Building up efficient and fair tax systems – lessons based on administrative tax data

The Uganda Revenue Authority Pay-As-You-Earn (PAYE) data

Gerald Agaba,¹ Quynh Tieu,² Maria Jouste,² and Brian Arinaitwe³

October 2023

Abstract: This technical note describes the Uganda Revenue Authority (URA) Pay-As-You-Earn (PAYE) data covering financial years 2013/14 to 2021/22. PAYE is paid by an employer who withholds personal income tax on employees' wages and other employment incomes like benefits and allowances. The data are constructed from monthly PAYE tax returns and assessments and contain 30 variables, including information on employees' incomes and taxes and employers' characteristics. These data can be merged with other available administrative tax data, such as the URA firm panel. Thus, it can be used to study formal employment and income distribution in Uganda. This note describes the main elements of the data and provides a list of variables. Finally, the note provides some summary statistics of the data.

Key words: administrative tax data, employment income, personal income tax, Uganda

JEL classification: H24, D31, J31, O23

Acknowledgements: We thank Norbert Afya and Nicholas Musoke for their valuable support on the construction of the data. We thank Amina Ebrahim, Bruce McDougall, and Kwabena Adu-Ababio for their helpful comments and feedback. This work is part of a more extensive research and capacity-building initiative between the Uganda Revenue Authority (URA) and UNU-WIDER.

Note: This study has received ethical approval by the Joint Ethical Review Board of the United Nations University (Ref No: 202104/01) on 11 May 2021.

Recommended citation for the dataset

Uganda Revenue Authority (2023). *The Uganda Revenue Authority Pay-As-You-Earn (PAYE) data* (version 1). Kampala: Uganda Revenue Authority.

This study has been prepared within the UNU-WIDER project Building up efficient and fair tax systems – lessons based on administrative tax data, which is part of the Domestic Revenue Mobilization programme. The programme is financed through specific contributions by the Norwegian Agency for Development Cooperation (Norad).

Copyright © UNU-WIDER 2023

UNU-WIDER employs a fair use policy for reasonable reproduction of UNU-WIDER copyrighted content—such as the reproduction of a table or a figure, and/or text not exceeding 400 words—with due acknowledgement of the original source, without requiring explicit permission from the copyright holder.

Information and requests: publications@wider.unu.edu

https://doi.org/10.35188/UNU-WIDER/WTN/2023-3

United Nations University World Institute for Development Economics Research – UNU-WIDER



Katajanokanlaituri 6 B, 00160 Helsinki, Finland

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

¹ Uganda Revenue Authority, Kampala; ² UNU-WIDER, Helsinki; ³ Kansas State University, Manhattan, Kansas (at time of writing: Uganda Revenue Authority); corresponding author: Maria Jouste, jouste@wider.unu.edu

1 Introduction

The Uganda Revenue Authority Pay-As-You-Earn (PAYE) tax data are an employer panel dataset consisting of all PAYE returns that employers remit to the Uganda Revenue Authority on behalf of the employee. The monthly data cover years from 2013/14 to 2021/22 based on the country's financial year that runs from 1 July in the previous year to 30 July in the subsequent year. The data have limitations related to employee identifications; therefore, they cannot be interpreted as an employer-employee panel (see more details in Section 2).

PAYE tax is a form of personal income tax, where tax is based on employment income. PAYE is one of the vital tax revenue sources—it is the largest source of direct taxes in Uganda (see Figure 4 in Jouste et al. 2023). Table 7 shows that observations and PAYE revenues have increased steadily since 2013/14.

PAYE is deducted from the employee's chargeable income following the personal income tax schedule.¹ Chargeable income² is a sum of basic salary and all allowances and benefits, such as medical and leave allowance and housing benefits, minus deductions, which currently include only Local Service Tax (LST) (for more information on LST, see Waiswa et al. 2023). The resident taxpayer's rates, which have been unchanged since July 2012, are shown in Table 1. Different rates are applied to non-residents as opposed to resident taxpayers. The tax rates for non-resident taxpayers can be found in Table A1 in the Appendix. In addition to progressive tax rates shown in Table 1, a flat tax rate of 30 per cent is applied to the secondary job. In this case, the first employer withholds tax at the standard progressive tax rate and the second at the flat rate.

