Final report

International Growth Centre

A study on personal income tax

Anders Jensen
Enrico Di Gregorio

October 2017

When citing this paper, please use the title and the following reference number:
S-41421-ZMB-1

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid
from the British people

ZAMBIA REVENUE AUTHORITY

RESEARCH AND POLICY DEPARTMENT

A STUDY ON PERSONAL INCOME TAX (PIT)

PRELIMINARY FINDINGS

OCTOBER

2017

ZAMBIA REVENUE AUTHORITY | Working To Serve You Efficiently

Table of Contents

List of Figures

List of Tables

Executive Summary

The study aimed at investigating the tax gap in the taxation of personal income tax in Zambia. The discourse focuses on the failure of tax agents that file and pay Turnover tax (TOT), to simultaneously file and pay the liabilities under Pay as You Earn (PAYE) particularly for firms with employees.

The overall objective of the study was to establish and unveil the potential tax payers that are registered for Turnover tax but are not registered for Personal Income under the PAYE tax type. Specifically, the study sought to; establish the potential firms to target for PAYE registration; estimate the potential number of employees that would be registered; estimate the potential PAYE tax from the potential employees; and recommend evidence based PAYE tax policy proposals.

In achieving the said objectives, the study adopted a quantitative research approach that employed micro-level data from the Turnover and PAYE tax returns from October 2013 to May 2017. The study further uses STATA matching techniques to create a counterfactual population of potential tax payers by profiling the characteristics of the compliant turnover tax population and extrapolating their characteristics to the nonPAYE paying Turnover tax population. This then enables the researchers to establish the potential tax that can be collected from this group.

In terms of the **results**, we find the following:

1. In 2016, only **11.8** percent of all returns in the TOT database were compliant for PAYE. Companies were the predominant taxpayers registered for TOT and paying PAYE representing (**87.88**% of merged sample).
    a. In 2016, there are **1,958** unique complier firms, and **6,342** unique non-complier firms
2. The key characteristics of the complier group are
    a. Average number of employees: **8.3 employees.**
    b. **Average turnover of above K 7,000**
    c. Average amount of monthly PAYE: **K 1,232.50**
    d. Average ratio of monthly wages to monthly turnover: **16 percent**
    e. Total PAYE tax of **K 18,003,775** was paid in 2016 by TOT registered compliant taxpayers.
3. Under the assumption that non-complier firms are similar to complier firms (controlling for monthly turnover type of taxpayer), we estimate the potential tax in the non-compliant group as follows:

a. Total annual number of uncovered employees: **171,007 employees**
b. Total annual PAYE taxes: K **23,430,151**
c. Average annual PAYE per newly uncovered employee: K **137.01**

There are **two sets of recommendations** that emerge from the analysis:

a. In increasing compliance enforcement, there is need to target those taxpayers that file exact round-number amounts below K10,000 (that is 1000, 2000, 3000, … , 8000, 9000, 10,000)
b. Target identified non-compliant group of firms who file monthly turnover between K 90,000 and 150,000 Kwacha. This is because such taxpayers tend to under-report their turnovers while their compliant counterparts often pay PAYE for their employees.
c. Propose systematic TOT-PAYE cross-match to integrate TOT risk element into PAYE audits

Chapter One

1.0 Introduction

In many developing countries, Zambia included, personal income tax commonly known as Pay as you Earn (PAYE), collected from wages represents a large share of total personal income tax revenue. This is primarily because it is an easy to collect item due to the method employed in its collection that of withholding the tax when it is earned (Van der Heeden, 1998). Moreover, the administrative costs associated with the tax are limited, particularly because filing of returns focuses on employers, rather than the individual employees. This allows tax authorities to focus their limited resources on compliance controls of a small number of employers who take up the responsibility of withholding the tax and remitting it to the tax authority instead of dealing with a larger number of individual employees. Despite this, PAYE tax compliance is an area of concern for tax authorities worldwide in that it has never been easy to persuade all taxpayers to comply with the regulations of a PAYE tax system (Mengere, 2013). Akhand (2002) agrees that spurring tax compliance is a commonly experienced challenge for tax authorities in both developing and developed countries.

