their average number between 2019 and 2021, the same adjustment as in Table A.6. To combine 2022 survey data with 2024 tax data, we apply the wage growth given by average income in the administrative data between 2022 and 2024, and population growth estimates derived from official numbers on size of the labour force. We use the 2019 implicit tax-exemption threshold as a merging point to adequately capture the impact of the tax structure in that year.

[25] Since we use 2024 data, the calculation of taxable income and reliefs follows 2024 rules (NSSF contributions deductible, relief on NHIF and AHL contributions). For an assessment of the deductibility of statutory contributions introduced at the end of 2024, see the last row of each panel in Table 4.

Finally, we incorporate the Kakwani index ($K_{sy}$) which quantifies the progressivity of a tax structure $s$ by comparing the distribution of tax payments $[T(s,y)]$ to the distribution of pre-tax income ($y$). Formally, it is defined as the difference between the concentration index of taxes, which is constructed by ranking individuals by pre-tax income and plotting the cumulative share of taxes paid against the cumulative share of the population, and the Gini coefficient of pre-tax income:

$$\%\Delta I_s = 100 \times \frac{I_s^{net} - I_{cur}^{net}}{I^{gross} - I_{cur}^{net}}$$

The concentration index ($C_{T(s,y)}$) captures how tax liabilities are distributed across the income distribution under tax structure $s$. A positive Kakwani index indicates that taxes are more concentrated among higher-income individuals than income itself, that is, progressivity, while a value of zero implies proportionality and a negative value, regressivity. This is illustrated in the lower sub-figures of Figure 6. We see there that for both the current and the 2019 PIT structures, the PIT liability curve is to the right of the Lorenz curve, reflecting the progressivity of the system. Consequently, higher values of the Kakwani index can be interpreted as reflecting stronger progressivity.

$$K_{sy} = C_{T(s,y)} - G(y)$$

Several patterns emerge from the analysis of the PIT structures in Table 4. First, the effect of the tax scheme in reducing inequality as measured by the Gini coefficient is more nuanced when looking at the full income distribution (3.8 points) than when focusing on formal workers (5.2 points). This can be seen by comparing the Lorenz curves in the upper sub-figure of Figure 6, as changes are somewhat more perceptible for formal workers (see also Figure A.1, which plots the gap between the pre-tax and net income Lorenz curves). On the other hand, the Kakwani index is much larger when considering all Kenyan income earners, since the bottom 70% of the income distribution does not pay any income tax (see Figure 6, bottom-left). The impact of any changes to the tax structure is much more pronounced when analyzing progressivity across formal workers than when looking at all income earners, since only the portion of workers paying income tax sees their net incomes altered by PIT reforms.

**Table 4: Effective progressivity analysis**

*Panel A: All income earners*

| | Gini | | Top 1% share | | Kakwani | |
|---|---|---|---|---|---|---|
| PIT structure | Value | %Δ | Value | %Δ | Value | %Δ |
| Gross income | 0.642 | - | 0.209 | - | - | - |
| Current | 0.604 | 0 | 0.172 | 0 | 0.300 | - |
| 2019 | 0.605 | 3.6 | 0.176 | 9.0 | 0.288 | −4.1 |
| Current, 2019 relief | 0.602 | −4.5 | 0.174 | 4.9 | 0.279 | −7.1 |
| Current, extra band | 0.604 | 2.0 | 0.170 | −6.6 | 0.304 | 1.1 |
| Current, no top tax | 0.605 | 3.4 | 0.175 | 7.3 | 0.299 | −0.5 |
| Current, contributions | 0.593 | −28.2 | 0.162 | −27.9 | 0.268 | −10.6 |
| Current, deductible contributions | 0.596 | −19.6 | 0.166 | −17.6 | 0.269 | −10.5 |

*Panel B: Formal workers*

| | Gini | | Top 1% share | | Kakwani | |
|---|---|---|---|---|---|---|
| PIT structure | Value | %Δ | Value | %Δ | Value | %Δ |
| Gross income | 0.617 | - | 0.188 | - | - | - |
| Current | 0.565 | 0 | 0.162 | 0 | 0.200 | 0 |
| 2019 | 0.573 | 14.6 | 0.169 | 24.8 | 0.170 | −15.1 |
| Current, 2019 relief | 0.573 | 16.0 | 0.167 | 17.4 | 0.146 | −26.9 |
| Current, extra band | 0.565 | −1.2 | 0.161 | −4.0 | 0.210 | 4.9 |
| Current, no top tax | 0.568 | 5.7 | 0.167 | 19.7 | 0.195 | −2.4 |
| Current, contributions | 0.570 | 10.0 | 0.168 | 22.6 | 0.116 | −41.9 |
| Current, deductible contributions | 0.574 | 16.1 | 0.168 | 22.8 | 0.118 | −41.0 |

Notes: Panel A shows results for the full income distribution of Kenyan income earners, obtained using Method 1B above based on administrative tax data for 2024 and KCHS 2022 (reweighted, see Table A.6), adjusted for wage and population growth. Panel B presents results for individuals in tax data. "Gross income" refers to pre-tax market income inequality values. "Current" denotes the PIT schedule in force in October 2026 (unchanged since July 2023); "2019" denotes the 2019 schedule. All other rows refer to net income after applying the respective PIT structure (see Tables A.8 and A.9 for detailed descriptions) to formal workers' pre-tax income, using individual data on tax deductions and reliefs for 2024. The last two rows additionally subtract NSSF, SHIF and AHL contributions at 2026 rates. "Deductible contributions" makes SHIF and AHL deductible from taxable income and removes the corresponding reliefs, as under the rules in force since December 2024. %Δ indicates the change in inequality or progressivity resulting from comparing the corresponding structure to the current structure, as a share of the current structure's effect on pre-tax inequality (cols. 2 and 4) or Kakwani index (col. 6). A positive %Δ means higher inequality under the corresponding PIT schedule than under the current one in columns 2 and 4, but higher progressivity in column 6.