Table 1: Individual income tax rates and thresholds for monthly chargeable income for resident taxpayers in Uganda for the data years

Monthly chargeable income	Tax rate
Not exceeding 235,000	0%
Over 235,000, but not exceeding 335,000	10% of the amount exceeding 235,000
Over 335,000, but not exceeding 410,000	10,000 plus 20% of the amount exceeding 335,000
Over 410,000, but not exceeding 10,000,000	25,000 plus 30% of the amount exceeding 410,000
Over 10,000,000	2,902,000 plus 40% of the amount exceeding 10,000,000

Note: all monetary values are in UGX.

Source: author's compilation based on the Income Tax (Amendment) Act 2012.

https://www.ura.go.ug/resources/webuploads/INLB/Taxation%20Handbook%204th%20Edition%202022_10.02. 2022.pdf (accessed 1 September 2023).

¹ The Income Tax Act Cap 340 and all its amendments are accessible on the Parliament of Uganda website https://www.parliament.go.ug/documents/acts. In addition, more information about taxation in Uganda can be found in the yearly Taxation Handbook published on the URA website. The most recent version of the Taxation Handbook is available here

² Chargeable income is also referred to as taxable income in this technical note.

2 Key elements of the panel

Taxpayers are expected to file 'returns' for their tax position; however, they might occasionally fail to do this. In such cases, a URA Tax Officer will issue them an 'assessment' for their tax position. Therefore, we end up with two series of datasets: Returns and Assessments. Returns are entirely self-reported by an employer, while tax officers do assessments on behalf of the URA. Therefore, in the panel, the rows for tax liability can be returns or assessments, depending on which series the datasets belong to. The former is always self-reported, while the Revenue Authority imputes the latter.

The data collected from the returns and assessments are relatively rich and have been captured over time. As such, these can be used to study various topics, such as the employment status and wage income distribution in Uganda, the growth rate of this number, firm survival, firm hiring and remuneration trends, and several other potential topics. Because the employees and their taxable income are also measured, the labour force and the income distribution can be examined, particularly in the aggregate. Some individual level analysis may also be possible, although identification problems are noted below.

Furthermore, the datasets contain identifiers that allow them to be merged with other tax datasets at the URA. For example, the return or filing number can be used together with the Tax Identification Number (TIN) of the employer to combine the Corporate Income Tax (CIT) and Value-Added Tax (VAT) datasets. This is possible as there are no duplicates in terms of employer-level identification based on return identifier (return ID), anonymized firm identifier (firm ID), return month, and return year variables. The merging between datasets can be based on return ID, anonymized firm ID, return month, and return year variables. In all PAYE datasets, the general location of the tax office is available from the variable *c_currentstationname*. More detailed information on location (e.g., district, sub-district, county) or other firm characteristics variables (e.g. sector) can also be acquired by matching PAYE with CIT and/or registration data using anonymized firm ID.

The URA anonymizes the key identifier in the dataset (the TIN of an employer/firm) before data developers/processors can access it for further work. This is to prevent breach of privacy rules such that no firm is tracked based on their actual TIN. Moreover, the employer/firm TIN is uniquely identified and has a harmonized 20-digit format. By 'uniquely identified', we mean that each firm has one unique anonymized TIN value across all filing/return years. This is possible as each employer is assigned one unique TIN number by the URA. Data developers ensure that return year filings are devoid of duplicate employer TIN values. The same anonymization rules apply to the TIN of an employee/individual. However, it is not uniquely identified and has no harmonized format because many employees do not have a TIN. For those employees who have a TIN, their TIN is anonymized under the 20-digit format. For employees who do not have a TIN, an anonymized pseudo-TIN based on a 30-digit format is generated for them. In any case whether an employee has or does not have an official TIN, the employer reports the tax liability on behalf of all of their employees using the employee's name (and the employee's official TIN, if any), for all Returns datasets. Due to the lack of TINs for a portion of employees, the PAYE datasets are only uniquely identified at the employer level (named c_firm_id in the datasets) and not at the employee level (named *c_tin_of_employee* in the datasets), since the harmonization from the former is based on a 20-digit format, but that of the latter is based on both a 20- and 30-digit format.

