WIDER Working Paper 2022/6

Income distribution in Uganda based on tax registers: what do top incomes say?

Markus Jäntti,[1] Milly Isingoma Nalukwago,[2]
and Ronald Waiswa[2]

January 2022

Abstract: We use income data from tax registers at the Uganda Revenue Authority from 2011 to 2017 to estimate top income inequality, focusing on the very top—the top 1, 0.1, and 0.01 per cent of the income distribution. The focus on the extreme top is facilitated by access to population data on formal sector income. The microdata from tax registries, submitted monthly to the Uganda Revenue Authority by employers, are supplemented by national accounts and population data that are used for control totals. Our results suggest that over the period we examine, incomes became substantially more heavily concentrated at the very top of the distribution.

---

[1] Stockholm University, Sweden, corresponding author: markus.jantti@sofi.su.se; [2] Uganda Revenue Authority, Kampala, Uganda

1   Introduction

Economic inequality and poverty are important social statistics that are monitored along with other social and economic indicators, such as economic growth and labour market participation, in both rich and developing societies. National governments, through their statistical agencies, as well as international organizations, such as the World Bank, publish inequality and poverty statistics for countries in sub-Saharan Africa, including for Uganda. Inequality and poverty statistics for Uganda along with many others appear in research-based compilations such as UNU-WIDER (2021). These are, as a rule, based on consumption rather than, as is the case for rich countries, on income. It can no doubt be argued that poverty as measured in terms of (monetized) consumption is the most important distributional statistic. It does not follow, however, that other dimensions of either economic well-being or its distribution are of no interest. As economies grow and the formal sector becomes more prominent, the role of income, its distribution, and its relation to consumption becomes of greater interest.

In this paper, we focus on an aspect of economic inequality that has increasingly come into focus across the past few decades, namely the share of income accruing to the very top of the distribution. In so doing, we are able to draw on a rich source of information, namely the income and tax registers collected by the Uganda Revenue Authority (URA). While income tax registers, by definition, include formal sector activities, we combine the tax register-based information with other sources to generate economy- and population-wide estimates of top income inequality.

Thus, here we examine the evolution of top incomes in Uganda across the years 2010 to 2018, using URA tax registers—the 'Pay-As-You-Earn' ('PAYE') and annual personal income tax (PIT) registers—complemented with imputations and estimates from national accounts and population data. We focus on the very top of the distribution of income in that we report the trends in distributional statistics for the top 1, 0.1, and 0.001 per cent of the population, including their income shares, average income, and taxes paid. While Uganda has high-quality household surveys, given the limited number of the economically active who are in formal employment, surveys typically include very few formally employed workers and are thus unlikely to provide very accurate estimates of the top of the distribution.

As with top income research in general, estimates are obtained by combining tax registers with aggregates from national accounts and population data.[1] In particular, we combine information from URA with estimates of gross domestic product or household final consumption expenditure (FCE) and the number of working-age adults, so-called control totals, to be able to estimate from URA data the top distribution. In contrast to much (but not all) top income research, we rely on unit record data for the top earners rather than tabulations of those at the top (for such research on Uganda, see Atkinson 2015).

The paper is structured as follows. After this introduction, we discuss some relevant research literature in Section 2. Section 3 presents our data, and Section 4 presents our main results.

---

[1] Some, such as the multi-country study by Chancel et al. (2019), combine surveys with similar external sources as we do; in this paper, we do not further rely on surveys.

2   Inequality in Uganda in earlier studies

Table 1: Earlier sources for income distribution in Uganda

Source	Resource	Period
Brunori et al. (2018)	Consumption and inequality of opportunity	2009–2010
Brunori et al. (2019)	Consumption and inequality of opportunity	2009
Atkinson (2015)	Income (top income, 0.1 and 0.05 per cent)	1948–1970
Chancel et al. (2019)	Income (top 1 per cent)	1989–2019

Source: authors' elaboration.

There are many estimates of inequality for Uganda for consumption, mainly based on Uganda household and panel surveys. These are the data that underlie, e.g., World Development Indicators estimates of both poverty and inequality.