Second, the reforms introduced since 2019 have globally increased the effective progressivity of the PIT structure as compared to the system in place back then, both among Kenyan income earners and specifically among taxpayers. The effects are particularly large in terms of reducing top income shares.[26] The same statements hold specifically for one element of the reforms, the increase of top marginal tax rates, which has been, however, more effective in reducing post-tax inequality than in increasing progressivity as measured by the Kakwani index.

[26] When effects are described in generic terms, they affect both formal workers and income earners overall. Exceptions are explicitly discussed.

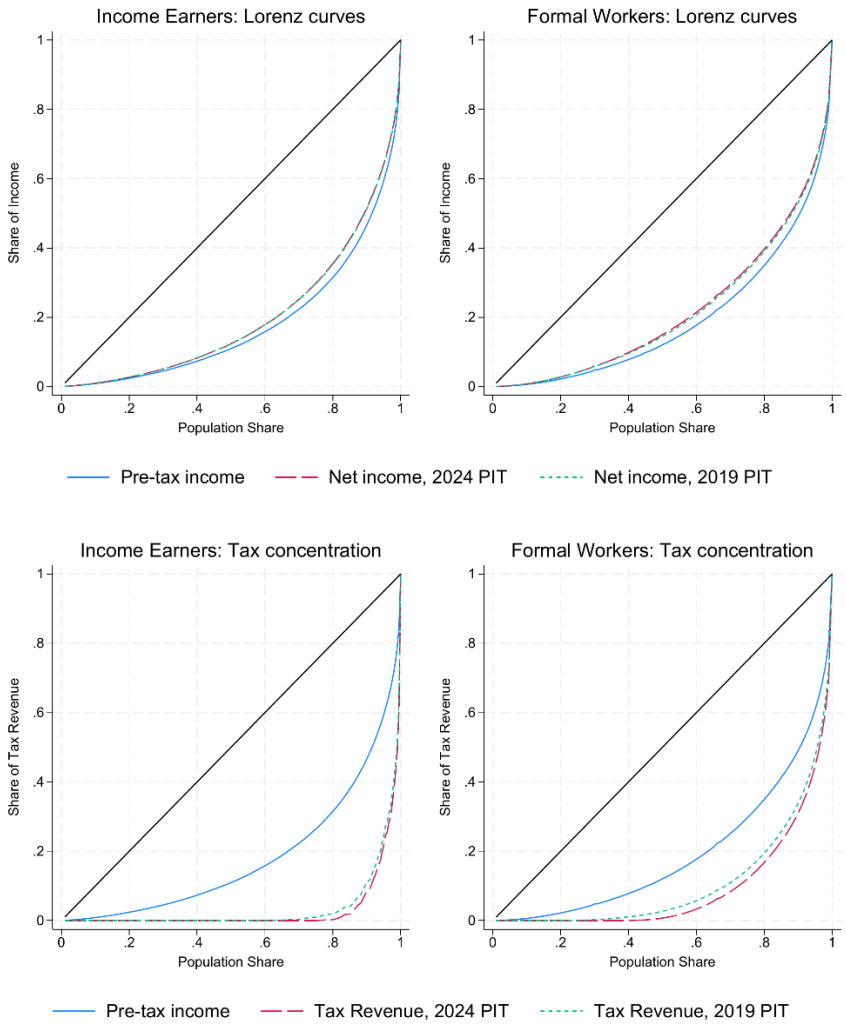



**Figure 6: Effective progressivity: 2019 vs current PIT structure**

Notes: The upper sub-figures show Lorenz curves for pre-tax and net income resulting from applying the 2019 and current PIT structures to 2024 data (referred to as "2024 PIT"). The lower sub-figures compare the pre-tax income distribution to the distribution of tax liabilities across income levels. The area between both curves is half the Kakwani index. The left sub-figures show results for the income distribution of all income earners, obtained through the data combination procedure described in Section 3 as Method 1B. The right sub-figures show results for the formal-sector income distribution.

The effects of the other two components of recent tax reforms are more subtle. Increasing personal relief has slightly reduced the overall redistributive impact of the PIT structure (as measured by the Gini) but has actually reduced inequality among taxpayers. There are two reasons for this. First, newly exempt taxpayers who see their net incomes increase are among the formal workers with lower earnings, but not among the poorest income earners when also considering informal workers. Second, increasing personal relief reduces tax liabilities and increases net income for all formal workers, widening the gap between them and informal workers, who are generally less affluent. A higher relief does reduce top-income shares, as their net incomes are barely affected, while other formal workers see their take-home pay

increase relatively more. Finally, higher personal relief unequivocally increases the Kakwani index by intensifying the concentration of the PIT burden on a smaller share of taxpayers and shifting the tax distribution to the right (see the bottom sub-figures in Figure 6).

Similarly to increasing personal relief, adding an extra 20% band to the current structure slightly increases overall inequality, as measured by the Gini, but decreases inequality among formal workers. Top income shares are reduced, as middle-class net incomes increase relatively more than those at the top. This reform increases progressivity according to the Kakwani index by reducing tax liabilities more substantially in relative terms for the first effective taxpayers and thereby increasing the share of tax revenue collected from high-income earners. Our two effective progressivity measures thus give apparently contradictory information for Kenyan income earners. This is because the Kakwani index focuses on how a given tax burden is distributed across the population, while the difference in Gini coefficients also considers the total size of that tax burden. The fact that an additional tax band with lower MTR reduces tax revenue thus implies a negative redistributive impact overall, even if the extra band increases progressivity according to the Kakwani index.