Besides, for the variable c_tin_of_employee in all Returns datasets, it should be noted that this variable is for now not an interest of and not validated by the URA. The URA does not check if an employer has put a wrong or right TIN for an employee even if that employee has an employee

TIN registered with the URA. This results in many employees appearing to have the same TIN, no matter if the TIN is a 20-digit anonymized TIN from the real TIN or a 30-digit pseudo-TIN. The suggestion here for researchers is to not use the variable $\varepsilon_{tin_{tot}}$ analyses, because it is not validated by the URA and unreliable to be used. Therefore, we do not recommend using the datasets as employee-level panel. However, each observation still pertains to unique tax liability filling for each employee working in the firms for a particular return month and year. Employees with multiple jobs are possible, but due to the identification of employee TINs, they cannot be tracked in the datasets. In terms of missing values, there is one missing value in the number of observations for all financial years, so we dropped that one missing value observation for all Returns and Assessments datasets.

Lastly, two papers have already been published based on a previous version of the PAYE dataset. These are the evaluation of personal income tax reform in 2012 by Jouste et al. (2023) and the investigation of the income distribution of top incomes by Jäntti et al. (2022). The previous version of PAYE data was extracted only for two research projects. It covered years 2010–18, included only returns, and it was not documented or made publicly available. Moreover, the years before 2013 were not comprehensive because the full roll-out of electronic filing was ready in the end of 2012.

3 Data quality checks and adjustments

In the Returns and Assessments datasets, there are some anomalies where employees' tax liability is larger than their incomes. There are three variables relevant to this discussion: PAYE deducted (named sch1_paye_deducted in the datasets), tax on total income (named sch1_tax_on_total_income), and total taxable income (named sch1_total_taxable_income).

One way in which this type of anomaly arises is when both PAYE deducted and tax on total income are greater than total taxable income. These seem highly problematic, and we decide to drop these anomalous cases. Fortunately, these cases are relatively rare: the number of such dropped cases is documented per year in Table 2 below.

Table 2: Number of dropped anomalies from the datasets by financial year

Financial year	A: Returns – Number of dropped anomalies	B: Assessments – Number of dropped anomalies
2013/2014	44	1
2014/2015	53	0
2015/2016	134	0
2016/2017	381	1
2017/2018	625	0
2018/2019	76	0
2019/2020	430	6
2020/2021	463	0
2021/2022	185	0

Source: authors' calculations using the URA PAYE data.

Another way in which this type of anomaly can arise is when one of the two tax liability variables is larger than taxable income. To check this, we generated a tag variable (named *tagPAYEdeducted* in the datasets) to indicate whether PAYE deducted is greater than the variable total taxable income. Because we have already removed cases where both variables are greater than total taxable income, in these remaining tagged cases, the variable tax on total income is less than total taxable

income. Tables 3 and 4 present the summary of tax on taxable income (sch1_tax_on_total_income) and total taxable income (sch1_total_taxable_income) for these tagged cases.

Table 3: Returns data—summary of tax on taxable income and total taxable income on those having PAYE deducted greater than the total taxable income by financial year

Financial year	Number of observations having tagPAYEdeducted = 1	Average of tax on taxable income (sd)	Average of total taxable income (sd)
2013/2014	2,746	73,695 (390197)	324,387 (1222574)
2014/2015	3,554	49,259 (313497)	260,070 (961481)
2015/2016	6,864	67,226 (295371)	385,303 (924366)
2016/2017	5,358	62,887 (289432)	316,020 (982052)
2017/2018	8,138	55,444 (287470)	294,999 (924052)
2018/2019	9,484	68,806 (423721)	329,573 (1262174)
2019/2020	20,389	31,238 (542719)	158,843 (1464557)
2020/2021	23,281	30,943 (468536)	144,112 (1272284)
2021/2022	23,311	36,346 (306848)	197,578 (912693)

Notes: standard deviation in brackets. All monetary values are in UGX.

