WIDER Working Paper 2023/142

Differential bunching impacts across the income distribution

Evidence from Zambian tax administrative data

Samuel Bryson,[1] Kwabena Adu-Ababio,[2] Evaristo Mwale,[3] and John Rand[1]

December 2023

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: We investigate the behavioural responses of individual taxpayers to changes in marginal personal income tax rates applying empirical bunching methodology to tax administrative data from Zambia over the period from 2014 to 2021. We find evidence for excess bunching at the first kink in the tax schedule for all years but less evidence of bunching at the second and third thresholds. While bunching is considerable and behavioural responses are observed to changes in the location of the kinks over time, bunching at reference points ('round-number bunching') also appears large. Implied elasticities of taxable income are however not remarkable, and comparing actual and estimated counterfactual wage distributions reveal that missed tax revenue arising from the excess bunching is limited. This is consistent with the observed bunching reacting sharply and immediately to changes in the location of the kink points over time, suggesting that observed behavioural change is driven by reporting behaviour rather than real economic responses.

Key words: tax bunching, personal income taxation, Zambia, income distribution, tax data

JEL classification: H24, H26, H31

Acknowledgements: The project team acknowledge the collaboration with the University of South Africa (UNISA), with a special thanks to PhD candidate Evaristo Mwale's supervisor Professor Robinson Zurika (UNISA). We are grateful for constructive comments and suggestions by Professor Jukka Pirtillä. We also thank the Illicit Financial Flows initiative at UNU-WIDER and participants at the WIDER Development Conference 'Revving up revenue for development—the role of domestic resource mobilization', 6–8 September 2023, Oslo, Norway, for useful comments and feedback.

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

[1] University of Copenhagen, Copenhagen, Denmark; [2] UNU-WIDER, Helsinki, Finland; [3] Zambia Revenue Authority, Lusaka, Zambia. Corresponding authors: samuel.bryson@lessing-ffm.net and evamwale@yahoo.com.

This study has been prepared within the UNU-WIDER project Detecting and countering illicit financial flows that is implemented in collaboration with the University of Copenhagen. The project is part of the Domestic Revenue Mobilization programme, which is financed through specific contributions by the Norwegian Agency for Development Cooperation (Norad).

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-450-2

https://doi.org/10.35188/UNU-WIDER/2023/450-2

Typescript prepared by Mary Boss.

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.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

A central challenge for developing and emerging economies is to generate sufficient tax income to secure a revenue base that can finance government developmental activities. Part of the domestic resource mobilization challenge is to ensure that any distortions or incentives for illicit behaviour created by the tax system are minimized and do not hinder economic activity. In order to understand the different welfare impacts and the efficiency of tax policy, it is therefore critical to have reliable evidence on the link between tax policy changes and the resulting behavioural responses of taxpayers. Changes in tax legislation could trigger real economic responses and affect the labour supply of individuals. At the same time, it could give rise to illicit behaviour in the form of reporting responses (evasion/avoidance). In developed countries, it has become common to use comprehensive taxpayer data to examine the consequences of changing tax policies, and the literature has been reviewed and summarized by Saez et al. (2012) or Kleven (2016). Studies that provide similar evidence using data from developing countries, however, remain relatively scarce.

In this paper, we therefore zoom in on the case of Zambia by investigating changes in tax thresholds in the personal income tax using the original bunching approaches developed by Saez (2010) and Chetty et al. (2011). The Pay As You Earn (PAYE) income tax in Zambia is a graduated system where tax liability increases progressively and each bracket is associated with a fixed marginal tax rate. This produces discontinuous jumps in tax liability at the cutoffs, and the kinks in the marginal tax rate therefore create strong incentives for bunching just below these thresholds. Using individual PAYE data for Zambia over the period 2014–21, our paper adds to a growing but still relatively limited empirical literature that applies bunching approaches to tax administrative data in developing countries. In addition, our paper addresses the important question on the impact of changing tax thresholds on overall domestic resource mobilization.

Methodologically, our paper is closest to Kleven and Waseem (2013) and Bell (2020) by analysing behavioural responses of taxpayers to discontinuous jumps in the personal income tax rate, accounting for issues related to the reference point problem. Some of the thresholds in Zambian personal income tax are round numbers and therefore present natural focal points for reasons other than the financial incentive ('round-number bunching'). Similar to optimization frictions such as switching costs and uncertainty, this implies that it can drive a wedge between the structural elasticity that is important for long-run welfare analysis and the observed elasticity that is estimated from the short-run variation in micro-data (Kleven and Waseem 2013). Unlike optimization frictions, such reference point effects amplify bunching and make the observed elasticity overstate the structural elasticity (Kleven 2016). We disentangle excess bunching at the kinks in the PAYE tax schedule from 'round-number bunching' responses, and like Bell (2020), our analysis relies on kinks in the personal income tax rate and makes use of the bunching approach developed by Saez (2010) and Chetty et al. (2011) and applies it to tax administrative data from a developing country.[1]

This is not the only study to take interest in developing countries. Kleven and Waseem (2013) apply the bunching approach to the personal income tax system of Pakistan and extend the original framework by looking at notches—discrete changes in the level of the choice sets of individuals or firms. Notches are conceptually different from kinks, which are discrete changes in the slope of the choice set (Kleven

[1] The surge in the use of the bunching approach in applied work in recent years can be explained by the increased availability of large administrative data sets. In his seminal contribution to this literature, Saez (2010) finds evidence of bunching at the first kink point of the US Earned Income Tax Credit (EITC) and shows that the compensated elasticity of reported taxable income can be estimated directly from the amount of bunching around the tax cutoffs. Chetty et al. (2011) provide an extension to this model by allowing for optimization frictions, specifically adjustment costs and hours constraints, that may prevent agents from bunching at kinks.

and Waseem 2013). In the context of taxes, this distinction corresponds to whether the discontinuity occurs in the marginal tax rate or in the average tax rate (Kleven and Waseem 2013). The authors find large excess bunching below every notch, and like Saez (2010), that bunching is larger for self-employed individuals than wage earners. Following Chetty et al. (2011), Kleven and Waseem (2013) also consider optimization frictions and find that absent frictions bunching would be 10 times larger for wage earners and 2–5 times larger for the self-employed. Importantly, the authors take into account reference points including those arising from 'number preferences'. Kinks and notches may represent natural focal points for taxpayers other than the financial incentive (e.g., if they are round numbers and hence constitute opposite optimization frictions). As a result, the observed elasticity may overstate the true structural elasticity (Kleven 2016).

Similar to Kleven and Waseem (2013), Bachas and Soto (2021) study notches where the average tax rate changes but in the case of the corporate tax system in Costa Rica. They find evidence of bunching below the thresholds and estimate larger elasticities than previous estimates for firms in developed countries. They argue that the response is driven partly by a change in revenue reporting but not by production responses.

Bell (2020) provides another example that considers the tax system of a developing country. The author investigates kinks in the South African marginal personal income tax rates and detects significant evidence of bunching only for the self-employed. The author finds small implied elasticities of taxable income from the bunching and that the responsiveness is due to both tax avoidance by income shifting and real labour supply responses. Another example from South Africa is Boonzaaier et al. (2019), where the bunching technique is also applied to corporate taxation. They discover significant bunching and large implied elasticities, providing further support to Bachas and Soto's (2021) result that elasticities may be larger for firms in developing countries than in developed countries.

Similar to Bell (2020), Bergolo et al. (2021) find only a small elasticity of taxable income at the first kink point (0.06) in their analysis of the personal income tax system in Uruguay. They find that the behavioural response is driven by a combination of labour income and deduction responses and that income is under-reported and deductions are used more frequently close to kink points. He et al. (2021), who investigate China's income tax schedule and use the fact that it has a graduated tax rate structure, find elasticity of taxable income estimates of between 0.09 and 0.41 for middle kinks and no evidence of bunching for bottom or top kinks.

Summarizing the developing country bunching literature illustrates several important points to consider when applying bunching methods in a developing country context: i) methodological distinction between kinks (marginal tax rate) and notches (average tax rate), ii) optimization frictions and reference points, including 'number preferences', iii) personal income tax: different behavioural effects for wage earners and the self-employed, and iv) distinction between changes in reporting behaviour and real responses (changes in labour supply/production).

In this study, we find significant evidence of excess bunching at the first kink in the PAYE schedule for all years over the period 2014–21 with an excess mass between 0.6 and 1.5. Since we detect bunching in all years, this indicates behavioural responses in adherence with the changes in the location of the kink. We find some indication for excess bunching at the second kink (but not such a strong response as at the first kink), and no evidence for bunching at the third (and highest) kink in the PAYE tax schedule. Throughout the period that we study, we observe 'round-number bunching' at natural focal points, but our excess bunching estimates remain significant after controlling for bunching at these reference points.

Our findings are in line with Boonzaaier et al. (2019) who also find that observed bunching reacts sharply and immediately to changes in the location of the kink points over time. This suggests that the behavioural response is driven by reporting responses rather than real economic responses. Real re-

sponses would result in a more scattered pattern around the kink thresholds due to inherent uncertainties in relation to real economic outcomes (e.g., adjustments in the labour supply). While this mechanism may be more relevant for firms, they may still be important in the context of workers, as adjustment of working hours of employees in response to changes in the PAYE tax rates may be more 'sticky' than adjusting reported income. According to Boonzaaier et al. (2019), reporting responses are less detrimental to welfare, compared to real economic responses, since the evasion/avoidance behaviour entails transfers to other economic actors.

Our results differ from those by Bell (2020) in the sense that the largest bunching responses are found at the highest kink, followed by the medium kinks and the smallest response at the lowest kink in the South African personal income tax schedule. We find the reverse in the case of Zambia. Moreover, for South Africa there is evidence for bunching only among self-employed workers and not for wage earners. We do not have information on whether workers are wage workers or self-employed in our data, but given that it is mostly firms that fill in the PAYE returns on behalf of their employees and self-employed workers typically file their returns under corporate income tax (CIT), we believe that our evidence of behavioural response applies to wage workers in the Zambian case. While Bell's (2020) finding of greater responses by self-employed workers is in line with what has typically been found in the literature (see, for example, Chetty et al. 2011; Kleven and Waseem 2013; Bastani and Selin 2014), our results fit with the findings of more recent papers that bunching responses are also observed for wage workers (Mortenson and Whitten 2020; Mavrokonstantis and Seibold 2022).

The findings from our study also fit with the results by Bachas and Soto (2021) in the sense that the behavioural response to the kinks is driven by a reporting response. Bachas and Soto (2021) find no evidence of production responses by Costa Rican firms. He et al. (2021), in contrast to our paper, do not find evidence of bunching at the bottom kink and only for the middle kinks in China's personal income tax schedule. Bergolo et al. (2021), like us, find evidence for bunching at the bottom kink in Uruguay, though. Moreover, they observe an increase in the amount of bunching over time, suggesting a learning process by individuals. While we do not see such an increase in bunching over time, we also see that individuals dynamically respond to changes in the location of the tax kinks in Zambia.

This article is organized as follows: Section 2 describes the institutional context, Section 3 develops the theoretical framework and empirical methodology as well as presents the data, Section 4 discusses the results, Section 5 contains robustness checks, Section 6 presents our estimations for the missed tax revenue, and Section 7 concludes.