Finally, mandatory contributions cause a substantial decline of overall income inequality because formal workers paying those contributions tend to have higher incomes than those outside the tax net.[27] However, the effect of PIT on inequality among formal workers is remarkably dampened by mandatory contributions. By increasing the tax liabilities of middle-class taxpayers relatively more strongly than those of high-income earners, contributions increase top income concentration among them and reduce the Kakwani index, ultimately undermining the progressivity of the system (see Figure A.2). Deductibility barely affects the Kakwani index but noticeably limits the inequality reduction achieved by contributions by reducing the total amount of tax liabilities.[28] As in our structural progressivity analysis, the effects of increasing mandatory contributions eclipse those of changes to the PIT schedule since 2019. While this increases the scope of overall redistribution due to the increase of the tax

[27] In our simulation, we assume that only formal workers pay statutory contributions. In practice, a small proportion of informal workers pay them as well, which may dampen somewhat the effects reported here.

[28] It is perhaps counterintuitive that deductibility looks regressive in the structural progressivity analysis but not here according to the Kakwani index. This is because our structural progressivity measure focuses on tax liabilities *relative to income*, which implies that impact of deductions is determined by the MTR. Here, the decisive element is how the savings implied by the deduction compare to the individual's tax liability.

burden on formal workers, it does so at the cost of increasing inequality among taxpayers and substantially undermining progressivity.

## 5.4 Beyond personal income

In our effective progressivity analysis, we focus on reported personal income that is subject to the PIT regime described in Section 5.1. However, some of the income that is part of our market income concept throughout the paper falls under different tax regimes, namely rental income under the MRI regime, and income that is subject to final withholding tax. Broadening our perspective to income that falls outside the scope of the PIT tax bands may nuance our perspective on the progressivity of income taxation overall.

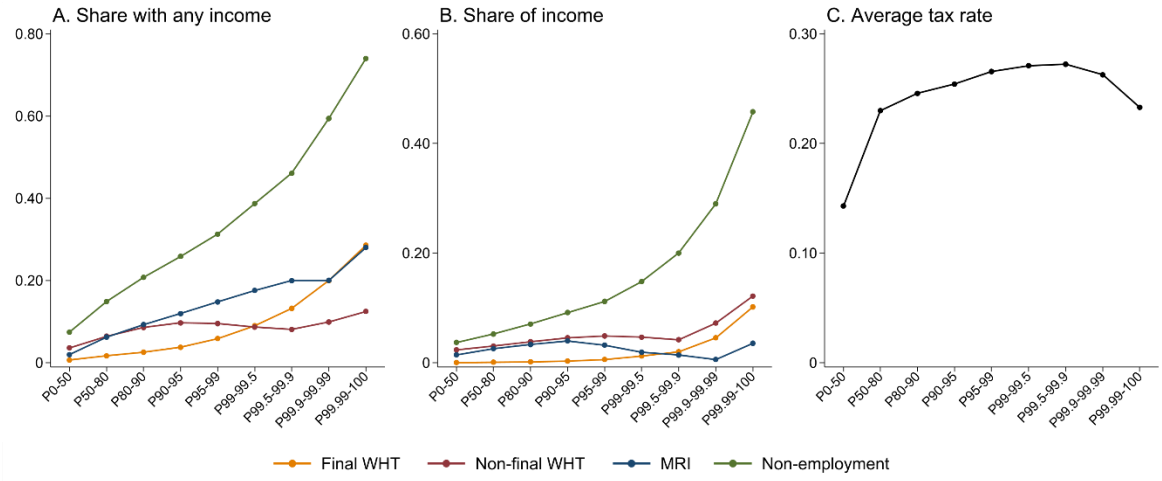



**Figure 7: Income sources along the income distribution**

Notes: Distribution percentiles computed from the combined survey and tax distribution resulting from Method 2. The figures show averages over years 2016 and 2019–2022. Non-employment income also includes that reported as such in ITRs. Tax liabilities are computed as a sum of the PIT applicable to taxable income as observed from the data, taxes on final withholding income and residential rental income, and the proportional part of non-final withholding taxes corresponding to the share of non-final withholding income not reported in the ITR.

Figure 7 illustrates the relevance of WHT and MRI income across different income levels over the study period in Section 4, focusing on the top of the distribution. While 7% of taxpayers in the bottom half of the distribution report any non-employment income, the share goes up to over 60% for the 0.1% richest individuals. The share of final withholding income rises monotonically across the income distribution, and very rapidly within the top 1%. The latter is also true for non-final withholding income; in particular, for the part not already reported in ITRs. According to our data, it is thus a very small share of taxpayers that benefits from low withholding taxes. On the other hand, residential rental income peaks at around 4% among those in the top 10% but not in the top 5%.

These substantial shares of income subject to taxes lower than top marginal PIT rates are sufficient to reverse progressivity for the top 0.1% of income earners, who face an effective tax rate 2 percentage points lower than the rest of the top 1% (see Figure 7). While this is a small share of taxpayers, it should be kept in mind that top earners account for a substantial share of tax payments. These results are likely to be a lower bound of the extent to which lower taxes on capital income undermine income tax progressivity, since a limited share of capital income can be directly attributed in our approach, which does not further allocate capital flows to firms and may also prevent us from identifying pure rentiers with no employment income in any of the years of the study period.

From a policy point of view, the challenges posed by non-final and final withholding tax income differ. When WHTs are non-final, the imperfect ability of the tax administration to link the payments to their actual recipients or to verify deductions claimed by them introduces a wedge between the tax that the policymaker intends to impose (namely PIT) and the effective rate faced by the agent, which will in many cases be the WHT rate. In those cases where income subject to WHT is likely to correspond to personal income, such as professional fees or commissions, it might make sense to close the gap between WHT and PIT rates to approximate the intended tax rate after reasonable deductions.[29] This would reduce incentives to shift income into categories subject to lower withholding rates, which increase with income under a progressive tax schedule.

When taxes are final, such as interest, dividends, or rental income under the monthly regime, it is a choice of the policymaker to tax these incomes at reduced rates. Whether this is desirable depends on a balance of normative arguments on equity and practical considerations on compliance and investment promotion. While assessing these trade-offs is outside the scope of the present study, reducing the gap between the relevant tax rates for these income types and top PIT rates would enhance the progressivity of income taxation in Kenya.