Source: authors' calculations using the URA PAYE data.

Table 4. Assessments data—summary of tax on taxable income and total taxable income on those having PAYE deducted greater than the total taxable income by financial year

Financial year	Number of observations having tagPAYEdeducted = 1	Average of tax on taxable income (sd)	Average of total taxable income (sd)
2013/2014	340	26,137 (198550)	102,202 (684595)
2014/2015	258	5,280 (57213)	23,918 (222971)
2015/2016	73	103,161 (152363)	567,178 (587052)
2016/2017	188	65,463 (206926)	306,773 (772342)
2017/2018	166	80,243 (293096)	343,788 (1052440)
2018/2019	342	46,052 (217559)	200,752 (790426)
2019/2020	290	177,632 (1440343)	682,374 (4812684)
2020/2021	185	45,650 (251338)	188,092 (894988)
2021/2022	27	592,045 (746193)	2,249,315 (2531153)

Notes: standard deviation in brackets. All monetary values are in UGX.

Source: authors' calculations using the URA PAYE data.

4 Data characteristics

This section provides some basic summary statistics and information on the two datasets. We begin in Table 5 with the number of observations in each. In Table 5, we observe persistent increases in the number of employer-reported returns and continuous decreases in the number of filings done by the Revenue Authority. This points to effective filing on the side of employers which decreases the amount of assessments the Tax Authority needs to do.

Table 5: Number of observations for the datasets by financial year

Financial year	A: Returns – Number of observations	B: Assessments – Number of observations
2013/2014	4,065,181	46,003
2014/2015	5,295,746	50,510
2015/2016	6,935,700	53,623
2016/2017	7,838,646	48,703
2017/2018	8,407,608	39,073
2018/2019	9,017,928	32,083
2019/2020	9,361,426	33,355
2020/2021	9,666,556	10,438
2021/2022	10,335,378	3,985

Source: authors' calculations using the URA PAYE data.

Table 6 shows the number of employers for all Returns and Assessments datasets by financial year. There is a consistent increase in the number of employers for all Returns datasets over the recorded financial years, with a peak of 32,787 employers for 2021/22. The number of employers who didn't file a tax return in time (and a URA Tax Officer had to step in and do their tax position assessment) for all Assessments datasets, on the other hand, increased and decreased over the recorded financial years, with a peak of 340 employers not filling their tax returns in time for the financial year 2019/20.

Table 6: Number of employers for all Returns and Assessments datasets by financial year

Financial Year	A: Returns – Number of employers	B: Assessments – Number of employers
2013/2014	12,097	246
2014/2015	13,560	223
2015/2016	14,939	147
2016/2017	17,312	106
2017/2018	20,072	124
2018/2019	23,047	153
2019/2020	25,379	340
2020/2021	27,523	132
2021/2022	32,787	65

Source: authors' calculations using the URA PAYE data.

Table 7 shows the amount of calculated cumulative tax on taxable income (Column A), the officially reported collected tax amount provided by the URA (Column B), and the difference between the officially reported amount and the calculated cumulative amount (Column C) in billion UGX. Column D, reports a ratio of the calculated tax amount to officially reported tax amount in percentage. There is an increasing trend in the calculated cumulative tax on taxable income amount over the recorded financial years, peaking at 3,251.087 billion UGX for 2021/22. The average difference amount is 262.134 billion UGX and the average ratio of the calculated tax amount to officially reported tax amount is 88.37 per cent.