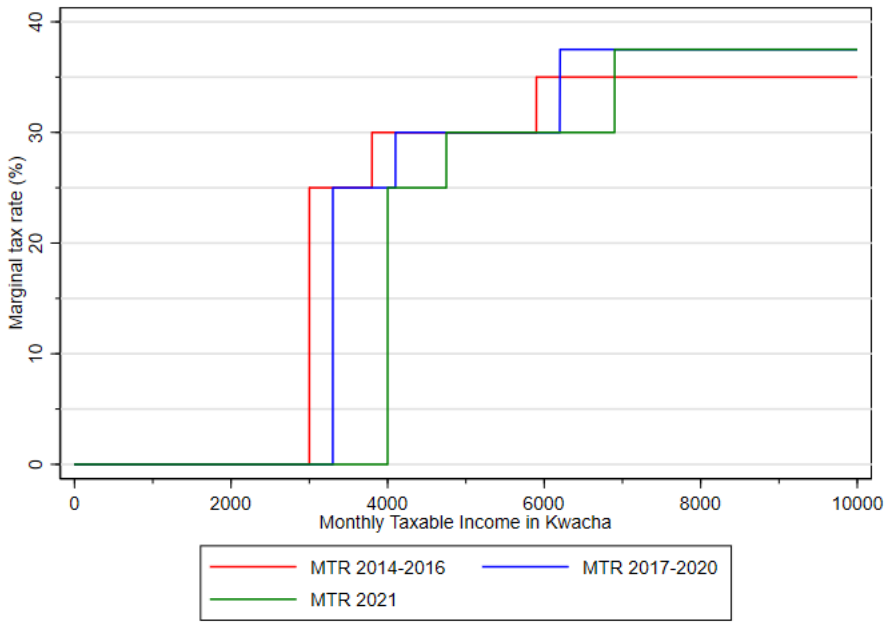
2 Institutional context

2.1 The Zambian personal income tax system

The tax system in Zambia is structured such that the Ministry of Finance and National Planning is responsible for the formulation of tax policy, and the implementing agency is the Zambia Revenue Authority (ZRA). Personal income tax is largely administered through the Pay As You Earn (PAYE) mechanism. PAYE is a method of deducting tax from employees' emoluments in proportion to what they earn. The system requires employers to calculate the tax payable by every employee, deduct tax due from the emoluments, and remit tax deducted to ZRA. As such, PAYE is administered as a withholding tax. Emoluments refer to the total earnings of an employee from employment, including wages, salaries, overtime, leave pay, commissions, fees, bonuses, and any other payments from employment or office (Section 2 of Income Tax Act (ITA) 2019) unless exempted by the ITA. Under the PAYE system, the amount of tax that the employer deducts from any pay depends on the employee's total gross pay and the applicable tax rates. The PAYE system of deducting tax from salaries and wages applies to all offices

and employments. Tax is deducted not only from monthly and weekly payments but also from daily, annual, or irregular payments; it applies to casual employees as well as full-time workers.

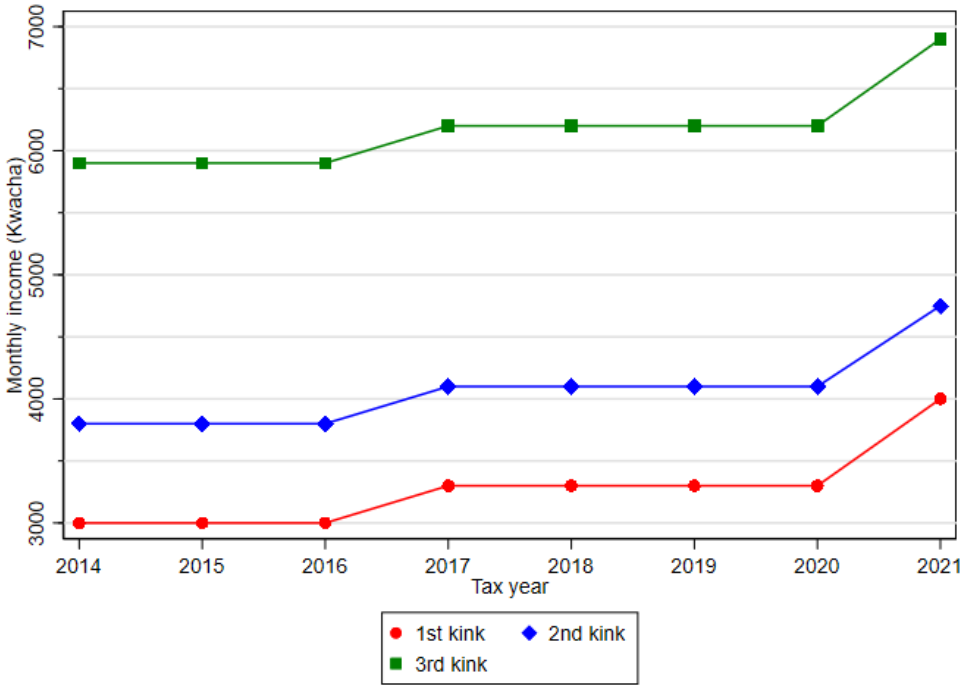
Figure 1: Personal income tax schedules in Zambia



Source: authors' illustration based on PAYE 2014–21 data.

The PAYE tax is designed as a graduated schedule with a fixed marginal tax rate in each bracket and therefore a kink (and not a notch) at each bracket cutoff. As depicted in Figure 1, the tax rate increases over three kinks from 0–25 per cent to 30 per cent and finally to 35 per cent before 2016 and to 37.5 per cent since 2017. These kinks create strong incentives since the tax rate jumps are large and they rise at higher income levels. The income level at which this first tax increase takes place—the value of the first kink point—has also increased steadily over the period 2014 to 2021. Figure 2 indicates that, over the period 2014–16, monthly income above ZMW3,000 became liable to tax. This increased to ZMW3,300 for the period 2017–20 and finally ZMW4,000 in 2021. The second kink in the PAYE schedule was at ZMW3,800 from 2014–16, increased to ZMW4,100 from 2017–20 and finally reached ZMW4,750 in 2021. The third and top kink was located at ZMW5,900, ZMW6,200, and ZMW6,900 during the same three time periods. In view of these kink locations, it is clear that special attention should be given to the first kink in 2014–16 and 2021 because of likely round-number bunching (ZMW3,000 and ZMW4,000, respectively) and therefore natural focal points for wage clustering other than the financial incentive. The first kink in 2017–20 and the second and third kinks are located such that we expect bunching due to financial incentives only.

Figure 2: Kink points in the Zambian personal income tax schedule (PAYE) 2014–21



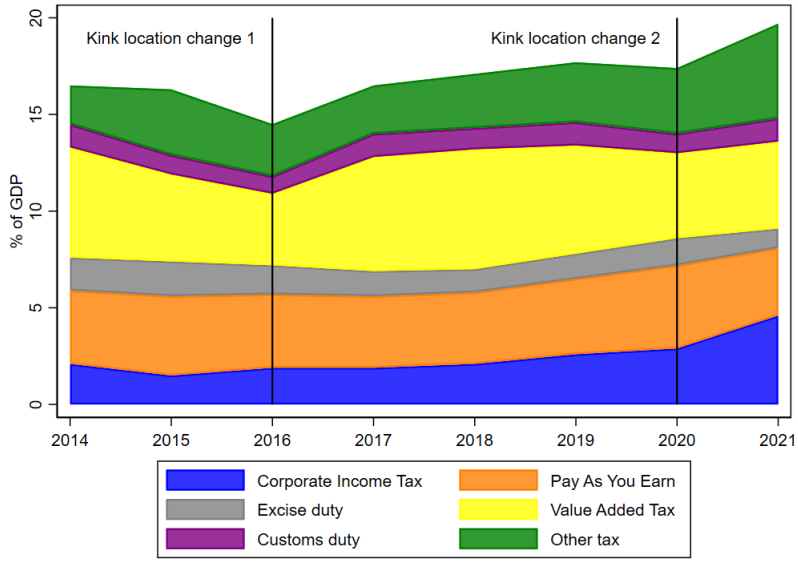
Note: the Zambia national poverty line is ZMW214 per adult per month (about US$1.53 per day in 2019 PPP terms). ZMW3,000 is US$979 in 2014 PPP and US$485 in 2021 PPP terms.

Source: authors' illustration based on PAYE 2014–21 data.

The personal income tax in Zambia currently raises revenue of 3.5 per cent of gross domestic product (GDP), or 18 per cent of total tax revenue (ZRA Annual Report 2021). The taxpayer population—those that have a Tax Payer Identification Number (TPIN)—in 2021 was 2,210,367, around 11 per cent of the total Zambian population (of around 19.5 million people). The number of taxpayers declaring PAYE increased from 511,445 in 2014 to 753,743 in 2021. Only VAT (22.6 per cent to total tax revenue) and tax revenue from company income tax (23.2 per cent of total) contributed more to total tax income than PAYE. Tax revenue accounted for 16 per cent of GDP in the beginning of the period, increasing to 19.7 per cent in 2021. However, the increase in 2021 is largely explained by increased revenues from mineral royalty tax and company mining tax as a result of increased copper prices on the global market. As indicated in Figure 3, the PAYE share has remained relatively constant during the period considered. To put this into context, the OECD PAYE average lies around 8 per cent over the same period.[2]

[2] Data extracted from OECD.Stat.

Figure 3: Tax type as a percentage of GDP 2014–21



Source: authors' illustration based on data from Zambia Revenue Authority (ZRA) and Tax Bulletin 2021.

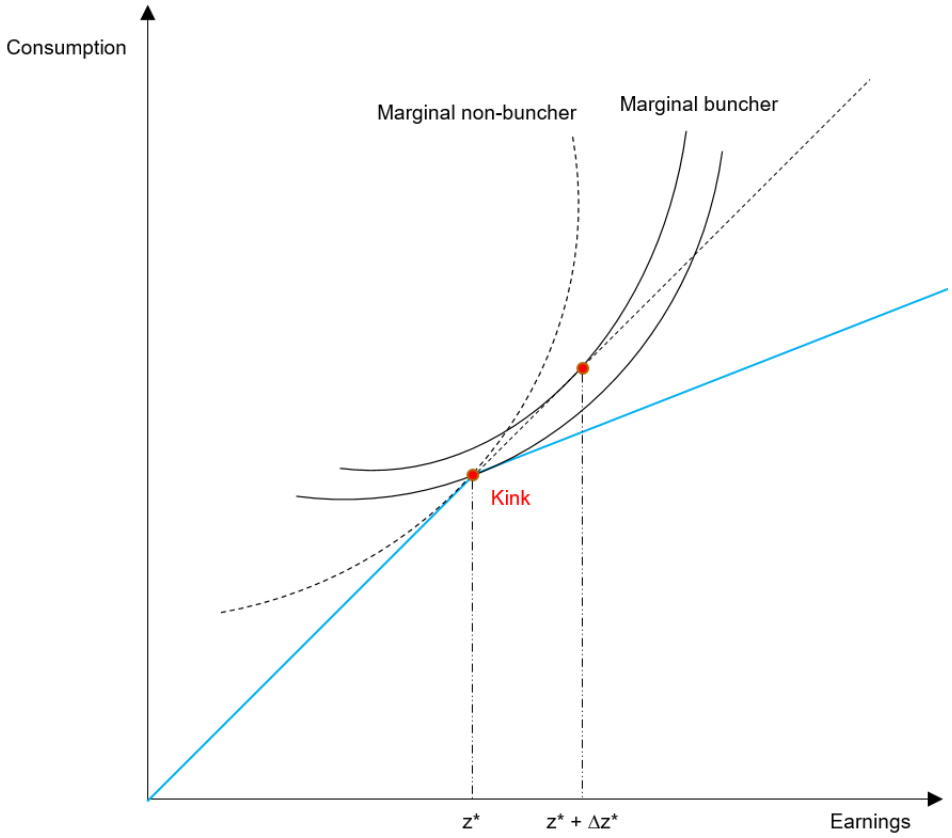
3 Methodology and data

3.1 Theory

The analysis of kink points created by discontinuities in marginal tax rates was initially developed by Saez (2010) and Chetty et al. (2011). The bunching methodology utilizes the predictions of a standard taxable income labour supply model. In the model, individuals' preferences are defined over after-tax income (consumption) and before-tax income (cost of effort). At the baseline, the tax system is smooth so that all individuals face the same marginal tax rate and individual optimization generates a smooth earnings distribution.