# 6 Conclusions and policy implications

The analysis of both structural and effective progressivity confirms that the Kenyan PIT system is progressive, but the intensity and distribution of progressivity vary across income levels. Structurally, the system is front-loaded: most of the increase in the effective ATR occurs

[29] Note that WHT and MRI are gross incomes, while personal income allows for deductions. It may thus be reasonable to calibrate taxes to account for this fact.

at the lower end of the income distribution, driven primarily by personal reliefs and narrow lower tax bands. Many taxpayers face a steep jump to the 30% rate soon after the exemption threshold. Progression across middle and higher formal-sector incomes is comparatively modest, producing a flatter effective tax schedule at higher levels. The limited number of tax bands prevents gradual tax increases that could better reflect taxpayers' ability to pay. This structure creates perceptions of unfairness, insufficient progressivity and a heavy burden on middle-class income earners.

This pattern is mirrored in effective progressivity measures: post-tax inequality is substantially reduced both among formal workers and income earners overall (Gini falls by 5.2 and 3.8 points, respectively), and the positive Kakwani index (0.20 for formal workers, and 0.30 for all income earners) indicates that the PIT induces meaningful redistribution. Conversely, increases in statutory contributions, though redistributive in aggregate, reduce effective progressivity because these contributions are proportional or capped. The findings indicate that the PIT system achieves meaningful redistribution by shielding lower-income earners but offers weaker incremental progressivity at levels corresponding to middle and higher formal-sector wages, highlighting both the system's achievements and the opportunities for targeted reforms to strengthen fairness and redistribution across the full spectrum of incomes.

Several policy implications emerge. **First, the results confirm a strong rationale for a progressive PIT structure.** The data combination exercise, even if mainly capturing employment income, reveals very high levels of interpersonal pre-tax income inequality (Gini coefficients consistently above 0.6, top 1% of income earners accounting for over 20% of total income), which contrasts with the notably lower figures for typically used measures based on consumption expenditure. These findings support the case for strengthening progressivity in future reforms.

**Second, personal relief remains a key driver of progressivity but also shapes the structure of the system.** The large personal relief implies that a significant share of the labor force remains outside the tax net, which is one of the main redistributive features of the current system. However, the structural progressivity analysis shows that most of the increase in the effective tax burden occurs at the point where the first positive tax liability begins. This explains why progressivity appears substantial in effective terms (as reflected in the Gini and Kakwani results), but relatively weak across the middle of the formal-sector income

distribution. In other words, the system redistributes meaningfully between the poor and those at higher income levels but becomes relatively flat afterwards.

**Third, the current structure creates scope for improving progressivity through lower-income rate design rather than further increases in relief.** Marginal tax rates at the bottom influence average tax rates across the whole distribution, so that increases in personal relief are rather costly as compared to other measures (see Kanina et al. (2025) for revenue implications). The simulations show that introducing additional lower tax bands, with modest rates below 25%, increases the Kakwani index, reduces top income shares, and strengthens redistribution among formal workers without excessively affecting overall inequality or revenue, if adequately calibrated. From a policy perspective, this suggests that lowering marginal rates for the first group of effective taxpayers and introducing more gradual progression would improve both structural and effective progressivity.

**Fourth, progressivity at higher income levels remains limited and can be strengthened.** The structural analysis shows that a large share of effective taxpayers face similar marginal tax rates, resulting in a relatively flat average tax rate beyond the lower-middle income range. From the effective progressivity standpoint, the redistributive impact of the system would also increase if progressivity were strengthened at the top. In practical terms, this means that increasing tax rates at the upper end of the current 30% band would both raise additional revenue and improve redistribution across formal workers and across the entire income distribution.

**Fifth, statutory contributions increase the overall tax burden but weaken effective progressivity.** While recent reforms to statutory contributions increase redistribution in aggregate terms, the effective progressivity results show a clear decline in the Kakwani index when these contributions are included. This reflects the fact that contributions such as NSSF are proportional or partially regressive due to minimum and maximum thresholds. From a policy perspective, this suggests the need to better align statutory contributions with the progressive structure of the income tax system.

**Finally, progressivity is also affected by the treatment of different income types.** Concerns about income shifting following the increase in top marginal PIT rates highlight the importance of aligning the taxation of labour income with the taxation of other income sources highly concentrated at the top of the distribution. Interest, dividends, residential rental income and capital gains are taxed at rates below top marginal PIT rates. The same often applies in

practice to commissions and professional fees, which are subject to non-final withholding tax. Narrowing this gap would strengthen effective progressivity while also reducing incentives for tax avoidance.

To further increase progressivity, policymakers may thus consider:

- Introducing a more gradual progression of tax rates among the first effective taxpayers with rates below 25% to balance structural progressivity and enhance redistributive impact.
- Strengthening progressivity at higher income levels by raising tax rates at the upper range of the current 30% band to generate additional revenue and improve redistributive performance.
- Reviewing statutory contributions to reduce regressivity and improve fairness among formal workers.
- Aligning taxation of other incomes with the progressive structure of the PIT system to strengthen progressivity and reduce incentives for tax avoidance.

# References


Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman. 2019. "Tax Evasion and Inequality." *American Economic Review* 109 (6):2073–2103.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. 2013. "The Top 1 Percent in International and Historical Perspective." *Journal of Economic Perspectives* 27 (3):3–20.

Assouad, Lydia. 2023. "Rethinking the Lebanese economic miracle: The extreme concentration of income and wealth in Lebanon, 2005–2014." *Journal of Development Economics* 161:103003.

Atkinson, Anthony B., and Thomas Piketty. 2007. *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*. Oxford: Oxford University Press.

Bartels, Charlotte and Maria Metzing. 2018. "An integrated approach for a top-corrected income distribution." *The Journal of Economic Inequality* 17 (2):125–143.

Blanchet, Thomas, Ignacio Flores, and Marc Morgan. 2022. "The weight of the rich: improving surveys using tax data." *The Journal of Economic Inequality* 20 (1):119–150.