One of the strategic objectives of Zambia Revenue Authority is to improve tax compliance across all tax types that the authority administers. One such tax type that over the years has performed better in terms of revenue contribution has been income tax in the form of PAYE. Despite this being the case, it has been observed that a segment of the taxpayers has not been compliant.

This paper analyses the Pay as You Earn (PAYE) and Turnover taxes in Zambia. PAYE is an employment tax whereby tax is deducted from employees' emoluments in proportion to what they earn. Under this system, the employer calculates from the emoluments of an employee, deducts and then remits the tax deducted to ZRA. Emoluments are defined to mean total earnings from employment. These include wages, salaries, overtime, leave pay, commissions, fees, bonuses, gratuities and any other payments from employment or office as defined in section 2 of the Income Tax Act. The amount of tax deducted depends on the employee's total gross pay, the applicable tax rates; and statutory deductions. To establish the potential number of firms and employees to be incorporated into this tax type, the paper shows the characteristic results of the merged firms from both PAYE and Turnover Tax returns and compares it to the non-merged firms registered for Turnover tax .The paper further goes on to estimate the potential revenue gain from the potential new employees. The paper begins by showing the descriptive statistics for PAYE and TOT retuns from October 2013 to June 2017. However, only the 2016 data is used to estimate the potential.

1.1 A Glimpse of PAYE

PAYE is one of the major contributors to total tax revenue in Zambia and it is arguably the most efficiently collected due to the fact that it is largely withheld at the source. However, evidence from the Zambia Institute for Policy Analysis and Research study on PAYE suggest that PAYE is also susceptible to tax evasion.

The table below reveals the compliance report for PAYE on filling between 2009 and 2015.

Table 1: PAYE FILLING

TAX	YEAR	EXPEXCTED RETURNS	RETURN RECEIVED WITHIN LEGAL LIMIT	LATE RETURNS	RETURNS FOR PRIOR YEAR	VOLUNTARY COMPLIANCE (%)	TOTAL COMPLIANCE (%)
PAYE	2009	9,892	2,856	964		29	39
	2010	12,017	2,560	3,267		21	48
	2011	10,899	3,798	796	4,035	35	79
	2012	16,457	3,378	632	1,776	21	35
	2013	19,518	503	7,620		-	41
	2014	165,762	52,192	22,608		31	44

Source: ZRA, Compliance report 2015.

The following table shows the compliance statistics as at end of quarter 3 in 2015.

Table 2: 2015 PAYE payment compliance statistics

	2015 Q1	2015 Q2	2015 Q3	AVERAGE
Number of timely payments	13,694	12,707	13,722	13,374
Number of late payments	5,045	4,633	5,164	4,947
Total number of payments	**18,739**	**17,340**	**18,,886**	**18,322**
Percentage on time	73.10%	73.30%	72.70%	73%
Percent late	26.90%	26.70%	27.30%	27%
Voluntary Compliance	19.80%	18.10%	19.20%	19%
Gross Compliance	27.10%	24.70%	26.50%	26%
Tax payer population	**69,148**	**70,202**	**71,268**	**70,469**

The average number of timely payments for PAYE in 2015 by quarter 3 was 13,374 which translated into 73 percent of the total PAYE payments. The average voluntary compliance rate was 19 percent while gross compliance rate stood at 26 percent.

In quarter three alone seven tax payers were audited and the assessment showed that K273,231 was to be raised as additional tax revenue with K137,515 as penalties.

In 2016, the total number of tax payers who were registered for PAYE was 27,377 as at the end of quarter 3. 27,355 of these were active while 22 were inactive and nine were de registered in September, 2016.

1.2 Statement of the Problem

The withholding of taxes at the source of income is the most basic of all compliance techniques and it increases the costs of evasion or avoidance to the taxpayer. It is widely used in developed and developing countries for collecting revenue with minimal use of administration resources. A popular mechanism is wage withholding, commonly known as pay-as-you-earn (PAYE), in which employers are withholding agents for personal income tax payable on the earnings of their employees. PAYE has therefore become the most common general income tax in the world. The widespread use of the withholding tax system and the PAYE is due in part to their perceived efficiency and effectiveness in raising tax revenue compared with other taxes. *However, like other taxes, the PAYE is vulnerable to tax evasion, tax fraud and poor enforcement thus contributes to the PAYE Gap*. ZIPAR (2014) and ZRA (2016) report highlights that there is a gap between tax reported by taxpayers and actual taxable income. Tax reported by the taxpayer is less than the tax payable under the law.