Suppose that a kink—a discrete increase in the marginal tax rate—is introduced at an earnings threshold z^*. Figure 4 illustrates the effects of this change in the tax rate. After the introduction of the kink, the individual initially located at $[z^* + \Delta z^*]$ is tangent to the kink point $z*$ and therefore moves down to the kink. This is the marginal bunching individual. All individuals initially located in the earnings interval $[z^*, z^* + \Delta z^*]$ move to the kink point z^*. Those individuals initially located above this interval reduce their earnings but stay in the interior of the upper bracket and do not move all the way to the kink point. All individuals earning below z^* continue to face the baseline marginal tax rate and, thus, absent any changes in incentives, the earnings distribution to the left of the kink is unaffected. Taken together, these responses produce excess bunching in the earnings distribution at the kink point. It does not produce a hole in the distribution above the kink, though, given the earnings response of the individuals initially located above the interval $[z^*, z^* + \Delta z^*]$ to the higher marginal tax rate who reduce their earnings and fill up the hole.

Figure 4: Bunching at a kink



Source: authors' illustration based on Kleven (2016).

3.2 Bunching estimation

The bunching estimation relies on calculating the earnings distribution that would have been observed in the absence of any kinks, i.e. the counterfactual earnings distribution. The standard approach, developed by Chetty et al. (2011), is to fit a flexible polynomial to the observed income distribution omitting observations located in a range around the kink and then extrapolate the fitted distribution to the threshold. More specifically, the counterfactual distribution is estimated by grouping individuals into earnings bins indexed by j and using a regression of the following form:

$$c_j = \sum_{i=0}^{\rho} \beta_i \cdot (z_j)^i + \sum_{r \in R} \rho_r \cdot 1[\frac{z_j}{r} \in \mathbb{N}] + \sum_{i=z_-}^{z_+} \gamma_i \cdot 1[z_j = i] + v_j \tag{1}$$

where c_j is the number of individuals in bin j, z_j is the earnings level in bin j, $[z_-, z_+]$ is the excluded range, and ρ is the order of the polynomial. The excluded range here is a narrow symmetric range around the threshold $z*$ ($[z_-, z_+] = [z^* - d, z^* + d]$). Both z_j and z are measured in units of the bin width, d. We use a seventh order polynomial as suggested by Chetty et al. (2011). We also use different bin widths to adjust the range of excluded data on either side of the kink point and obtain a more robust result.

In our case, we also need to control for round-number bunching since some kinks are located at a round number (e.g., the first kink between 2014 and 2016 is located at ZMW3,000). The tendency of taxpayers to report taxable income in round numbers creates mass points at round numbers in the empirical distribution (Kleven and Waseem 2013). This implies that typically obtained smooth counterfactuals (see, for example, Bell 2020) would be biased in our case—it would be overstating the behavioural responses to the kinks (Kleven 2016). To solve this issue, we follow Kleven and Waseem (2013) and

control for round-number bunching at kinks by using excess bunching at similar round numbers that are not kinks as counterfactuals. To construct these round number counterfactuals, we account for the patterns of rounding by estimating a set of round-number fixed effects and include these in (1). We ensure that these round-number fixed effects are able to account for differences in round numbers (some are rounder than others and therefore induce stronger bunching). In (1), $\mathbb{N}$ is the set of natural numbers, $R = 1.5K, 2K, 2.5K, 5K$ is a vector of round-number multiplies that capture monthly rounding. The counterfactual bin counts are obtained as predicted values from regression (1) omitting the effects of the dummies in the excluded range but not omitting the contribution of round-number dummies. The extent of excess bunching can then be calculated by taking the difference between the observed and counterfactual bin counts in the bunching range. Following Chetty et al. (2011), standard errors are estimated using a bootstrap procedure that generates a large number of earning distributions by randomly re-sampling the residuals from the above equation.

If we assume that the heterogeneity distribution of individuals is uniform around the kink, then we can relate our estimate of excess bunching to the compensated elasticity of taxable income, $e(z*)$, locally at the kink point $z*$ using the following equation:

$$e(z^*) = \frac{\hat{b}}{z^* \times log(\frac{1-t_1}{1-t_2})} \tag{2}$$

where $\hat{b}$ is the excess bunching mass, t_1 is the initial tax, and t_2 is the adjusted tax.

Importantly, the elasticity in the above equation cannot be treated as structural elasticity. Since we control for round-number bunching, but there may be optimization frictions such as search costs and uncertainty that prevent individuals from bunching, our estimated elasticity $e(z*)$ is likely to be smaller than the structural elasticity.

3.3 Data

The administrative personal income data

Our study is based on tax data from the Zambia Revenue Authority (ZRA) and covers the universe of personal income (PAYE) tax returns in Zambia filed between 2014 and 2021. It includes data on the employer TPIN, employee ID, the return year and month, gender, age, the sector of the firm, the jurisdiction that the firm falls under, gross emoluments, chargeable emoluments, and the total tax deducted. We do not have information on whether individuals are wage workers or self-employed. Therefore, we cannot test the typically found result in the literature for the case of Zambia that estimated elasticities are larger for self-employed workers. In total, the data set includes 44,768,466 observations of employee IDs. From 2014 to 2016, there are around 4,700,000 observations each year. Between 2017 and 2021, the number of observations increased from 5,300,000 to 6,500,000. This is also reflected in increasing monthly average observations over the period 2014–21, from around 400,000 in 2014–16 to around 450,000 in 2017, 500,000 in 2018–19, and the highest average in 2021 with 560,000.

The variable of interest, in our case, is chargeable emoluments, which under the PAYE system refers to emoluments from an employee's employment that are chargeable to income tax but do not include any amount that is exempt from income tax. Emoluments that are exempt, or otherwise not chargeable to income tax, and are therefore not included in the chargeable emoluments from which PAYE tax is to be deducted include medical expenses, accommodation provided by the employer, and labour day awards, among other things.

Cleaning steps

Two aspects of the data had to be adjusted: (i) creating a unique ID and (ii) identifying outliers. The data set includes both employer and employee IDs. The former is mandatory to fill in the PAYE returns and is a unique computer-generated 12-digit number (TPIN) that was allocated to the firm when it registered formally. The latter is not mandatory to fill in and is filled in by the firms on behalf of their employees. It is typically an eight-digit ID, but some firms chose other numbers including the National Registration Card (NRC) number. In this context, in a first step, we dropped all the employee IDs that are different from an eight-digit number. As indicated in Table 1, the dropped IDs are less than 1 per cent from 2014 to 2019. However, the number increased to 8.45 per cent in 2020 and 9.71 per cent in 2021. After this step, we combined the employer and employee IDs to generate a new unique 12-digit ID.

In regards to chargeable emoluments, we simply dropped outliers, i.e. values below the 1st percentile and above the 99th percentile. As shown in Table 1, we drop only very few observations below the 1st percentile since this is naturally bound at zero usually, as no reported negative chargeable emoluments are observed. This means that the majority of observations dropped are above the 99th percentile. After the cleaning process, we are left with more than 98 per cent, and closer to 99 per cent in most years, of the raw observations, with the exception of 2020 and 2021, where we retain around 90 per cent of the data.

Table 1: Overview of cleaning steps and observations dropped

Year	Raw	Drop ID	% of raw	Drop >p99	Clean	% of raw
2014	4,653,865	5,793	0.12	46,479	4,601,585	98.88
2015	4,770,504	5,761	0.12	47,647	4,717,096	98.88
2016	4,650,100	5,318	0.11	46,447	4,598,335	98.89
2017	5,301,171	2,618	0.05	52,985	5,245,568	98.95
2018	5,805,011	277	0.00	58,046	5,746,688	99.00
2019	5,933,279	31,158	0.53	59,021	5,843,100	98.48
2020	6,204,777	524,542	8.45	56,802	5,623,445	90.63
2021	6,482,701	629,247	9.71	58,534	5,794,920	89.39

Source: authors' calculations based on PAYE 2014–21 data.

Clean data

Table 2 indicates that there are no values of chargeable emoluments below zero in the cleaned data set. The highest emoluments were reported in 2020 with ZMW100,450. In the other years, the maximum chargeable emoluments lie at ZMW96,700 or lower. We find the highest mean level of income, as measured by chargeable emoluments, in 2020 was ZMW7,275. The mean level was below ZMW6,000 in 2014 and 2015, increased to below ZMW7,000 from 2016–19, and since 2020 lies above ZMW7,000. Similarly, the median value of chargeable emoluments, below which 50 per cent of the taxpayers declaring PAYE find themselves, increased since 2017 to ZMW2,976 in 2021. In all years, the median level lies below the first kink point, which means that at least 50 per cent of the taxpayer population declaring PAYE do not have to pay tax. In fact, over the last years, around 54–60 per cent of employees have typically belonged to the first income tax band paying 0 per cent (%K1), 4–7 per cent belonged to the second (%K2), 8–10 per cent to the third income tax band (%K3), and the remainder, 27–30 per cent, paid the top tax rate (%>K3).

Table 2: Chargeable emoluments: summary statistics

Year	Obs	% of LF	Min	Max	Mean	Median	% 0	% K1	% K2	% K3	% >K3
2014	4,601,585	6.03	0.00	83,789	5,572	2,249	5.37	57.28	5.51	9.88	27.33
2015	4,717,096	5.96	0.00	87,845	5,846	2,375	5.55	56.24	5.62	9.96	28.18
2016	4,598,335	5.61	0.00	94,975	6,418	2,566	6.16	54.83	5.91	9.86	29.41
2017	5,245,568	6.17	0.00	94,964	6,358	2,390	5.03	58.83	5.14	8.89	27.13
2018	5,746,688	6.52	0.00	96,683	6,694	2,600	4.98	57.26	5.64	8.99	28.12
2019	5,843,100	6.40	0.00	88,825	6,701	2,800	5.11	55.59	6.43	9.31	28.68
2020	5,623,445	5.99	0.00	100,450	7,275	2,953	5.98	54.21	6.63	9.00	30.17
2021	5,794,920	5.95	0.00	93,042	7,087	2,976	6.15	60.45	4.39	7.74	27.42

Note: average number of observations per month ranges from around 383,000 (in 2016) to around 478,900 (in 2018). Share of the labour force (LF) is based on this monthly average. The labour force comprises people ages 15 and older and is taken from the World Development Indicators.
Source: authors' calculations based on PAYE 2014–21 data.

Within each tax return year, there is no large variation in regards to the mean and median chargeable emoluments. The average reported chargeable emoluments are typically lowest in January and highest in December of each year, due to extra bonuses and gratuity at the end of the year. The picture is similar when looking at the quarterly numbers. The variation is not very large, and the lowest emoluments are reported in the first quarter while the highest are reported in the fourth quarter.