The difference between the amount of calculated cumulative tax on taxable income and the officially reported collected tax amount provided by the URA can be explained by various reasons, for example:

1. The tax on taxable income amount from the Returns datasets is entirely self-reported by firms and could be possibly under-reported. After tax audits, some firms would need to pay their accurate amount of tax liability to the URA. Therefore, due to audits, the officially

- reported collected tax amount would exceed the self-reported filed amount for a specific financial year.
- 2. The officially reported collected tax amount provided by URA is greater than the calculated cumulative tax on taxable income amount from the data because of the missing unreconciled data that are directly reported to the Bank of Uganda by the public sector office and are not channelled through the URA databases.

Table 7: Amount of calculated cumulative tax on taxable income, officially reported collected tax amount provided by the URA, and the difference between the officially reported collected tax amount and the calculated cumulative tax on taxable income amount from the data (in billion UGX)

Financial year	(A) Calculated cumulative tax on taxable income amount from the data (in billion UGX)	(B) Officially reported collected tax amount provided by the URA (in billion UGX)	(C) Difference between the officially reported collected tax amount and the calculated cumulative tax on taxable income amount from the data (in billion UGX)	(D) Column A divided by Column B in %
2013/2014	1,159.470	1,397.56	220.09	82.96%
2014/2015	1,340.769	1,613.24	272.471	83.11%
2015/2016	1,631.915	1,803.53	171.615	90.48%
2016/2017	1,916.031	2,114.99	198.959	90.59%
2017/2018	2,170.907	2,396.11	225.203	90.60%
2018/2019	2,480.721	2,811.30	330.579	88.24%
2019/2020	2,723.761	3,039.83	316.069	89.60%
2020/2021	2,868.095	3,109.14	241.045	90.25%
2021/2022	3,251.087	3,634.26	383.173	89.46%
Average:	2,171.417	2,435.551	262.134	88.37%

Source: authors' calculations using the URA PAYE data. The officially reported collected tax amounts are reported in URA Annual Data Book 2018/19 and 2021/22 and can be found in the Annual Revenue Performance Reports for financial years. All reports from 2015/16 onwards are available at www.ura.go.ug. The information for financial years 2013/14 and 2014/15 was received from the URA.

Table 8 shows the number of resident/non-resident taxpayers for all Returns datasets by financial year. The number of resident taxpayers consistently increased over the recorded financial years, with the peak of 10,313,168 resident taxpayers in 2021/22. The number of non-resident taxpayers, on the other hand, increased and decreased over the recorded financial years, with the peak of 51,328 non-resident taxpayers in 2017/18.

Table 8: Number of resident/non-resident taxpayers for all Returns datasets by financial year

Financial year	Number of resident taxpayers	Number of non-resident taxpayers
2013/2014	4,055,320	9,861
2014/2015	5,282,450	13,296
2015/2016	6,905,734	29,966
2016/2017	7,818,991	19,655
2017/2018	8,356,280	51,328
2018/2019	8,988,478	29,450
2019/2020	9,343,226	18,200
2020/2021	9,648,624	17,932
2021/2022	10,313,168	22,210

Source: authors' calculations using the URA PAYE data.

Table 9 shows the number of taxpayers who paid the flat PAYE tax rate for all Returns datasets by financial year. The number of taxpayers who paid the flat PAYE tax rate over the recorded financial years increased from the 2013/2014 to 2019/2020, then decreased from 2019/2020 to 2020/2021 and increased again from 2020/2021 onwards. The highest number of taxpayers who paid the flat PAYE tax rate was 241,733 in 2019/2020.

Table 9: Number of taxpayers who paid the flat PAYE tax rate for all Returns datasets by financial year

Financial year	Number of taxpayers who paid the flat PAYE tax rate
2013/2014	138,156
2014/2015	155,059
2015/2016	170,636
2016/2017	196,080
2017/2018	206,188
2018/2019	219,558
2019/2020	241,733
2020/2021	212,582
2021/2022	223,473

Source: authors' calculations using the URA PAYE data.

5 Future plan of updating the data and potential research topics using the data

For now, the observations were extracted from financial year 2013/14 up to financial year 2021/22. The data are planned to be extracted and updated again from March 2024 to May 2024 and then annually.