As mentioned in the introduction, Chancel et al. (2019) provide estimates for top incomes in Uganda along with many other African countries (Alvaredo et al. 2017) from 1989 to 2019 (the earliest and latest years are based on projections). The estimates are available through the WID.world database and the underlying distributional data are based on surveys. Atkinson (2015) provides estimates of top income shares for Uganda along with several other East African countries based on originally British colonial administration tabulated tax records for 1948–1990. We will discuss the results in both of these below.

While their focus is not on top income inequality, but on that part of overall (consumption) inequality that can be said to be due to inequality of opportunity, Brunori et al. (2018) provide estimates of the so-called growth incidence curves for 2009–2010, which shows how real incomes changed for the same households across the two years, including at the top. Their data suggest that while incomes declined across the whole distribution, the relative changes at the top were the smallest (see their figure 1). Brunori et al. (2019) in turn examine inequality and inequality of opportunity in several African countries, allowing one to note that Uganda is neither among the most nor the least unequal among those.

3   Data

The data we rely on to measure inequality of top incomes are the monthly submissions on incomes paid and taxes withheld that employers submit to the Uganda Revenue Authority (URA) and annual personal income tax submissions. In the process of collecting income taxes, URA receives from employers regular (i.e. monthly) submissions of wages and salaries paid to employees as well as taxes withheld on their behalf, the so-called Pay-As-You-Earn (PAYE).[2] URA also receives from individuals annual submissions of income from employment and other sources, such as rental income, and business income from business owners (this source is dubbed the Personal Income Tax (PIT) database henceforth). We use both sources of information in this report.

Both the PAYE and PIT databases, referred to in what follows as the URA registers or databases, contain a wealth of information.[3] For the purposes of this paper, one limitation is that most of the individuals in the PAYE database have not applied for their own personal Tax Identification Number (TIN), although

---

[2] 'Employer' is used here to refer to a formal sector employer who has a Tax Identification Number for 'non-individuals' ('TIN-e'). 'Individual' or 'employee' is used to refer to natural persons, some of whom have a Tax Identification Number for individuals ('TIN-i') and some do not. We will discuss the latter two groups in detail below.

[3] We use URA registers when referring to the combined data from the two databases, but occasionally refer to either one of them separately.

an increasing number do. We shall leave for future work a closer elucidation of what can be gleaned of the background characteristics of individuals who do not have a TIN.

We focus here on overall top income shares and use for that purpose the PAYE/PIT gross income variable. PAYE data are reported monthly. We aggregate gross income within the fiscal year.[4] For those individuals who are covered by both PAYE and PIT data, we use the PIT. We use the consumer price index to adjust for price level changes. All money amounts are expressed in 2017 prices. The PAYE data are available from 2009 to 2018, at the time of writing. However, as detailed in the results section, low coverage of PAYE data in 2009 and 2010 makes it difficult to estimate top shares. We start the reporting with the 2010 fiscal year.

The PIT data are reported for each fiscal year, i.e. they are annual to begin with. We combine the two data sources such that for individuals who are in both PAYE and PIT data, we use the PIT data, i.e. we replace their PAYE records with those from PIT.

Table 2: Sources for estimating top incomes: control totals

	What	Source
National accounts	Household expenditure, overall and formal sector GDP at market prices measured through expenditure	Uganda Bureau of Statistics (2019)
Population	Persons of economically active age (15–74)	U.S. Census (2020)

Source: authors' elaboration.

As Atkinson (2007) discusses, estimating income distribution statistics from tax data when only part of the population pays taxes requires several additional pieces of information. We need information on the total number of income recipients or tax units in each year, and we need an estimate of total taxable income that can be used to impute incomes to those who are not covered by the tax registers. These pieces can be combined with information on the distribution of income among those who pay taxes to generate estimates that apply to the whole population.

Much of the research on top incomes relies on tabulations of the incomes among the population that has been taxed. This, in general, covers those income earners whose incomes exceed the threshold for income to be taxed. The control total income, which is needed to calculate income shares of those who are taxed (and by implication, those who were not), is taken from an external source, such as national accounts. The overall adult population—or, if taxation is based on families, the estimated total number of families—is used to estimate the quantiles of the top end of the distribution.