4 Results

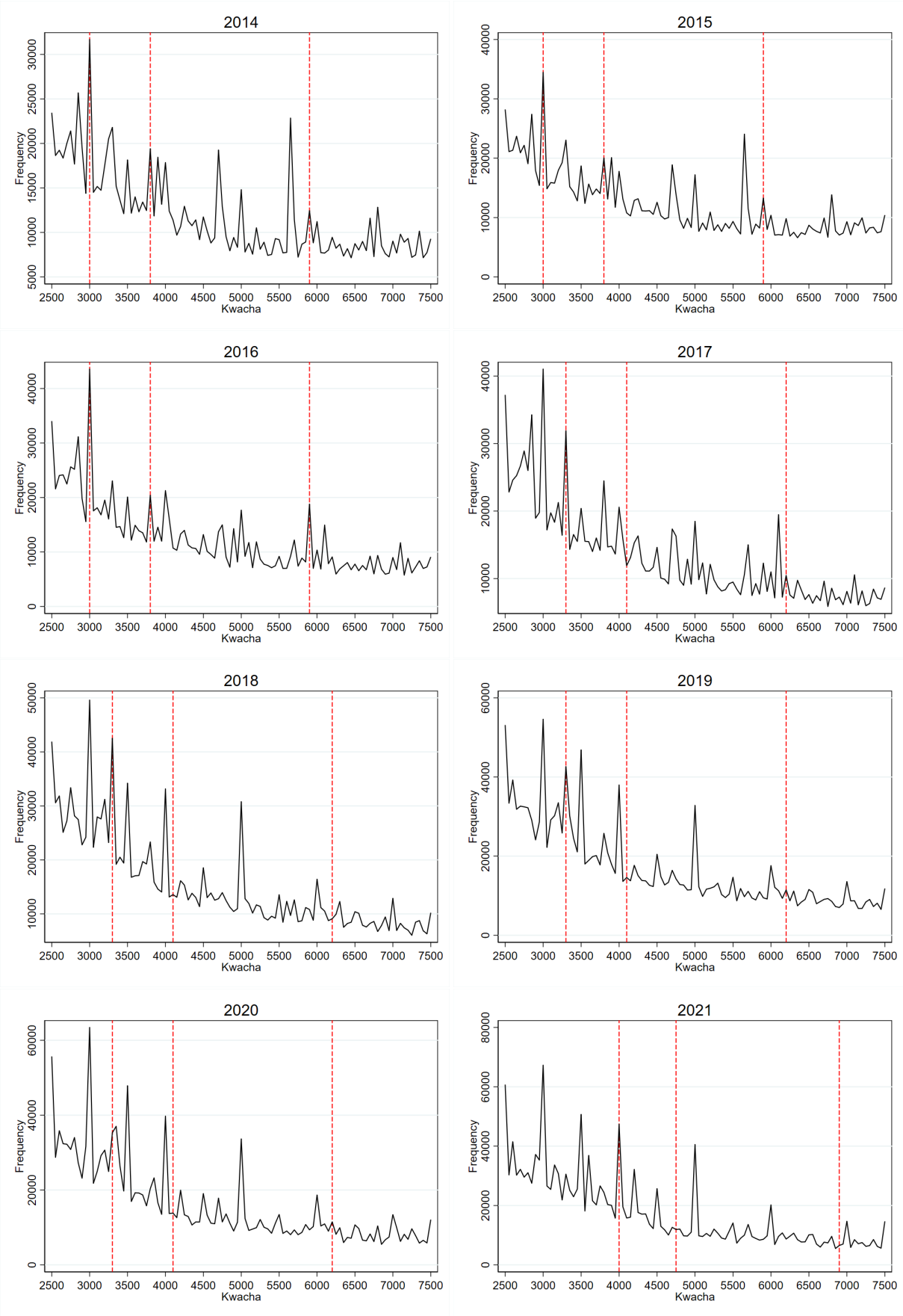
4.1 Graphical evidence

In order to obtain a first overview of the income distribution and to see whether any bunching behaviour can be discerned, we plot the income distribution in earnings bins with a width of 50 using line plots. The red lines indicate the location of the kinks. As explained previously, the location of the kinks changed in 2017 and again in 2021.

Figure 5 provides a first indication that there is excess bunching at the first kink and indicates that bunching is the most pronounced at this kink. From 2014 to 2016, the first kink is located at ZMW3,000, which makes it difficult to distinguish any bunching behaviour arising for tax reasons from the 'round-number-bunching' behaviour. From 2017 to 2020, however, when the first kink changes to ZMW3,300, the excess bunching becomes visible. Now, both the round-number bunching and the bunching at the kink become visible as two separate peaks. In 2021, the same issue as for the period 2014–16 arises since the first kink is now located at the round number ZMW4,000. Overall, bunching at the second and third tax kinks appear to be not as significant, if at all. In the years 2014–16, some bunching can be seen at the second kink and to a lesser and very small extent at the third kink. However, this disappears entirely for the years 2017–21.

The closer look at the distribution shows the existence of 'round-number bunching' at other round numbers. For example, clear candidates are 3,000, 4,000, 5,000, etc. There are also some other peaks at less round numbers such as 3,500 and 4,500. This becomes the most clear from the year 2018 onwards.

Figure 5: Earnings distributions with kinks 2014–21



Note: earnings distribution in bins with a bin width of 50. Dashed vertical red lines depict the first, second, and third tax kink, respectively.
Source: author's illustration based on PAYE 2014–21 data.

4.2 Main results

Since the graphical evidence suggests that the greatest responses are produced by the first tax kink in the Zambian PAYE schedule, we provide line plots with the counterfactual distribution as well as the estimates of the excess mass and the elasticity for this kink for selected years. Figure 6 presents a visual representation of the excess bunching around the first kink for the years 2016, 2017, 2019, and 2021. The blue line plots the empirical distribution of taxable income around the kink point (ZMW3,000 in 2016, ZMW3,300 in 2017 and 2019, and ZMW4,000 in 2021), while the red line represents the seventh order fitted polynomial, which excludes the observations in the small window of $[-ZMW50; +ZMW50]$ around the kink point and is estimated over a large window of ZMW10,000.

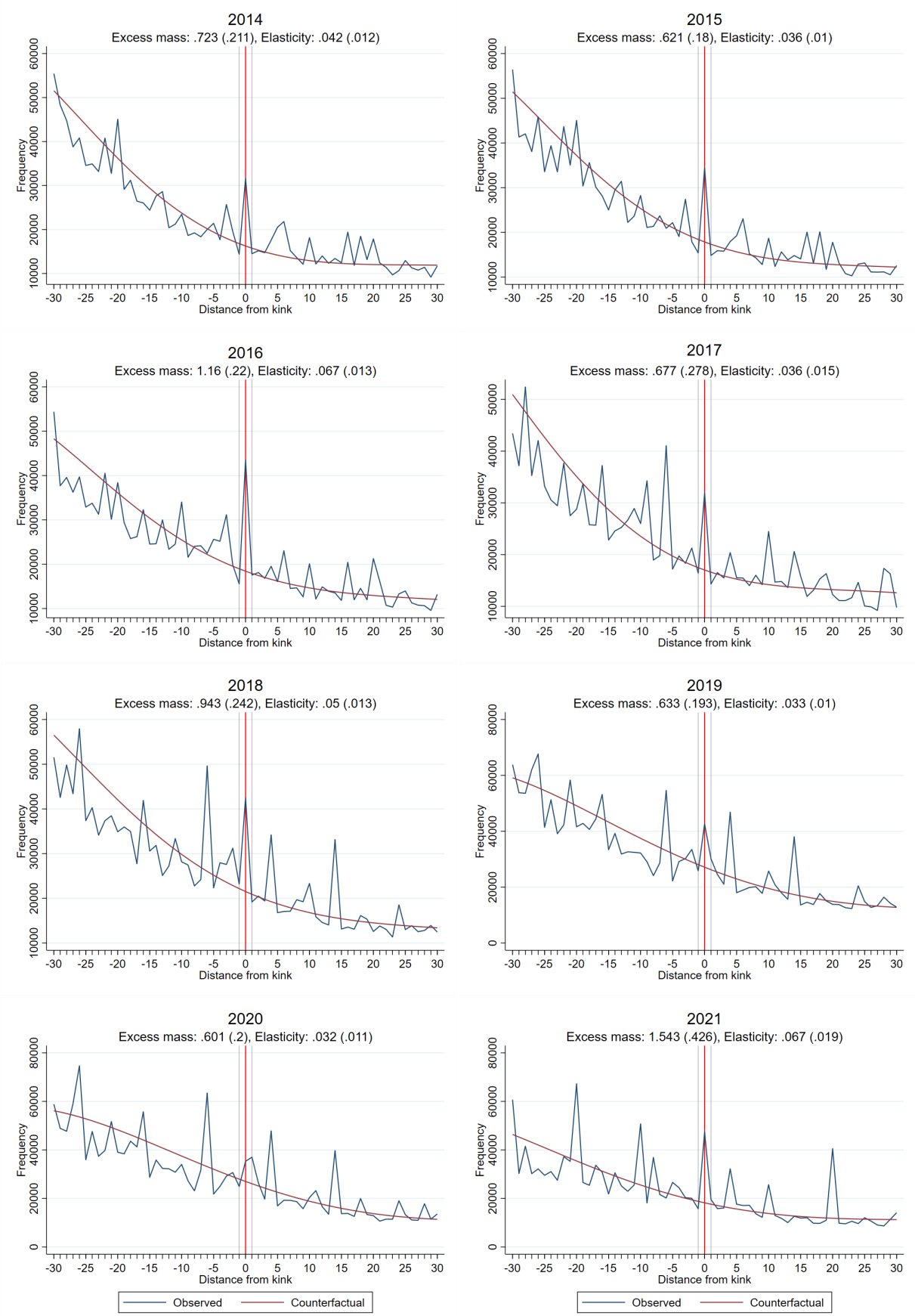
Over the selected years, it becomes evident that registered taxpayers in Zambia exhibit strong bunching behaviour at the first kink in the tax schedule, and the estimates of excess bunching and elasticity are positive and significant for all years over the period 2014–21.

Figure 6 illustrates that the excess bunching around the first kink remains significant over the entire sample period despite the changes in the location of the first kink. The estimates of excess bunching range from 0.601 to 1.543 and are significant at the 1 per cent level for all years besides 2017—here, at the 5 per cent level. Excess bunching at the first kink is not obviously larger in the period 2014–16 than in the period 2017–20, even though the kink represents a round number in the previous period. This strongly suggests that wages in Zambia are not only set according to natural focal points such as round numbers but also with a view to optimizing income by keeping them at or below tax kinks. Moreover, there may be a learning process by taxpayers in the sense that the estimate of excess bunching initially drops after the kink revision in 2017—from 1.16 in 2016 to 0.677 in 2017—but then increases again to 0.943 in 2018. In other words, the information about the new location of the kink may take some time to disseminate and taxpayers over time adjust to this. Overall, the largest estimates are generated for years where the first tax kink also represents a round number. In 2016, the estimate is 1.16, and the largest estimate is obtained for 2021, where the first tax kink is at ZMW4,000, with a value of 1.543.

Despite the significant observed bunching, the implied elasticities are not large. The elasticity for Zambian taxpayers in 2014 and 2021, where the greatest bunching is observed in Figure 6, is only 0.067 in both years. Overall, the elasticity estimates for the first kink range from 0.032 to 0.067.

Table 3 presents the estimates of excess bunching for all three kink points in the income tax distribution. In regards to the second and third kinks, the estimates confirm the graphical evidence discussed previously. We only find significant estimates of excess bunching at the second and third kinks for the period 2014–16 before the kink location adjustment. There is significant (at the 5 per cent level) excess bunching of around 0.57 at the second kink in 2014 and 2015, and this is matched by the visible evidence of excess bunching in Figure 5. This, however, decreases and turns insignificant in 2016 and after that even turns negative, i.e. the observed distribution lies below the predicted counterfactual income distribution. For the third kink, we only find significant and relatively large excess bunching of 0.974 in 2016, and this again becomes visible in Figure 5. In the two years before, there is no significant bunching, and afterwards, the estimates turn negative, as with the second kink.