Burdín, Gabriel, Mauricio De Rosa, Andrea Vigorito, and Joan Vilá. 2022. "Falling inequality and the growing capital income share: Reconciling divergent trends in survey and tax data." *World Development* 152:105783.

Chancel, Lucas, Denis Cogneau, Amory Gethin, Alix Myczkowski, and Anne-Sophie Robilliard. 2023. "Income inequality in Africa, 1990–2019: Measurement, patterns, determinants." *World Development* 163:106162.

Chancel, Lucas and Thomas Piketty. 2019. "Indian Income Inequality, 1922–2015: From British Raj to Billionaire Raj?" *Review of Income and Wealth* 65 (S1).

Chatterjee, Aroop, Léo Czajka, and Amory Gethin. 2023. “Redistribution without Inclusion? Inequality in South Africa Since the End of Apartheid.” URL https://leo-czajka.eu/wp-content/uploads/2023/10/ChatterjeeCzajkaGethin2023.pdf.

Czajka, Léo. 2017. “Income Inequality in Côte d’ Ivoire.” WID World Working Paper Series, 2017/8.

De Rosa, Mauricio, Ignacio Flores, and Marc Morgan. 2024. “More unequal or not as rich? Revisiting the Latin American exception.” *World Development* 184:106737.

Flores, Ignacio. 2021. “The capital share and income inequality: Increasing gaps between micro and macro-data.” *The Journal of Economic Inequality* 19 (4):685–706.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty. 2018. “Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA).” *Journal of Public Economics* 162:63–77.

Günther, Isabel and Andrey Launov. 2012. “Informal employment in developing countries.” *Journal of Development Economics* 97 (1):88–98.

Inchauste, G. and N. Lustig, editors. 2017. *The Distributional Impact of Taxes and Transfers: Evidence from Eight Low and Middle-Income Countries*. Washington DC: The World Bank.

Jäntti, Markus, Milly Isingoma Nalukwago, and Ronald Waiswa. 2022. “Income Distribution in Uganda Based on Tax Registers: What Do Top Incomes Say?” *WIDER Working Paper* 6/2022. Helsinki: UNU-WIDER.

Jenkins, Stephen P. 2017. “Pareto Models, Top Incomes and Recent Trends in UK Income Inequality.” *Economica* 84 (334):261–289.

Kanina, Jane, David Garcés Urzainqui, Josephine Mugure, Peter Fisker, and Jacob Nato. 2025. “Assessing the Impact of Personal Income Tax Reform in Kenya with Administrative Data: Behavioral Responses and Distributional Implications.” AERC Working Paper GOK-AERC-UCPH-DERG 005, African Economic Research Consortium, Nairobi.

Kleven, Henrik J., Martin B. Knudsen, Claus T. Kreiner, S. Pedersen, and E. Saez. 2011. “Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark.” *Econometrica* 79 (3):651–692.

KNBS. 2025. *2025 Economic Survey*. Nairobi: Kenya National Bureau of Statistics.

Lastunen, Jesse, Antoine de Mahieu, Katrin Gasior, H. Xavier Jara, and Jukka Pirttilä. 2024. “Microsimulation of tax-benefit systems in the Global South: a comparative assessment.” WIDER Working Paper 2024-35.

Lustig, Nora. 2019. “The Missing Rich in Household Surveys: Causes and Correction Approaches.” Commitment To Equity Working Paper 75.

Lustig, Nora and Andrea Vigorito. 2025. “The ‘Missing Rich’ in Household Surveys: Causes and Correction Approaches.” Tulane Economics Working Paper 2512, Tulane University.

OECD. 2023. *Taxing Wages 2023: Indexation of Labour Taxation and Benefits in OECD Countries*. OECD Publishing.

Pape, Utz Johann, and Nduati Maina Kariuki. 2019. “Technical Design of the Kenya Continuous Household Survey.” Washington, DC: World Bank. URL https://documents.worldbank.org/en/publication/documents-reports/documentdetail/995861636118553938.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. “Distributional National Accounts: Methods and Estimates for the United States.” *The Quarterly Journal of Economics* 133 (2):553–609.

Ravallion, Martin. 2016. *The Economics of Poverty. History, Measurement, and Policy*. Oxford University Press.

Zwijnenburg, Jorrit. 2019. “Unequal Distributions: EG DNA versus DINA Approach.” *AEA Papers and Proceedings* 109:296–301.

# Appendix A. Additional tables and figures and figures

**Table A.1: Withholding tax rates by type of payment**

| Category | Payment type | Examples | Resident rate | Non-resident rate | Final (residents) |
|---|---|---|---|---|---|
| *Professional and service fees* | Professional fees | Legal, accounting, audit, architectural, consulting | 5% | 20% | No |
| | Management and consultancy fees | Business or project management; advisory, technical or expert services | 5% | 20% | No |
| | Training fees | Training, seminars, workshops | 5% | 20% | No |
| | Contractual fees | Payments under service contracts | 5% | 20% | No |
| | Agency fees | Commission-based representation services | 5% | 20% | No |
| *Commissions and marketing* | Sales commission | Sales agents, brokers | 5% | 20% | No |
| | Insurance commission | Paid to insurance agents | 5% | 20% | No |
| | Marketing and advertising services | Promotions, brand marketing | 5% | 20% | No |
| *Investment income* | Dividends | Paid to shareholders | 5% | 15% | Yes |
| | Interest (general) | Bank interest, bonds, debentures | 15% | 15% | Yes |
| | Royalties | Use of copyrights, patents, trademarks | 5% | 20% | No |
| | Natural resource income | Mining, oil and gas royalties | 5% | 20% | No |
| *Rent* | Rent on immovable property | Rent on land and buildings | 10% | 30% | No |

Notes: Selected withholding tax (WHT) rates under the Income Tax Act as currently in force; categories applying only to non-residents or not part of our income variable are omitted. “Final (residents)” indicates whether the WHT settles the tax liability of resident individuals; otherwise, the amount withheld is credited against the annual income tax liability and the income must be declared in the individual return. Rates on dividends apply to resident individuals and to resident companies holding less than 12.5% of voting power (holdings above that threshold are exempt); interest refers to the general rate (special rates apply to e.g. housing bonds and bearer instruments). Non-resident rates may be reduced under double taxation agreements. Several rates and categories changed over the study period.