Our setting here is somewhat different. First, we have microdata for the people employed in the formal sector on whose earnings their employers file a PAYE declaration or who have filed a PIT submission, regardless of whether any taxes were in fact withheld.Second, we have no information in URA data of earnings in the informal sector, so whether or not someone is observed depends on belonging to the formal or informal sectors, not necessarily or primarily on the amount of earnings. What we do is that we work out what number of adults are not in the formal sector and impute to them the average income that is implied by taking into account the average income among those in the formal sector (i.e. who are in URA data) and the national accounts average income. In particular, let v_F, μ_F denote the number of those working in the formal sector and their mean income, and v_I, μ_I that of those who are not, with $v_F + v_I \equiv 1$ and $\mu = v_F \mu_F + v_I \mu_I$. Note that we have omitted the time index for brevity; population weights are based on total target population (N) and individuals within URA (N_F) so that $v_I \equiv (N - N_F)/N$. Then we work

---

[4] This is achieved by identifying individuals as persons who with the year have the same employer, and the same TIN or pseudo-TIN and the same name.

out the mean of the informal sector as

$$\mu_I = \frac{\mu - v_F \mu_F}{v_I}.$$ (1)

Our target population is the total number of potential tax payers. We approximate that with the population of economically active age, which we take to be the standard age of those in the labour force (16–64) plus those who are slightly above (65–74). Top income research typically includes all adults, starting at 18 or 15. In sensitivity analysis, we have experimented with a few different age ranges; while levels of estimates depend on these marginally, the changes across time do not. We use data from the International Database (IDB) of the U.S. Census (2020) on the population measured in 5-year age intervals to measure the size of the adult population.

To estimate total income, we rely on national accounts. The income control would ideally be the total income accruing to private households (sector S14). Sector accounts are not available for Uganda. We follow Atkinson (2015) and start from overall gross domestic product (GDP) (Uganda Bureau of Statistics 2019). These are available at market, rather than factor prices. Not all of GDP ends up as factor payments to households, so we must rely on different pieces to adjust GDP downwards. To start with, we deduct net exports, expenditure by the non-profit sector, and changes in inventories. We approximate factor payments abroad by net exports (which are negative, so reduce income). We are then left with household final expenditure (FCE), which equals household disposable income less savings, and capital formation, which incorporates some of the savings, and government FCE. While not ideal, we settle for approximating household sector income by household final consumption expenditure (HFCE). The consequence of this choice could be examined by utilizing survey data to conduct sensitivity analyses.

The key piece of information, however, is that we take as a control total that part of GDP that is generated in the formal sector. That is, we assume that the share of HFCE that is in the formal economy is the same as that for overall GDP. This share varies around 40 per cent, starting below 39 per cent in 2009 and increasing to around 41 per cent by 2012 (see lower panel in Figure 2). In using as the control total formal sector GDP (less the items detailed above), we are restricting our focus to formal sector incomes. This means that the not-covered population, called 'informal' above, does *not* consist of the the entire informal sector, but of those who are formally employed but whose employers do not engage in PAYE activities.