Figure 6: Bunching at the first kink in the income tax schedule



Source: authors' calculations based on PAYE 2014–21 data.

Table 3: Estimates of excess bunching at three kink points in the income tax distribution

	First kink		**Second kink**		**Third kink**	
Year	b	se	b	se	b	se
	(1)	(2)	(3)	(4)	(5)	(6)
2014	0.723***	(0.211)	0.568**	(0.269)	0.252	(0.341)
2015	0.621***	(0.180)	0.575**	(0.282)	0.246	(0.380)
2016	1.160***	(0.220)	0.207	(0.242)	0.974**	(0.494)
2017	0.677**	(0.278)	-0.027	(0.279)	-0.068	(0.448)
2018	0.943***	(0.242)	-0.490**	(0.236)	-0.254	(0.398)
2019	0.633***	(0.193)	-0.536**	(0.210)	-0.389	(0.349)
2020	0.601***	(0.200)	-0.488**	(0.241)	-0.339	(0.385)
2021	1.543***	(0.426)	-0.232	(0.326)	-0.944**	(0.391)

Note: *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations based on PAYE 2014–21 data.

Table 4 depicts the elasticity estimates for all three kinks. In line with the bunching estimates, positive and significant elasticity estimates are obtained only for the years 2014 and 2015 for the second kink and for 2016 for the third kink. These are significant at the 5 per cent and 1 per cent level for the second and third kink, respectively, and compared to the estimates for the first kink are very large. For the second kink, the elasticity is 0.108 and 0.11 in 2014 and 2015, respectively. For the third kink, the elasticity is 0.111 in 2016—the highest elasticity obtained overall.

Table 4: Elasticity estimates at three kink points in the income tax distribution

	First kink		**Second kink**		**Third kink**	
Year	e	se	e	se	e	se
	(1)	(2)	(3)	(4)	(5)	(6)
2014	0.042***	(0.012)	0.108**	(0.052)	0.029	(0.039)
2015	0.036***	(0.010)	0.110**	(0.054)	0.028	(0.044)
2016	0.067***	(0.013)	0.039	(0.046)	0.111*	(0.057)
2017	0.036**	(0.015)	-0.005	(0.049)	-0.005	(0.032)
2018	0.050***	(0.013)	-0.087**	(0.042)	-0.018	(0.028)
2019	0.033***	(0.010)	-0.095**	(0.037)	-0.028	(0.025)
2020	0.032***	(0.011)	-0.086**	(0.043)	-0.024	(0.027)
2021	0.067***	(0.019)	-0.035	(0.050)	-0.060**	(0.025)

Note: *** p<0.01, ** p<0.05, * p<0.1.

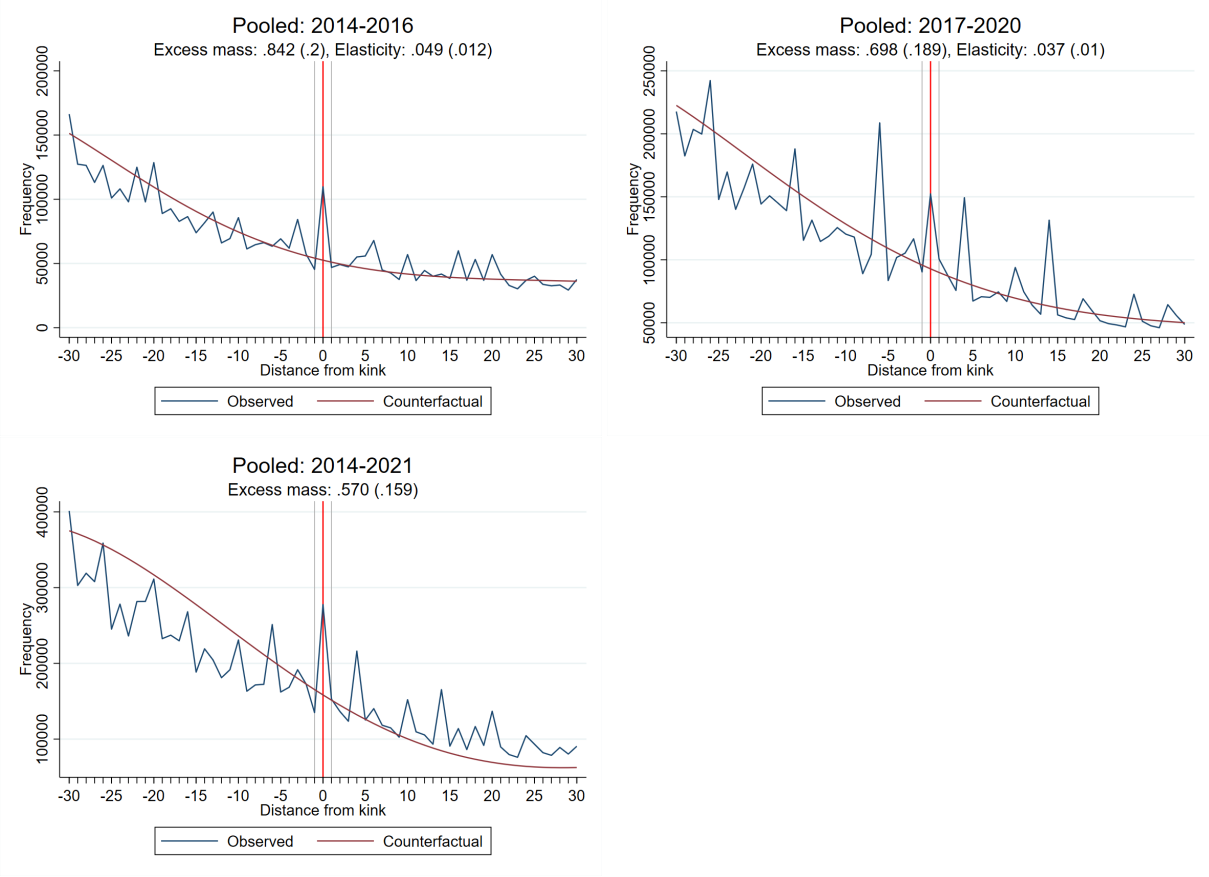
Source: authors' calculations based on PAYE 2014–21 data.

We also calculate the estimates of excess mass and elasticity after combining the data for the different years. We combine the data sets by first re-defining the taxable income variable to reflect the distance from the bracket cut-off point in that year so that it takes on a value of zero at the kink point and then pooling the data.

Figure 7 gives a visual representation of the excess mass and the elasticity for the first kink for three different pooled data sets—2014–16 (before the first change in the kink location), 2017–20 (after the change and before the second change in the kink location), and 2014–21 (the entire period). Across all three periods, we confirm the results discussed above. The excess bunching is significant at the 1 per cent level and has a value of 0.842 for the period 2014–16, 0.689 for the period 2017–20, and 0.57 overall. In terms of magnitude, this is in line with the results presented in Table 3 as it constitutes roughly the average of the estimates of excess bunching for the individual years. When looking at the entire period, the excess bunching estimate is slightly lower.

As before, the implied elasticities are not large but well determined, with a value of 0.49 and 0.37 for the period 2014–16 and 2017–20, respectively. Due to the kink location changes, we were not able to calculate the elasticity for the combined data set for the entire period.

Figure 7: Bunching at the first kink in the income tax schedule—pooled data



Source: authors' calculations based on PAYE 2014–21 data.

Table 5 depicts the bunching and elasticity estimates for all three kinks for the pooled data sets. We find significant excess bunching of 0.459 and a corresponding significant elasticity estimate of 0.088 for the second kink in the period 2014–16. While we find significant excess bunching and a significant elasticity for the third kink in the year 2016 individually, this turns insignificant when using the combined data set for the period 2014–16. This makes sense, given that the estimates for the third kink in 2014 and 2015 are insignificant. For the period 2017–20, we find both a significant negative excess bunching and elasticity estimate for the second kink, which matches the result found for the years 2018, 2019, and 2020 in Tables 3 and 4. The estimates for the third kink for this period are insignificant. When looking at the entire period, the estimates of excess bunching are significant for both the second and third kinks, with a value of 0.687 and -0.746, respectively.

These estimates are again in line in terms of magnitude with the results in Tables 3 and 4 and have a value corresponding roughly to the average of the estimates for the individual years.

Table 5: Elasticity estimates at three kink points in the income tax distribution—pooled data

	Excess mass		**Elasticity**	
Year	b	se	e	se
	(1)	(2)	(3)	(4)
A: 2014–16				
Kink 1	0.842***	(0.200)	0.049***	(0.012)
Kink 2	0.459*	(0.248)	0.088*	(0.047)
Kink 3	0.479	(0.388)	0.055	(0.044)
B: 2017–20				
Kink 1	0.698***	(0.189)	0.037***	(0.010)
Kink 2	-0.404**	(0.203)	-0.071**	(0.036)
Kink 3	-0.269	(0.315)	-0.019	(0.022)
C: 2014–21				
Kink 1	0.570***	(0.159)	-	-
Kink 2	0.687**	(0.340)	-	-
Kink 3	-0.746***	(0.182)	-	-

Note: *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations based on PAYE 2014–21 data.

4.3 Comparison of main results with other studies

Our findings are the reverse of those found by Bell (2020) in the sense that we detect the greatest bunching at the lowest kink in the Zambian personal income tax schedule and the smallest bunching behaviour at the highest kink. Our estimates of excess bunching are also slightly higher with around 0.6 to 1.5 at the first kink, while Bell (2020) finds an excess mass of around 0.1 to 0.7 at the highest kink in the South African personal income tax schedule. Our estimates of excess mass are also larger than those found by Bergolo et al. (2021) but relatively small compared to those found by Kleven and Waseem (2013) of 1.7 to 5.5 across the different notches.