**Table A.2: Withholding tax income by type of recipient, 2022 (% of income in each category)**

| | Individuals | Firms | Not matched | Total |
|---|---|---|---|---|
| *Panel A: Final withholding income* | | | | |
| Interest | 3.3 | 30.0 | 66.7 | 100.0 |
| Dividends | 1.6 | 3.3 | 95.1 | 100.0 |
| All final withholding income | 2.4 | 15.7 | 81.9 | 100.0 |
| *Panel B: Non-final withholding income* | | | | |
| Professional fees | 12.6 | 55.2 | 32.2 | 100.0 |
| Commission | 35.4 | 58.9 | 5.7 | 100.0 |
| Rent | 61.6 | 1.8 | 36.6 | 100.0 |
| Royalties | 2.9 | 25.2 | 71.9 | 100.0 |
| Contractual fees | 2.9 | 95.4 | 1.7 | 100.0 |
| Marketing and sales | 1.4 | 0.6 | 98.0 | 100.0 |
| All non-final withholding income | 8.5 | 73.1 | 18.4 | 100.0 |
| All withholding income | 6.7 | 55.6 | 37.7 | 100.0 |

Notes: Share of gross income reported in withholding tax (WHT) returns for 2022 in each income category, by type of recipient. Shares exclude payment categories that never apply to resident individuals, but may otherwise include payments to non-residents. "Individuals": recipient tax ID appears in PAYE returns or ITRs. "Firms": ID appears in corporate income tax returns only. "Not matched": ID cannot be matched to PAYE, individual or corporate income tax records.

**Table A.3: Income imputation in the survey data**

| | 2016 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|
| *Panel A: Employees* | | | | | |
| Number of observations | 8,095 | 10,422 | 9,344 | 7,772 | 6,367 |
| Adjusted $R^2$ | 0.67 | 0.50 | 0.55 | 0.60 | 0.57 |
| Share imputed (%) | 4.5 | 10.0 | 10.7 | 1.7 | 6.6 |
| *Panel B: Self-employed* | | | | | |
| Number of observations | 17,398 | 12,425 | 11,718 | 10,129 | 8,408 |
| Adjusted $R^2$ | 0.34 | 0.18 | 0.22 | 0.18 | 0.20 |
| Share imputed (%) | 31.3 | 8.6 | 28.2 | 25.0 | 20.9 |

Notes: Income models are estimated separately for employees and self-employed workers in each survey round (2016 refers to KIHBS 2015/16, all other years to KCHS surveys). Number of observations is all workers in each category. Share imputed is the share of workers with unreliable income data for a given category and year, for whom incomes are imputed based on a model estimated on workers with reliable income data. Adjusted $R^2$ refers to the predictive power of that regression. For a small number of workers, incomes are imputed with a reduced model because covariates of the main model are missing.

**Table A.4: Method 2 estimates under homogeneous merging points**

| | Merging point | 2016 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|
| Gini | Low (13.5) | 0.640 | *0.616* | *0.666* | 0.632 | 0.650 |
| | High (14.2) | 0.641 | 0.612 | 0.660 | *0.634* | *0.652* |
| Top 1% share | Low (13.5) | 0.241 | *0.225* | *0.252* | 0.229 | 0.240 |
| | High (14.2) | 0.239 | 0.232 | 0.266 | *0.225* | *0.236* |
| Top 5% share | Low (13.5) | 0.424 | *0.407* | *0.456* | 0.417 | 0.436 |
| | High (14.2) | 0.425 | 0.410 | 0.466 | *0.416* | *0.434* |
| Top incomes: Tax − Survey | Low (13.5) | 30,721 | 225,608 | 344,431 | −35,685 | 59,645 |
| | High (14.2) | 59,285 | 97,124 | 133,765 | 20,038 | 111,926 |

Notes: Method 2 estimates use a common merging point for all years, expressed in log annual income: 13.5 (low, approximately KSh 729,000) or 14.2 (high, approximately KSh 1.47 million). Values in italics correspond to the baseline merging point and coincide with the Method 2 estimates in Table 2 (low in 2019 and 2020, high in 2021 and 2022). "Top incomes: Tax − Survey" is the difference between the number of individuals above the merging point in tax data and in the survey, as in Table 1. In 2021, the survey contains more individuals than tax data above the low merging point.

**Table A.5: Method 2 estimates for 2021 under alternative merging points**

| | Merging point (log income) | | |
|---|---|---|---|
| | 14.2 (baseline) | 15 | 17 |
| Gini | 0.634 | 0.635 | 0.646 |
| Top 1% share | 0.225 | 0.227 | 0.250 |
| Top 5% share | 0.416 | 0.419 | 0.436 |
| Top incomes: Tax | 290,203 | 78,766 | 3,690 |
| Top incomes: Survey | 270,165 | 75,587 | 1,336 |

Notes: Method 2 estimates for 2021 using alternative merging points, expressed in log annual income. The baseline merging point is 14.2 (approximately KSh 1.47 million), as used in Table 2; 15 and 17 correspond to approximately KSh 3.27 million and KSh 24.2 million. The last two rows report the number of individuals above the merging point in tax data and in the survey, as in Table 1.

**Table A.6: Combined inequality estimates for 2022, reweighted survey**

| | Method 1A | Method 1B | | | Method 2 |
|---|---|---|---|---|---|
| | Pre-tax | Pre-tax | Net | Δ | Pre-tax |
| Gini | 0.644 | 0.643 | 0.606 | −0.037 | 0.643 |
| Top 1% share | 0.209 | 0.208 | 0.171 | −0.037 | 0.231 |
| Top 5% share | 0.408 | 0.406 | 0.361 | −0.045 | 0.416 |

Notes: Estimates for 2022 after adjusting KCHS 2022 weights to correct the likely undercount of informal workers. Under adjusted weights, the size of the informal sector in KCHS 2022 corresponds to the average for the 2019, 2020 and 2021 surveys.