4

Figure 1: Control total for population of tax payers
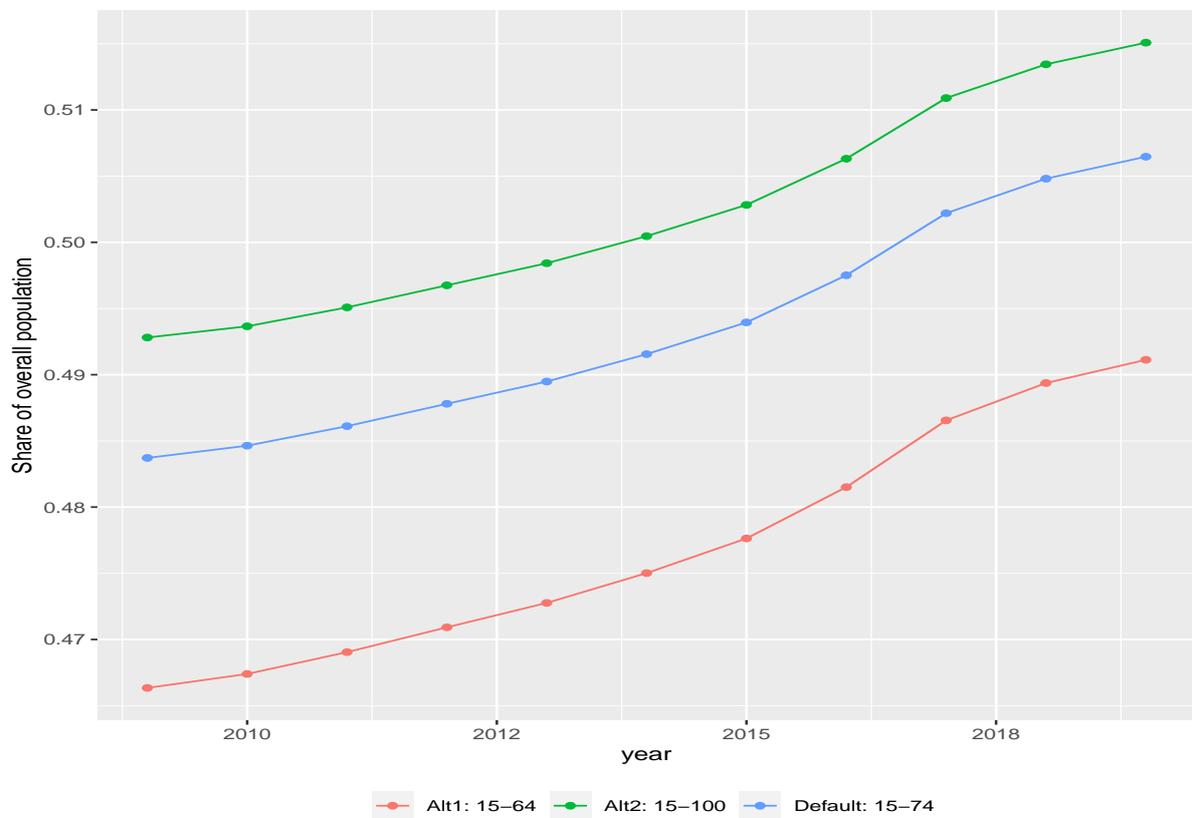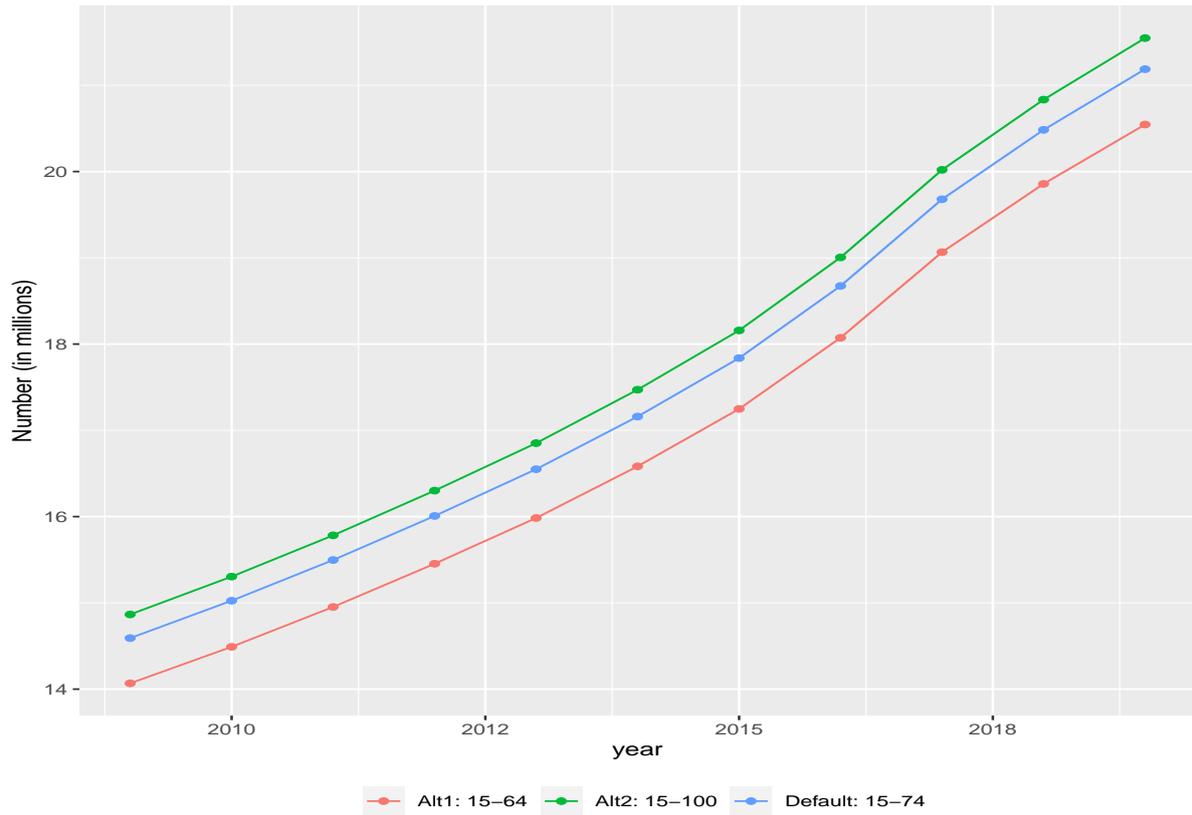
Source: authors' elaborations based on sources detailed in Table 2.