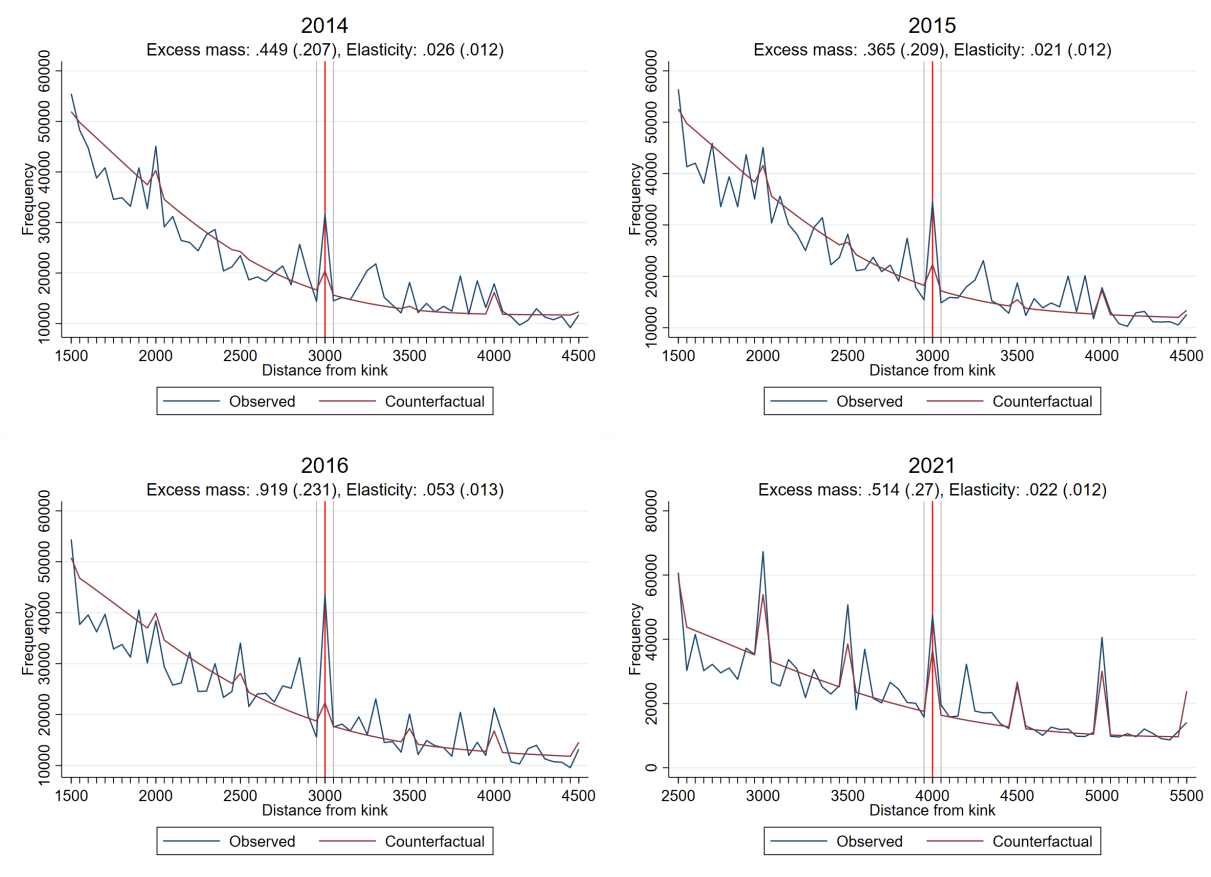
In line with these other studies, our estimates of the implied elasticity are relatively small. The implied elasticities in Chetty et al. (2011), Bastani and Selin (2014), and Bell (2020), who use a similar estimation method as we do, are 0.02, 0.24, and 0.08, respectively. Our elasticity estimate for the year where we detect the largest excess mass is 0.067, which fits with the results from these other studies. The elasticity estimates found by Kleven and Waseem (2013) and Bergolo et al. (2021) fall into this range of magnitude and fit with our overall estimates.

4.4 Round-number bunching

In this section, we account for potential bias arising from round-number bunching at the first tax kink in the years 2014–16 and 2021. We follow Kleven and Waseem (2013) by estimating a set of round-number fixed effects that depend on the degree of the roundness (multiples of 1,000 or 500).

Figure 8 shows the excess mass around the first kink in these years as well as the counterfactual distribution that has been estimated while including round-number fixed effects at multiples of 1,000 or 500. It becomes clear that the largest round-number bunching appears in 2021—this is where the spike in the counterfactual distribution is largest.

Figure 8: Accounting for round-number bunching at the first kink in 2014, 2015, 2016, and 2021



Source: authors' calculations based on PAYE 2014–21 data.

Table 6 shows how the main results for the excess mass and elasticity at the first kink in the affected years compare with the results after controlling for round-number bunching. Only the estimates for 2016 remain significant at the 1 per cent level and only decrease by a small amount. The excess mass decreases from 1.16 to 0.919 and the elasticity from 0.067 to 0.053. For the years 2014 and 2015, the estimates of the excess mass shrink by around 0.25 points and the elasticity estimates by around 0.015 points. They also become less significant but remain statistically significant. The estimates for 2021, the year where we detected the greatest excess mass, decrease drastically from 1.543 to 0.514 and 0.067 to 0.022 in terms of the elasticity and also become less significant. This is in line with the visual evidence in Figure 8.

These results suggest that, even though bunching behaviour at the first kink is the greatest in 2021, this is driven to a large extent by round-number bunching. We find the strongest evidence of bunching for tax reasons for the year 2016. Here the excess mass and elasticity estimates do not change much after controlling for round-number bunching.

Table 6: Accounting for round-number bunching at the first kink

	d = 50		**RN**	
Year	b	e	b	e
	(1)	(2)	(3)	(4)
2014	0.723***	0.042***	0.449**	0.026**
	(0.211)	(0.012)	(0.207)	(0.012)
2015	0.621***	0.036***	0.365*	0.021*
	(0.180)	(0.010)	(0.209)	(0.012)
2016	1.160***	0.067***	0.919***	0.053***
	(0.220)	(0.013)	(0.231)	(0.013)
2021	1.543**	0.067***	0.514*	0.022*
	(0.426)	(0.019)	(0.270)	(0.012)

Note: *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations based on PAYE 2014–21 data.

5 Robustness check

In this section, we test the sensitivity of our results to different bin widths and estimation windows. We also tried to test the robustness of our results to keeping the outliers above the 99th percentile. However, these results turned out to be flawed, since the entries above the 99th percentile were implausibly large and thus problematic and were therefore not included.

Table 7 presents excess mass and elasticity estimation results for three other bin widths—10, 25, and 100 (two smaller and one larger than the baseline bin width)—over the whole sample period. Columns (1) and (2) use a bin width of 50 and hence reproduce the baseline result.

The value of the estimates of excess bunching depend on the bin width, so the results in columns (3), (5), and (7) should be different to the baseline result in column (1) of Table 7. But they should be qualitatively similar, and the elasticity estimates should remain similar. This is indeed what we find. We find a greater excess mass using a smaller bin width, and vice versa. The excess mass estimate for the first kink in 2016 increases to 5.667 and 2.393 using a bin width of 10 and 25, respectively, and decreases to 0.69 using a bin width of 100. The elasticity estimate varies only by a small amount—0.001 points when using the smaller bin widths and 0.011 points when using the bin width of 100. The similar effect can be seen for the first kink in 2021, where we obtained the largest overall estimate of excess mass. This increases to 7.979 when using a bin width of 10 but decreases to 0.874 for a bin width of 100.

While the excess mass results change quantitatively, they remain similar qualitatively in the sense that the significance of the obtained result does not change. The only small effect is that some results for the second and third kinks that are significant at the 5 per cent level are even better determined when using a smaller bin width. Moreover, the elasticity estimates do not vary much in magnitude and keep their same significance levels.

Table 7: Robustness check with different bin widths

	d = 50		**d = 10**		**d = 25**		**d = 100**	
Year	b	e	b	e	b	e	b	e
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2014								
Kink 1	0.723***	0.042***	3.895***	0.045***	1.488***	0.043***	0.447***	0.052***
	(0.211)	(0.012)	(0.680)	(0.008)	(0.312)	(0.009)	(0.095)	(0.011)
Kink 2	0.568**	0.108**	3.004***	0.115***	1.162***	0.111***	0.274**	0.105**
	(0.269)	(0.051)	(0.970)	(0.037)	(0.448)	(0.043)	(0.120)	(0.046)
Kink 3	0.252	0.029	1.203	0.028	0.549	0.031	0.159	0.036
	(0.341)	(0.039)	(1.046)	(0.024)	(0.525)	(0.030)	(0.143)	(0.033)
2015								
Kink 1	0.621***	0.036***	3.606***	0.042***	1.315***	0.038***	0.406***	0.047***
	(0.180)	(0.010)	(0.608)	(0.007)	(0.299)	(0.009)	(0.072)	(0.008)
Kink 2	0.575**	0.110**	2.968***	0.113***	1.194***	0.114***	0.254***	0.097***
	(0.282)	(0.054)	(0.954)	(0.036)	(0.393)	(0.038)	(0.086)	(0.033)
Kink 3	0.246	0.028	1.503	0.034	0.642	0.037	0.188	0.043
	(0.380)	(0.044)	(1.189)	(0.027)	(0.548)	(0.031)	(0.134)	(0.031)
2016								
Kink 1	1.160***	0.067***	5.667***	0.066***	2.393***	0.069***	0.690***	0.080***
	(0.220)	(0.013)	(0.768)	(0.009)	(0.332)	(0.010)	(0.089)	(0.010)
Kink 2	0.207	0.039	1.484*	0.057*	0.541	0.052	0.183**	0.070**
	(0.242)	(0.046)	(0.793)	(0.03)	(0.396)	(0.038)	(0.092)	(0.035)
Kink 3	0.974**	0.111*	5.421***	0.124***	1.996**	0.114**	0.529***	0.121***
	(0.494)	(0.057)	(1.671)	(0.038)	(0.784)	(0.045)	(0.184)	(0.042)
2017								
Kink 1	0.677**	0.036**	3.961***	0.042***	1.513***	0.04***	0.374***	0.039***
	(0.278)	(0.015)	(0.824)	(0.009)	(0.392)	(0.010)	(0.123)	(0.013)
Kink 2	-0.027	-0.005	0.153	0.005	-0.003	0.000	-0.096	-0.034
	(0.279)	(0.049)	(0.786)	(0.028)	(0.400)	(0.035)	(0.107)	(0.038)
Kink 3	-0.068	-0.005	-0.068	-0.001	0.178	0.006	0.069	0.010
	(0.448)	(0.032)	(1.365)	(0.019)	(0.629)	(0.022)	(0.177)	(0.025)
2018								
Kink 1	0.943***	0.050***	5.091***	0.054***	1.991***	0.052***	0.452***	0.048***
	(0.242)	(0.013)	(0.846)	(0.009)	(0.408)	(0.011)	(0.101)	(0.011)
Kink 2	-0.490**	-0.087**	-1.675**	-0.059**	-0.798**	-0.071**	-0.150*	-0.053*
	(0.236)	(0.042)	(0.698)	(0.025)	(0.341)	(0.030)	(0.082)	(0.029)
Kink 3	-0.254	-0.018	-1.155	-0.016	-0.342	-0.012	-0.048	-0.007
	(0.398)	(0.028)	(1.271)	(0.018)	(0.632)	(0.023)	(0.166)	(0.024)
2019								
Kink 1	0.633***	0.033***	2.181***	0.023***	0.879***	0.023***	0.367***	0.039***
	(0.193)	(0.010)	(0.693)	(0.007)	(0.300)	(0.008)	(0.073)	(0.008)
Kink 2	-0.536**	-0.095**	-1.945**	-0.069**	-0.922***	-0.082***	-0.153**	-0.054**
	(0.210)	(0.037)	(0.797)	(0.028)	(0.318)	(0.028)	(0.076)	(0.027)
Kink 3	-0.389	-0.028	-0.808	-0.011	-0.377	-0.013	-0.100	-0.014
	(0.349)	(0.025)	(1.328)	(0.019)	(0.557)	(0.020)	(0.128)	(0.018)
2020								
Kink 1	0.601***	0.032***	0.560	0.006	0.408	0.011	0.367***	0.039***
	(0.200)	(0.011)	(0.640)	(0.007)	(0.285)	(0.008)	(0.071)	(0.007)
Kink 2	-0.488**	-0.086**	-1.648*	-0.058*	-0.737*	-0.065*	-0.157*	-0.056*
	(0.241)	(0.043)	(0.890)	(0.032)	(0.401)	(0.035)	(0.084)	(0.030)
Kink 3	-0.339	-0.024	-0.412	-0.006	-0.235	-0.008	-0.081	-0.081
	(0.385)	(0.027)	(1.573)	(0.022)	(0.720)	(0.026)	(0.135)	(0.135)
2021								
Kink 1	1.543***	0.067***	7.979***	0.069***	3.076***	0.067***	0.874***	0.076***
	(0.426)	(0.019)	(1.788)	(0.016)	(0.725)	(0.016)	(0.169)	(0.015)
Kink 2	-0.232	-0.035	-0.170	-0.005	-0.225	-0.017	-0.249	-0.076
	(0.326)	(0.050)	(1.305)	(0.040)	(0.554)	(0.042)	(0.204)	(0.062)
Kink 3	-0.944**	-0.060**	-3.814***	-0.049***	-1.636***	-0.052***	-0.253*	-0.032*
	(0.391)	(0.025)	(1.281)	(0.016)	(0.582)	(0.019)	(0.133)	(0.017)

Note: *** p<0.01, ** p<0.05, * p<0.1.
Source: authors' calculations based on PAYE 2014–21 data.