**Table A.7: Robustness to the definition of income in tax data**

| | Method 1B | | | Method 2 | |
|---|---|---|---|---|---|
| | Gini | Top 1% | Post-tax Gini | Gini | Top 1% |
| *Panel A: 2022* | | | | | |
| Baseline | 0.649 | 0.212 | 0.611 | 0.652 | 0.236 |
| Full tax data | 0.653 | 0.217 | 0.614 | – | – |
| Personal income | 0.647 | 0.209 | 0.608 | 0.647 | 0.230 |
| Employment income | 0.645 | 0.204 | 0.606 | 0.642 | 0.221 |
| PAYE income | 0.643 | 0.200 | 0.604 | 0.639 | 0.216 |
| *Panel B: Top 1% share* | | | | | |
| | 2016 | 2019 | 2020 | 2021 | 2022 |
| *Method 1B* | | | | | |
| Baseline | – | 0.220 | 0.244 | 0.242 | 0.212 |
| Full tax data | – | 0.226 | 0.249 | 0.244 | 0.217 |
| Personal income | – | 0.217 | 0.240 | 0.239 | 0.209 |
| Employment income | – | 0.212 | 0.235 | 0.236 | 0.204 |
| PAYE income | – | 0.210 | 0.234 | 0.234 | 0.200 |
| *Method 2* | | | | | |
| Baseline | 0.238 | 0.225 | 0.252 | 0.225 | 0.236 |
| Personal income | 0.230 | 0.218 | 0.246 | 0.218 | 0.230 |
| Employment income | 0.221 | 0.208 | 0.236 | 0.209 | 0.221 |
| PAYE income | 0.214 | 0.205 | 0.233 | 0.204 | 0.216 |

Notes: Panel A reports 2022 estimates; Panel B reports top 1% income shares for all years. “Baseline” corresponds to Tables 2 and 3. “Full tax data” also includes individuals in tax data who are not found in PAYE records; it only affects Method 1. “Personal income”: reported income subject to PIT. “Employment income”: employment income combining PAYE and ITR information. “PAYE income”: employment income as reported in PAYE returns. Post-tax Gini corresponds to the “Combined” columns in Table 3.

**Table A.8: PIT structures since 2019**

| 2019 | | 2020 (April) | | 2021 and 2022 | | Since July 2023 | |
|---|---|---|---|---|---|---|---|
| Income | MTR | Income | MTR | Income | MTR | Income | MTR |
| First 147,580 | 10 | First 288,000 | 10 | First 288,000 | 10 | First 288,000 | 10 |
| Next 139,043 | 15 | Next 200,000 | 15 | Next 100,000 | 25 | Next 100,000 | 25 |
| Next 139,043 | 20 | Next 200,000 | 20 | Over 388,000 | 30 | Next 5,612,000 | 30 |
| Next 139,043 | 25 | Over 688,000 | 25 | | | Next 3,600,000 | 32.5 |
| Over 564,709 | 30 | | | | | Over 9,600,000 | 35 |
| Pers. relief | 16,896 | Pers. relief | 28,800 | Pers. relief | 28,800 | Pers. relief | 28,800 |

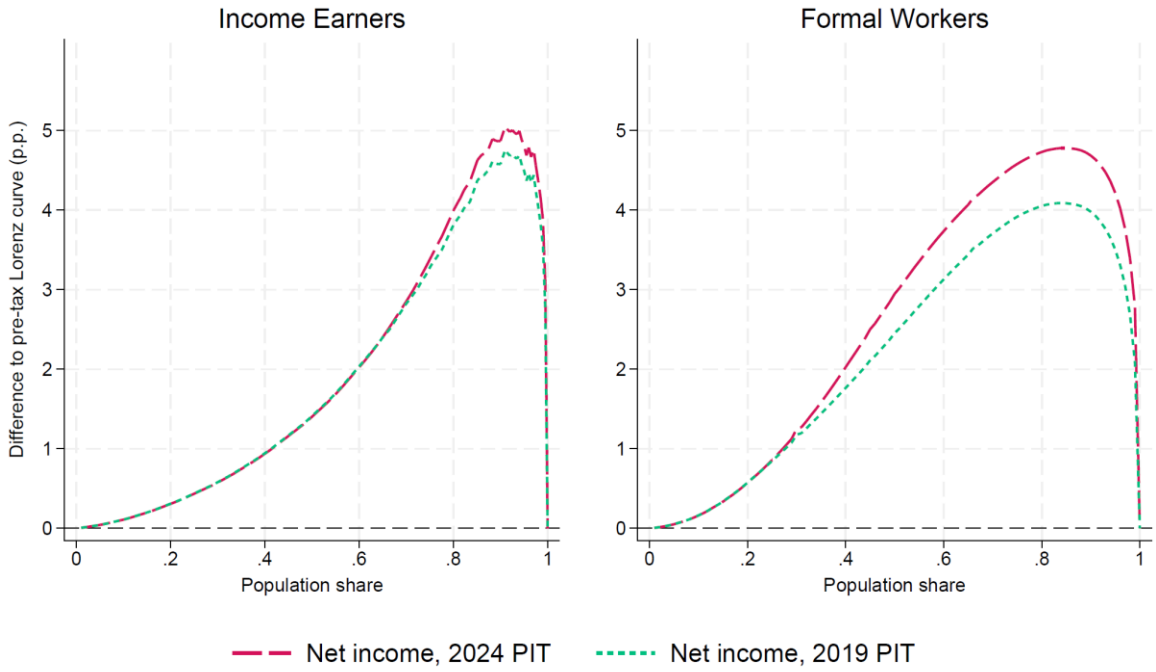
Notes: Annual taxable income and personal relief in KSh. MTR: marginal tax rate (%). “2020 (April)”: schedule in force since April 2020.