Figure 2: Control totals for income: household expenditure's share of GDP (upper panel) and formal sector share of overall GDP (lower panel)
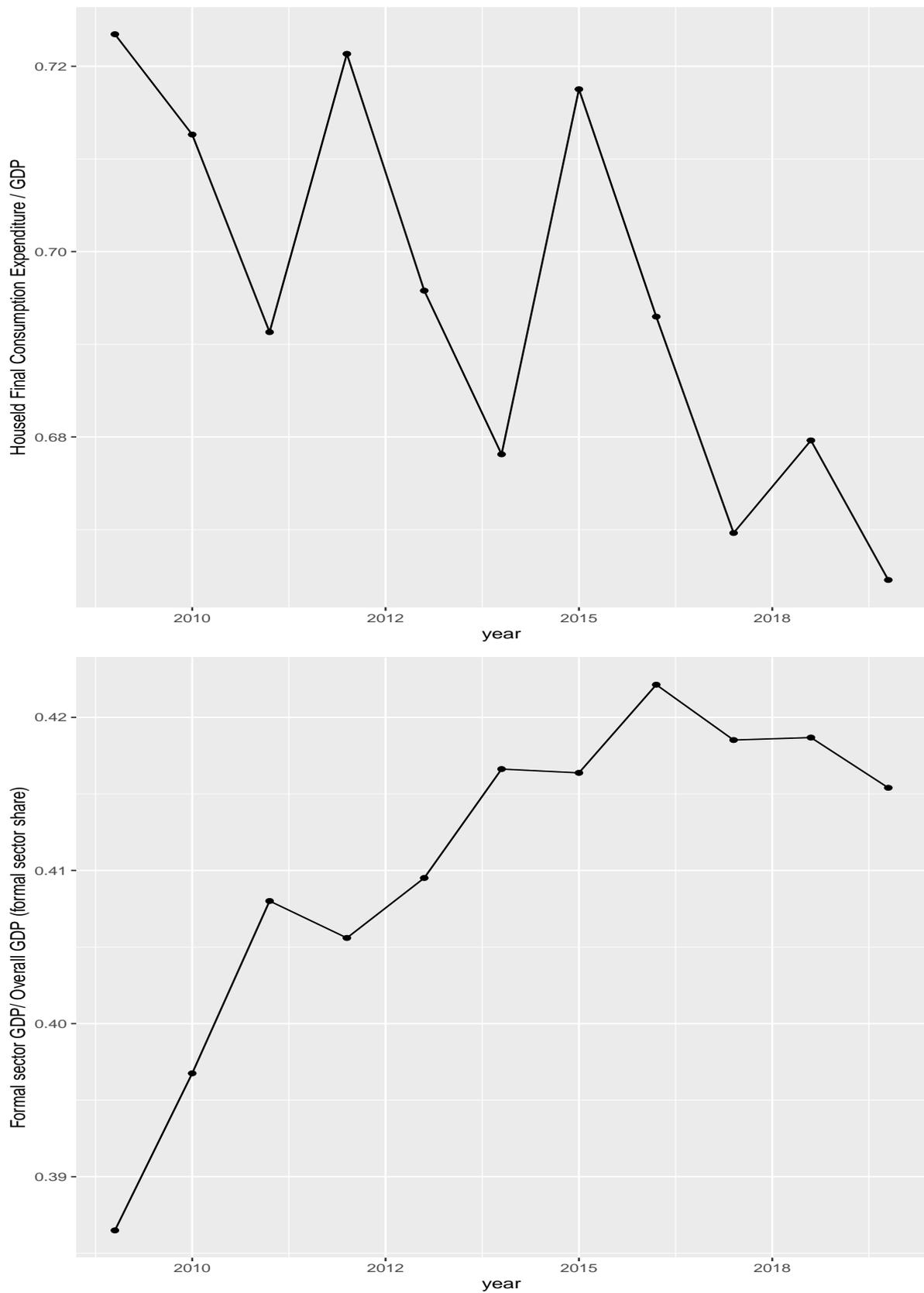
Source: authors' elaborations based on sources detailed in Table 2.