Table 8 presents excess mass and elasticity estimation results for two other estimations windows—a symmetric window around the kink corresponding to the value of the kink on each side (i.e. for the first kink this is the range $[0; 6,000]$ for the years 2014–16, $[0; 6,600]$ for the years 2017–20, and $[0; 8,000]$ for 2021) and the window $[0; 15,000]$ (one smaller and one larger than the baseline estimation window)—over the whole sample period and while keeping the bin width fixed at ZMW50. Columns (1) and (2) use an estimation window of $[0; 10,000]$ and hence reproduce the baseline result.

The results are quantitatively slightly different but qualitatively similar when using the smaller symmetric and larger estimation window of $[0; 15,000]$ in columns (3) and (5). The excess mass estimate for the first kink in 2016 increases to 1.314 and decreases to 0.767 while keeping the same significance level when using the smaller symmetric and larger estimation window, respectively. The elasticity estimate increases by 0.009 points and decreases by 0.023 points in columns (4) and (6). Similarly, the results for the first kink in 2021 keep the same significance levels for both alternative estimation windows. The excess mass estimate decreases to 1.524 using the smaller window and decreases to 1.088 using the larger estimation window, while the elasticity estimate decreases to 0.066 and to 0.047, respectively.

The estimates in Table 8 indicate that the results are generally robust to the size of the estimation window. The significance of the obtained result does not change when using the smaller and larger estimation window.

Table 8: Robustness check with different estimation windows

	[0;10,000]		**Symmetric**		**[0;15,000]**	
Year	b	e	b	e	b	e
	(1)	(2)	(3)	(4)	(5)	(6)
2014						
Kink 1	0.723***	0.042***	0.842**	0.049**	0.232*	0.013
	(0.211)	(0.012)	(0.329)	(0.019)	(0.140)	(0.008)
Kink 2	0.568**	0.108**	0.379	0.072	1.818***	0.347***
	(0.269)	(0.051)	(0.317)	(0.061)	(0.429)	(0.082)
Kink 3	0.252	0.029	-0.326*	-0.037*	-0.184	-0.021
	(0.341)	(0.039)	(0.195)	(0.022)	(0.204)	(0.023)
2015						
Kink 1	0.621***	0.036***	0.711**	0.041**	0.208	0.012
	(0.180)	(0.010)	(0.322)	(0.019)	(0.132)	(0.008)
Kink 2	0.575**	0.110**	0.433	0.083	1.574***	0.300***
	(0.282)	(0.054)	(0.321)	(0.061)	(0.401)	(0.077)
Kink 3	0.246	0.028	-0.300	0.034	-0.151	-0.017
	(0.380)	(0.044)	(0.200)	(0.023)	(0.204)	(0.023)
2016						
Kink 1	1.16***	0.067***	1.314***	0.076***	0.767***	0.044***
	(0.220)	(0.013)	(0.368)	(0.021)	(0.159)	(0.009)
Kink 2	0.207	0.039	0.102	0.019	0.887***	0.169***
	(0.242)	(0.046)	(0.297)	(0.057)	(0.273)	(0.052)
Kink 3	0.974*	0.111*	0.356	0.041	0.506*	0.058*
	(0.494)	(0.057)	(0.289)	(0.033)	(0.303)	(0.035)
2017						
Kink 1	0.677**	0.036**	0.667*	0.035*	0.564***	0.030***
	(0.278)	(0.015)	(0.351)	(0.019)	(0.208)	(0.011)
Kink 2	-0.027	-0.005	-0.109	-0.019	1.491***	0.264***
	(0.279)	(0.049)	(0.293)	(0.052)	(0.459)	(0.081)
Kink 3	-0.068	-0.005	-0.740***	-0.050***	-0.776***	-0.055***
	(0.448)	(0.032)	(0.197)	(0.014)	(0.182)	(0.013)
2018						
Kink 1	0.943***	0.050***	1.134***	0.060***	0.766***	0.040***
	(0.242)	(0.013)	(0.380)	(0.020)	(0.170)	(0.009)
Kink 2	-0.490**	-0.087**	-0.555**	-0.098**	0.065	0.011
	(0.236)	(0.042)	(0.239)	(0.042)	(0.246)	(0.044)
Kink 3	-0.254	-0.018	-0.533**	-0.038**	-0.536***	-0.038***
	(0.398)	(0.028)	(0.235)	(0.017)	(0.200)	(0.014)
2019						
Kink 1	0.633***	0.033***	0.961***	0.051***	0.474***	0.025***
	(0.193)	(0.010)	(0.307)	(0.016)	(0.129)	(0.007)
Kink 2	-0.536**	-0.095**	-0.571**	-0.101**	-0.627***	-0.111***
	(0.210)	(0.037)	(0.235)	(0.042)	(0.124)	(0.022)
Kink 3	-0.389	-0.028	-0.073	-0.005	-0.016	-0.001
	(0.349)	(0.025)	(0.279)	(0.020)	(0.301)	(0.021)
2020						
Kink 1	0.601***	0.032***	0.988***	0.052***	0.438***	0.023***
	(0.200)	(0.011)	(0.346)	(0.018)	(0.142)	(0.007)
Kink 2	-0.488**	-0.086**	-0.532**	-0.094**	-0.755***	-0.133***
	(0.241)	(0.043)	(0.264)	(0.047)	(0.128)	(0.023)
Kink 3	-0.339	-0.024	0.265	0.019	0.363	0.026
	(0.385)	(0.027)	(0.470)	(0.034)	(0.404)	(0.029)
2021						
Kink 1	1.543***	0.067***	1.524***	0.066***	1.088***	0.047***
	(0.426)	(0.019)	(0.457)	(0.020)	(0.224)	(0.01)
Kink 2	-0.232	-0.035	-0.318	-0.049	-0.342	-0.052
	(0.326)	(0.05)	(0.329)	(0.050)	(0.210)	(0.032)
Kink 3	-0.944**	-0.060**	-0.759**	-0.049**	-0.771***	-0.049***
	(0.391)	(0.025)	(0.336)	(0.022)	(0.287)	(0.018)

Note: *** p<0.01, ** p<0.05, * p<0.1.
Source: authors' calculations based on PAYE 2014–21 data.

6 Missed tax revenue

We perform a simple back-of-the-envelope calculation to estimate the missed tax revenue arising from the excess bunching at the three kinks in the Zambian PAYE schedule that we can attribute to tax avoidance. We do this by simply utilizing the counterfactual distribution to calculate the number of excess individuals at these kinks and apply the average amount of tax paid by individuals just above the tax kink. We assume this is where the excess individuals should be located in the earnings distribution, following the bunching theory laid out previously.

Table 9: Back-of-the-envelope calculation: missed tax revenue

	Bunching mass (1)	**Counterfactual mass** (2)	**Excess mass** (3)	**Avg tax missed** (4)	**Missed tax total** (5)
2014					
K1	60,573	52,694	7,879	22.39	176,424
K2	43,729	36,768	6,961	17.87	124,424
K3	30,208	27,869	2,339	30.19	70,620
2015					
K1	64,727	57,703	7,024	12.66	88,918
K2	47,216	39,623	7,593	28.19	214,067
K3	29,522	27,285	2,237	29.87	66,835
2016					
K1	76,703	58,721	17,982	17.75	319,183
K2	44,242	41,390	2,852	27.26	77,737
K3	33,984	25,654	8,330	27.95	232,802
2017					
K1	62,624	51,095	11,529	18.33	211,261
K2	40,863	41,239	-376	24.00	-
K3	25,276	25,862	-586	25.42	-
2018					
K1	84,894	64,594	20,300	14.91	302,725
K2	39,782	47,541	-7,759	24.84	-
K3	27,763	30,330	-2,567	18.75	-
2019					
K1	98,633	81,439	17,194	13.24	227,600
K2	41,998	51,138	-9,140	21.68	-
K3	29,512	33,903	-4,391	28.13	-
2020					
K1	97,265	81,034	16,231	24.38	395,624
K2	40,164	47,961	-7,797	22.80	-
K3	28,653	32,305	-3,652	25.20	-
2021					
K1	82,719	70,625	12,094	15.97	193,114
K2	36,610	39,671	-3,061	19.22	-
K3	19,080	27,836	-8,756	25.58	-
Missed tax income:					**2,701,336**

Note: total tax missed based on the group (bin width 50) just above the kink, which is calculated using the average earnings in this group multiplied by the excess mass.
Source: authors' calculations based on PAYE 2014–21 data.

Column (3) in Table 9 indicates that the number of excess individuals ranges from around 7,024 in 2015 to 20,300 in 2018 for the first kink. Since the average tax paid missed per individual is on the margin (ranging from 12.66 in 2015 to 24.38 in 2020), the total tax missed is negligible. Column (5) shows a combined loss of ZMW2.7 million due to excess bunching (excluding the kinks where there is negative bunching), or around 0.003 per cent of the overall PAYE tax collected by ZRA over the period considered. This illustrates that, although behavioural responses to changes in kinks are observed, the

missed tax revenue due to 'strategic' bunching just below the tax kinks should not be overstated, at least not in the case of Zambia.

7 Conclusion

In this paper, we delve into the intriguing realm of how individual taxpayers in Zambia react when faced with alterations in their marginal personal income tax rates. To unravel this complex behavioural phenomenon, we employ an empirical bunching methodology, scrutinizing tax administrative data spanning from 2014 to 2021. Our findings shed light on several key aspects.

We show that taxpayers exhibit a propensity to 'bunch' their income near the initial kink in the tax schedule across all the years under examination. This clustering of income suggests that individuals aim to optimize their tax liability by aligning their earnings with this critical threshold. Moreover, we observe that taxpayers quickly and sharply adapt their behaviour in response to changes in the location of the kink over time. However, when it comes to subsequent kinks in the tax schedule, we observe less solid evidence of bunching behaviour. It appears that the tax system's initial tax threshold should be the focal area of tax collector attention—also in terms of increasing the amount of individuals liable for taxes.

Despite a notable presence of bunching behaviour also when corrected for 'round-number bunching', we find that the implied elasticities of taxable income, a measure of how sensitive taxpayers are to tax rate changes, do not exhibit remarkable deviations from the norm. This implies that, while taxpayers may engage in strategic income manipulation, the magnitude of their responses to tax rate adjustments is not exceptionally pronounced.

Moreover, when we compare the actual tax revenue collected with our estimates of what it would be in a hypothetical scenario without bunching, we discover that the missed tax revenue attributable to excess bunching is relatively limited. This finding suggests that the government's revenue losses due to bunching are not as extensive as one might expect.

In summary, our study suggests that the behavioural changes we observe in taxpayers are primarily driven by reporting behaviour rather than substantial alterations in their economic activities. Taxpayers appear to be quick at adapting their income declarations to minimize their tax burden, especially near key kink points, but the overall impact on total tax revenue of this strategic tax avoidance behaviour seems moderate. It is important to keep in mind, however, that the bunching method tends to produce small elasticities due to optimization frictions, and they could be larger. Therefore, our results could be understating the tax avoidance behaviour in Zambia. These insights into taxpayer behaviour provide valuable guidance for policy makers as they consider tax policy adjustments and aim to strike a balance between revenue generation and taxpayer compliance.