The exact schedule to extract and update the PAYE data is planned as follows:

- Step 1: Extract: in March annually;
- Step 2: Clean: in April annually;
- Step 3: Document: in May annually.

In terms of potential research topics, the PAYE data can be used to answer the following research questions:

- 1. What are income shares and trends of formal employment?
- 2. How has income inequality developed in formal employment?

- 3. How progressive is the PAYE schedule?
- 4. What are PAYE revenue trends?

In addition, the dataset can be used, for example, to answer research questions related to firm and employment if it is merged with the URA firm panel. This list of questions is not exhaustive.

References

- Jouste M., T. Kaidu Barugahara, J. Okello Ayo, J. Pirttilä, and P. Rattenhuber. (2023). "Taxpayer Response to Greater Progressivity: Evidence from Personal Income Tax Reform in Uganda'. WIDER Working Paper 2023/66. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2023/374-1
- Jäntti M., M. I. Nalukwago, and R. Waiswa. (2022). Income Distribution in Uganda Based on Tax Registers: What Do Top Incomes Say?' WIDER Working Paper 2022/6. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2022/137-2
- Waiswa, R., J. Okello Ayo, S. Kavuma, D. McLennan, M. Noble, and G. Wright (2023). SOUTHMOD Country Report Uganda UGAMOD v2.0. Helsinki: UNU-WIDER.

Appendix A

Table A1: Individual income tax rates and thresholds for monthly chargeable income for non-resident taxpayers in Uganda for the data years

Monthly chargeable income	Tax rate
Not exceeding 335,000	10%
Over 335,000, but not exceeding 410,000	33,500 plus 20% of the amount exceeding 335,000
Over 410,000, but not exceeding 10,000,000	48,500 plus 30% of the amount exceeding 410,000
Over 10,000,000	2,925,500 plus 40% of the amount exceeding 10,000,000

Note: all monetary values are in UGX.

Source: author's compilation based on the Income Tax (Amendment) Act 2012

Appendix B: variables description

There are 30 variables in both Returns and Assessments datasets for all financial years from 2013/14 to 2021/22. These are organized to follow the logic of the PAYE return form and it is recommended that users use this dataset alongside the PAYE return form, in order to best understand the breakdown of variables.³ All variables in the table below are from Schedule 1 of the PAYE return form.⁴

Table A2: PAYE Schedule 1 variable list

Name	Variable label
c_fy	Financial Year
c_source_type	Type of Return
c_return_no	Return Number
c_currentstationname	Current Station Name
c_tin_of_employee	Anonymised TIN of Employee
c_firm_id	Anonymised Firm ID
c_return_date	Return Date
c_returnfromdate	Return From Date
c_returntodate	Return To Date
sch1_basic_salary	Basic Salary
sch1_housing_allowance	Housing Allowance
sch1_transport	Transport Allowance
sch1_medical	Medical Allowance
sch1_leave	Leave Allowance
sch1_over_time	Over Time Allowance
sch1_other_taxable_allow	Other Taxable Allowance
sch1_housing	Housing Benefit
sch1_motor_vehicle	Motor Vehicle Benefit
sch1_domestic_servants	Domestic Servants Benefit
sch1_other_taxable_benf	Other Value of Benefits
sch1_gross_income	Gross Income
sch1_allowable_deductions	Allowable Deduction
sch1_total_taxable_income	Chargeable Income (i.e. taxable income)
sch1_is_flat_rate	Does employee fall under fixed tax rate of 30%
sch1_residency_status	Residency Status
sch1_tax_on_total_income	Tax on Chargeable Income
sch1_paye_deducted	PAYE deducted
tagPAYEdeducted	gen tagPAYEdeducted = 1 if sch1_paye_deducted > sch1_total_taxable_income

Source: authors' listing from the URA PAYE data.

³ The latest version of the PAYE return form can be found on www.ura.go.ug at the bottom of the page under the title 'Download Forms'.

⁴ All variables with amounts are denoted in Ugandan Shillings (UGX).