We show in Figures 1 and 2 some alternative control totals. As the default case, we have chosen to use 15–74 as the target population and household final consumption expenditure as the control total for calculating μ and v_I. Obviously, the more expansive age ranges in Figure 1 imply larger target populations. However, these grow at roughly the same rates. The population share of 15–64 year-olds is an exception in that it grows more rapidly across the time period we study than the share of the two broader age ranges, including our default case, 15–74 year-olds. The formal sector (see lower panel in Figure 2) grows initially as a share of GDP, implying faster growth than in the informal sector. It should be noted that the share of household final consumption expenditure jumps around a lot (see upper panel in Figure 2); the reasons for that jumpiness are beyond the scope of this paper.

4 Top incomes in Uganda

Table 3: Overall, non-URA and URA mean incomes and fraction of population in URA (in 2017 PPP USD)

Fiscal year	Overall	Average Not in PAYE	In PAYE	Prop in PAYE
2010	1160	1088	5394	1.7
2011	1227	1072	6664	2.8
2012	1280	1103	5567	4.0
2013	1249	1022	6517	4.1
2014	1255	1007	6774	4.3
2015	1343	1079	5800	5.6
2016	1316	1023	5191	7.0
2017	1243	922	5823	6.5

Source: authors' calculations based on URA registers, population data, and aggregate FCE from national accounts with sources detailed in Section 3 (see Table 2).

We begin the reporting of results by showing in Table 3 overall average income (μ), the average implied for the adult population not included in URA tax registers (URA, for short; μ_I), and the average of real gross income among the population included in URA (μ_F) in each of the years. The final column gives the population share of those included in URA as a fraction of the overall adult population, the estimated weight v_F. The monetary amounts are expressed in USD purchasing power parity (PPP) at 2017 prices.[5]

The overall mean is only a little higher than the mean implied for the informally active population, which is accounted for by the fact that the fraction that is included in URA is very small.

The overall mean is relatively stable at around 1,100–1,300 dollars per year. The average for the informal sector is about 100–300 dollars below that, varying around 1,000 dollars per year, i.e. showing little or no growth. The proportion in PAYE relative to the overall population increases from a little less than 2 per cent in 2010 to between 6 and 7 per cent by the end of the period. The average income within the covered population fluctuated between 5,000–7,000 dollars over the period, with little tendency to increase over time. This most likely reflects changes in selection into the formal sector across the years.

---

[5] Monetary amounts are first inflated to 2017 prices using price deflators from the World Development Indicators and then converted to US PPP dollars using the base year PPP conversion factor.