References

Bachas, P., and M. Soto (2021). 'Corporate Taxation Under Weak Enforcement.' *American Economic Journal: Economic Policy*, 13(4): 36–71. https://doi.org/10.1257/pol.20180564

Bastani, S., and H. Selin (2014). 'Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule'. *Journal of Public Economics*, 109(2014): 36–49. http://dx.doi.org/10.1016/j.jpubeco.2013.09.010

Bell, N. P. (2020). 'Taxpayer Responsiveness to Taxation: Evidence from Bunching at Kink Points of the South African Income Tax Schedule'. WIDER Working Paper 2020/68. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/825-2

Bergolo, M., G. Burdin, M. De Rosa, M. Giaccobasso, and M. Leites (2021). 'Digging Into the Channels of Bunching: Evidence from the Uruguayan Income Tax'. *The Economic Journal*, 131(639): 2726–62. https://doi.org/10.1093/ej/ueab002

Boonzaaier, W., J. Harju, T. Matikka, and J. Pirttilä (2019). 'How Do Small Firms Respond to Tax Schedule Discontinuities? Evidence from South African Tax Registers'. *International Tax and Public Finance*, 26(5): 1104–36. https://doi.org/10.1007/s10797-019-09550-z

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). 'Adjustment Costs, Firm Responses, and Micro Vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records'. *The Quarterly Journal of Economics*, 126(2): 749–804. https://doi.org/10.1093/qje/qjr013

He, D., L. Peng, and X. Wang (2021). 'Understanding the Elasticity of Taxable Income: A Tale of Two Approaches'. *Journal of Public Economics*, 197: 104–375. https://doi.org/10.1016/j.jpubeco.2021.104375

Kleven, H. J. (2016). 'Bunching'. *Annual Review of Economics*, 8: 435–64. https://doi.org/10.1146/annurev-economics-080315-015234

Kleven, H. J., and M. Waseem (2013). 'Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan'. *The Quarterly Journal of Economics*, 128(2): 669–723. https://doi.org/10.1093/qje/qjt004

Mavrokonstantis, P., and A. Seibold (2022). 'Bunching and Adjustment Costs: Evidence from Cypriot Tax Reforms'. *Journal of Public Economics*, 214(2022): 104727. https://doi.org/10.1016/j.jpubeco.2022.104727

Mortenson, J. A., and A. Whitten (2020). 'Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule'. *American Economic Journal: Economic Policy*, 12(3): 402–32. https://doi.org/10.1257/pol.20180054

Saez, E. (2010). 'Do Taxpayers Bunch at Kink Points?' *American Economic Journal: Economic Policy*, 2(3): 180–212. https://doi.org/10.1257/pol.2.3.180

Saez, E., J. Slemrod, and S. H. Giertz (2012). 'The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review'. *Journal of Economic Literature*, 50(1): 3–50. http://doi.org/10.1257/jel.50.1.3

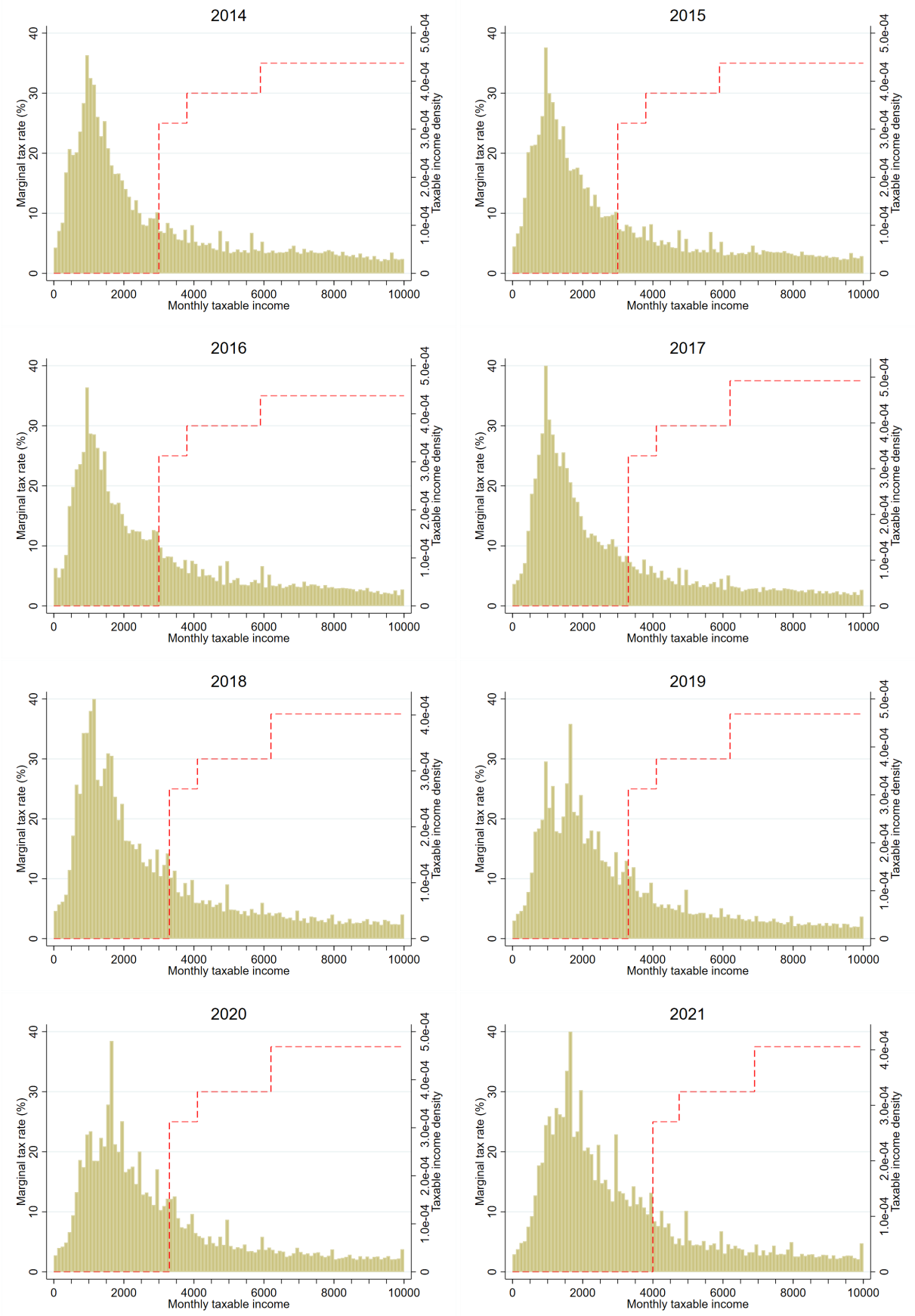
Appendix: PAYE rates

Table A1: Income tax bands and tax rates

Period	Monthly income	Annual income	Tax rate
2014–16	0–3,000	0–36,000	0%
	3,000.01–3,800	36,000.01–45,600	25%
	3,800.01–5,900	45,600.01–70,800	30%
	Above 5,900	Above 70,800	35%
2017–20	0–3,300	0–39,600	0%
	3,300.01–4,100	39,600.01–49,200	25%
	4,100.01–6,200	49,200.01–74,400	30%
	Above 6,200	Above 74,400	37,5%
2021	0–4,000	0–48,000	0%
	4,000.01–4,750	48,000.01–57,600	25%
	4,750.01–6,900	57,600.01–82,800	30%
	Above 6,900	Above 82,800	37.5%

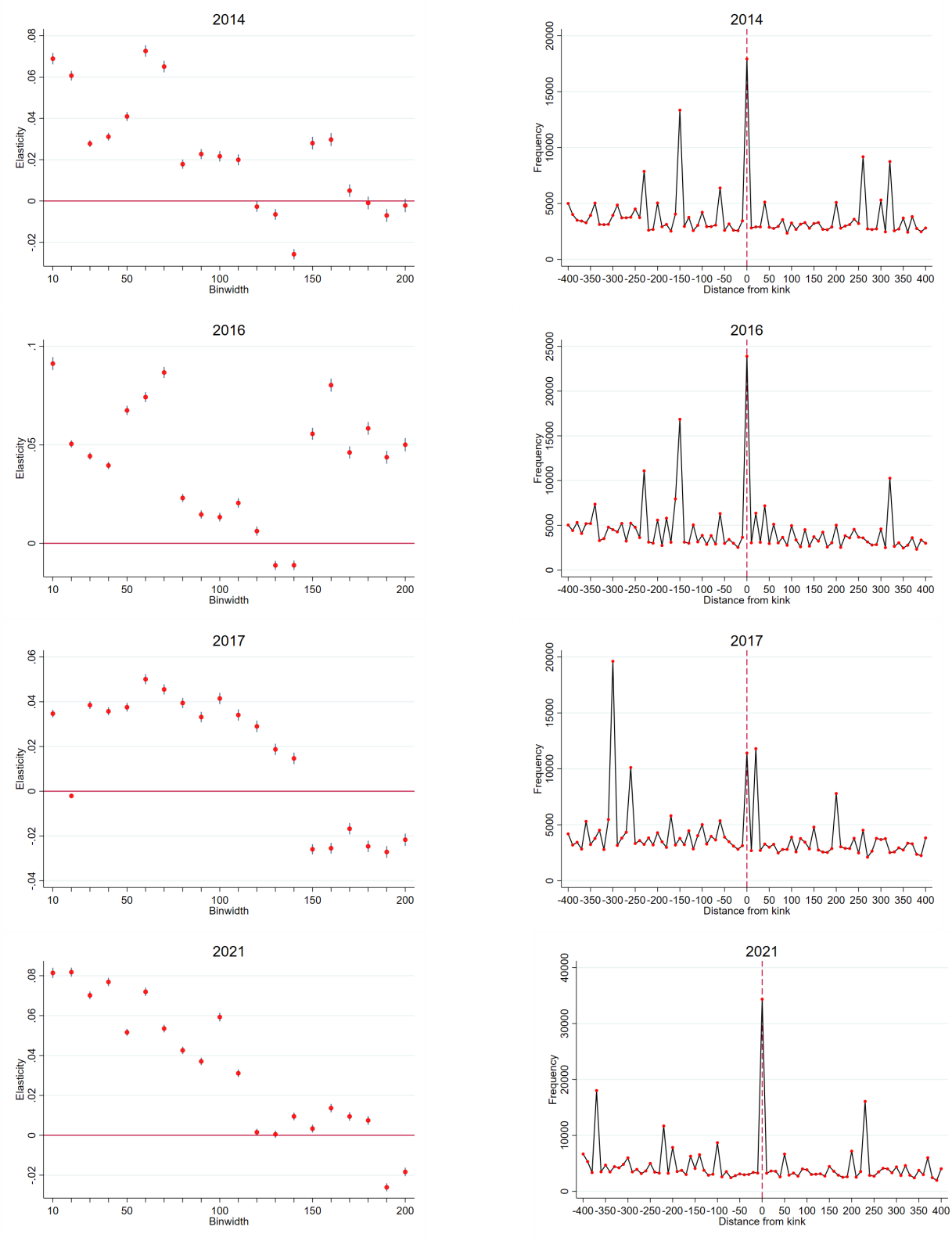
Source: authors' compilation based on Zambia Revenue Authority (ZRA) data.

Figure A1: Marginal tax rates and taxable income distribution



Source: authors' illustration based on PAYE 2014–21 data.

Figure A2: Elasticity estimation using the Saez (2010) method



Note: the plots on the left illustrate the varying elasticity estimates depending on the bin width. The plots on the right show the earnings distribution around the kink in earnings bins to help understand the fluctuations in the elasticity estimates using the Saez (2010) method. This method compares the number of individuals in a chosen bunching window with the number in a lower- and upper-surrounding band and therefore is highly dependent on the choice of bin width and bunching window.
Source: authors' calculations based on PAYE 2014–21 data.