**Table A.9: PIT structures used in effective progressivity simulations**

| Current, 2019 relief | | Current, extra band | | Current, no top tax | |
|---|---|---|---|---|---|
| Income | MTR | Income | MTR | Income | MTR |
| First 168,960 | 10 | First 288,000 | 10 | First 288,000 | 10 |
| Next 219,040 | 25 | Next 100,000 | 20 | Next 100,000 | 25 |
| Next 5,612,000 | 30 | Next 100,000 | 25 | Over 388,000 | 30 |
| Next 3,600,000 | 32.5 | Next 5,512,000 | 30 | | |
| Over 9,600,000 | 35 | Next 3,600,000 | 32.5 | | |
| | | Over 9,600,000 | 35 | | |
| Pers. relief | 16,896 | Pers. relief | 28,800 | Pers. relief | 28,800 |

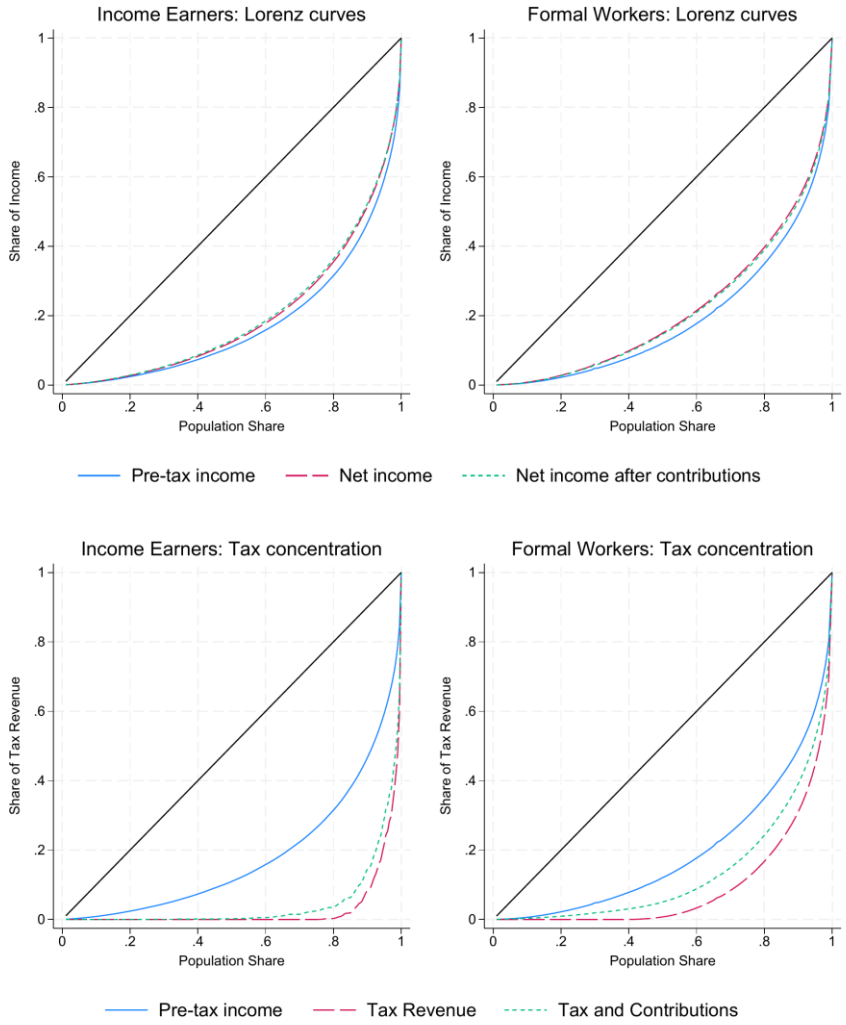
Notes: Annual taxable income and personal relief in KSh. MTR: marginal tax rate (%).

**Figure A.1: Lorenz curve gap: 2019 vs current PIT structure**




Notes: The figure shows the difference between the Lorenz curve of net income and the Lorenz curve of pre-tax income at each population share, in percentage points, under the current ("2024") and 2019 PIT schedules applied to 2024 data. The area under each curve equals half the reduction in the Gini coefficient achieved by the respective schedule. The left panel shows results for all income earners, according to the combination of tax and survey data using Method 1B; the right panel shows results for formal workers in tax data.

**Figure A.2: Effective progressivity and mandatory contributions**




Notes: The upper sub-figures show Lorenz curves for pre-tax income, net income under the current PIT structure, and net income after additionally subtracting NSSF, SHIF and AHL contributions at 2026 rates, using 2024 data. Contributions are deductible from taxable income, as under the rules in force since December 2024. The lower sub-figures compare the pre-tax income distribution to the distribution of PIT liabilities and of PIT liabilities plus contributions across income levels. The left sub-figures show results for the income distribution of all income earners, obtained through the data combination procedure described in Section 3 as Method 1B. The right sub-figures show results for the formal-sector income distribution.




# Mission

To strengthen local capacity for conducting independent,
rigorous inquiry into the problems facing the management of economies in sub-Saharan Africa.

The mission rests on two basic premises: that development is more likely to occur where there is sustained sound management of the economy, and that such management is more likely to happen where there is an active, well-informed group of locally based professional economists to conduct policy-relevant research.

Bringing Rigour and Evidence to Economic Policy Making in Africa

- Improve quality.
- Ensure Sustainability.
- Expand influence.

www.aercafrica.org

## Learn More

www.facebook.com/aercafrica

www.instagram.com/aercafrica_official/

twitter.com/aercafrica

www.linkedin.com/school/aercafrica/

Contact Us
African Economic Research Consortium
Consortium pour la Recherche Economique en Afrique
Middle East Bank Towers,
3rd Floor, Jakaya Kikwete Road
Nairobi 00200, Kenya
Tel: +254 (0) 20 273 4150
communications@aercafrica.org