Table 4: Top income shares: income shares of top 1, 0.1, and 0.01 per cent of the population

Fiscal year	Income share of top		
	1 %	0.1 %	0.01 %
2010	7.4	3.8	1.5
2011	13.9	7.0	2.7
2012	15.2	7.5	2.9
2013	18.7	9.0	3.4
2014	19.6	9.2	3.3
2015	18.9	8.8	3.1
2016	20.3	9.7	3.4
2017	22.1	10.1	3.7

Source: authors' calculations based on URA registers, population data, and aggregate FCE from national accounts with sources detailed in Section 3 (see Table 2).

Only a small (albeit growing) fraction of the population is covered by URA registers, from which we can estimate top income shares or other statistics. The informal sector gets represented by a point-mass distribution at μ_I with density equal to v_I. Distributional characteristics can be estimated for those in the URA microdata whose income exceeds μ_I and where the combined cdf has $F > 1 - p$, where p is the relevant top group percentage. After some experimentation, we concluded that we are (mostly) able to estimate top shares for the top 1 and higher groups, so we decided to examine the top 1, 0.1, and 0.01 per cent groups. It is useful to recall that, although we are looking at unusually small fractions of the population ordered by income, our estimates for those groups are based on population unit-record data (as opposed to the usual tabulations in the top-income literature).

We show the results for top income shares in Table 4 and within-group means in Table 5. Note that the groups are not mutually exclusive, i.e. the top 0.01 per cent is included in the top 0.1 per cent, which in turn is included in the top 1 per cent.

The income shares do not exhibit a clear trend. Our first estimate in 2010 of the share of the top 1 per cent is 7.4 per cent of total income. In 2011, this almost doubles to 13.9 per cent. In 2012, we estimate the top 1 per cent share to increase further to 15.2 per cent, after which it increases to 18.7 per cent in 2013. The share peaks at 22.1 per cent in 2017, the last year in our data. The share of the top 0.1 per cent rises similarly from 3.8 per cent in 2010 to a high of 10.1 per cent in 2017. The pattern of the top 0.01 per cent is similar again, rising from 1.5 per cent 2010 to 3.7 per cent in 2017. In all three cases, concentration at the top is increasing across time.

There is substantial inequality in the top group. Moving from the top 1 per cent past 0.9 per cent, we see that the top 0.1 per cent share is roughly one half of the whole group's share. That is, the top 99–99.9 per cent earn roughly as much as the top 99.9–100 per cent of the population. The 99.90 to 99.99 per cent, in turn, earn slightly less than two thirds of the very top 99.99–100's income share, with between 4 and 10 per cent of all income for the whole top 0.1 per cent and between 1.5 and 4 per cent for the top 0.01 per cent.

Table 5: Top income mean: mean incomes of top 1, 0.1, and 0.01 per cent of the population (in 2017 PPP USD)

	Average income of top		
Fiscal year	1 %	0.1 %	0.01 %
2010	8614	44421	170043
2011	17037	86151	327642
2012	19412	96518	375891
2013	23362	112450	424462
2014	24617	115476	412894
2015	25332	117708	423077
2016	26677	127490	449823
2017	27448	125710	456166

Source: authors' calculations based on URA registers, population data, and aggregate FCE from national accounts with sources detailed in Section 3 (see Table 2).

Turning to average gross income at the top, shown in Table 5, we see some variation across the years but mostly quite steep increases in average income at the top. Income in the top 1 per cent varied, after initial low estimates of 8,614 and 17,037 dollars in 2010 and 2011–2012, to a high of 27,448 dollars in 2017. The average of the top 0.1 per cent varied from around 44,421 dollars in 2010 to a high of about 127,000 dollars in 2016, after which it declined to about 126,000 dollars; as a multiple of the overall top 1 per cent, this varies from 4.5 to just over five. The top 0.01 per cent, in turn, earns just under four times as much on average as the whole 0.1 group. Although the increase is not monotonic across time within all groups, average top incomes do increase substantially across the period.

Table 6: Top income mean: mean incomes of top 1–0.9, 0.100–0.099, and 0.099-0.001 per cent of the population (in 2017 PPP USD)

	Average income of top		
Fiscal year	1–0.9 %	0.1–0.099 %	0.099-0.001 %
2010	4635	30468	170043
2011	9358	59330	327642
2012	10845	65492	375891
2013	13419	77954	424462
2014	14518	82410	412894
2015	15069	83793	423077
2016	15476	91682	449823
2017	16530	88992	456166

Source: authors' calculations based on URA registers, population data, and aggregate FCE from national accounts with sources detailed in Section 3 (see Table 2).

Indeed, if we look at the fractile groups' income means disjointly in Table 6, we can observe some variation in inequality at the top. The top 0.100–0.099 per cent earns about six times as much as the 1–0.9 per cent, and the top 0.099–0.001 about five times as much as the 0.100–0.099 per cent. However, despite some tendency for top income inequality to increase in that income shares indicate increased concentration to the top, as seen in Table 4, there is no tendency for increased inequality within the top 1 per cent in the sense that real income growth was not systematically higher among the very rich than the two other groups. Year-to-year real income growth varies about as much within groups across the years as it does across groups.

5 Discussion and concluding comments

To examine the level of and change in top income inequality in Uganda, we have used data that employers submit on behalf of their employees on a monthly basis to the Uganda Revenue Authority to

estimate top income shares and average incomes in Uganda for fiscal years 2011–2018 combined with national accounts and population data. Our estimates for the top 1, 0.1, and 0.01 per cent suggest around 8–23, 4–10, and 2–4 per cent income shares for the groups. Our estimates suggest sharp increases in the concentration of incomes to the top of the distribution.

The tax-record-based estimates by Atkinson (2015) for 1948–1970 for top 0.1 per cent are substantially lower than ours. But once we recognize that they apply not to the share of formal sector income but rather overall GDP, they are roughly comparable to ours, varying between 3.3 and 4.6 per cent (recall that around 40 per cent of GDP is in the formal sector). By contrast, the more recent estimates by Chancel et al. (2019) and available at the WID.world database for 1989–2019, based on adjusted survey data, do not provide estimates for the share of the very top. A more detailed comparison with their broader estimates, which presupposes supplementing our data with surveys, is an important next step.

References

Alvaredo, F., Atkinson, A. B., Piketty, T., Saez, E., and Zucman, G. (2017). 'The World Wealth and Income Database'. Statistical database. World Inequality Database (WID.world). Available at: http://wid.world (accessed 3 April 2017)

Atkinson, A. B. (2007). 'Measuring Top Incomes: Methodological Issues'. In A. B. Atkinson and T. Piketty (eds), *Top Incomes Over the 20th Century* (pp. 18–42). Oxford: Oxford University Press.

Atkinson, A. B. (2015). 'Top Incomes in East Africa Before and After Independence'. WID.world Working Paper 2015/2. WID.world. Available at: https://wid.world/document/wid_working_paper_2015_2_east_africa/ (accessed in January 2022)

Brunori, P., Palmisano, F., and Peragine, V. (2018). 'Consumption Dynamics and Inequality of Opportunity with an Application to Uganda'. *Review of Development Economics*, 22(2): 632–57. https://doi.org/10.1111/rode.12357

Brunori, P., Palmisano, F., and Peragine, V. (2019). 'Inequality of opportunity in sub-saharan africa'. *Applied Economics*, 51(60): 6428–58. https://doi.org/10.1080/00036846.2019.1619018

Chancel, L., Cogneau, D., Gethin, A., and Myczkowski, A. (2019). 'How Large Are African Inequalities? Towards Distributional National Accounts in Africa, 1990-2017'. WID.world Working Paper 2019/13. WID.world. Available at: https://wid.world/document/cgm2019-full-paper/ (accessed in January 2022)

Uganda Bureau of Statistics (2019). 'AGDP Publication Tables FY - Uganda 18/19 (Rebased)'. Statistical database. Available at: https://www.ubos.org/wp-content/uploads/statistics/AGDP_Publication_Tables _Oct_201819.xls (accessed 2 December 2019)

UNU-WIDER (2021). 'World Income Inequality Database (WIID)'. Version 31 May 2021. https://doi.org/10.35188/UNU-WIDER/WIID-310521

U.S. Census (2020). 'International Data Base'. Statistical database. Available at: https://www.census.gov/data/ developers/data-sets/international-database.html (acessed 27 October 2020)