The Authority has missed its revenue targets in some years thus compelling the government to increase the level of borrowing to finance its expenditures. *The failure to meet the targets has been largely attributed to the revenue gaps arising from non-compliance in tax types such as PAYE*. Whereas several reforms have been adopted in an effort to enhance compliance such as adopting online services to make it easier for taxpayers to comply, the tax gaps still remain a major concern to the Authority. Thus focusing on the strategies that can seal the revenue gaps will go a long way in providing an insight to the policy makers on the key strategies towards enhancing revenue collection. Therefore, to arrive at the strategies, a clear understanding of the characteristics of the PAYE non-compliant firms that have the potential to be brought in this tax net is desired.

1.3 General Objective

The overall objective of the study is to establish and unveil the potential tax payers that are registered for Turnover tax and are not registered for Personal Income Tax under the PAYE tax type and specific objectives are as follows:

1.4 Specific objectives

1. To establish the potential firms to target for PAYE registration.
2. To estimate the potential number of employees that would be registered.
3. To estimate the potential PAYE tax from the potential employees.
4. Recommend evidence based PAYE tax policy proposals

1.5 Rationale

The importance of tax equality, equity and fairness has long been recognized in taxation. Perhaps among the most critical matters to equity in taxation are two factors i.e. inter-group equity and fairness; and compliance equity and fairness. It is important that no group of taxpayers appear to be favoured to the detriment of another without good cause. Actual or perceived inequities as is the case with some non-registered firms on PAYE tend to demoralize the compliant ones. It is important therefore that all taxpayers pay what they owe on a timely basis. In the long run, significant non-compliance depresses perceptions of equity, increases tax administration costs, shift tax burdens, and enlarges the tax gap. In view of the foregoing and in the spirit of fairness in taxation, there is an immediate need to develop strategies to net in these individuals.

Chapter Two

2.0 Review of related literature

Having identified that there is a tax gap in the PAYE tax, this section will utilise the literature review relating the tax gap in Personal Income tax.

A tax gap analysis is not just about examining revenue coming in, but how it is collected, who collects it and what happens to it once it is collected. All countries have a tax gap. No government ever collects all revenues due because the administration costs of attaining a100 percent compliance would outweigh the revenue gains available. It is in this regard that this section presents other studies that have tried to analyse the Tax Gap. First the theoretical part of the paper presents definitions and methods as well as the factors fostering and deepening the tax gap and practical measurement methods of the tax gap.

Tax has been defined by various authorities and professionals in various ways. Conceptually, tax can be defined or seen as a compulsory transfer of resources from the private to the public sector (Uremadu, 2000). According to Lymer and Oats (2009) tax is defined as a compulsory levy, imposed by government or other tax raising body, on income, expenditure, or capital assets for which the taxpayer receives nothing specific in return.

For decades, policy makers and politicians have railed against the tax gap. However, the tax Gap defined as the difference between what taxpayers are legally obligated to pay in taxes and what they actually pay in taxes has begun to gain recognition as a powerful policy tool. In this regard, Adams (1921) defines the tax gap as the difference between the tax which would be raised under a hypothetical, perfect enforcement of tax laws and the actual tax payments. In other words, the tax gap is the difference between the tax revenue that the government should collect from the taxes owed in the country if everyone complied with the law and the amount of tax revenue the government does actually collect.

Although the determination of the tax gap is not easy, a rough estimation is usually done. The calculated gap can be further broken down into areas such as barely legitimate tax avoidance, fraud, serious non-compliance, error, and debt. This provides additional information on areas of risks that should be addressed.

Researchers have often used the term "non-compliance" to characterise the intentional or unintentional failure of taxpayers to pay their taxes correctly, and the term will be used in this way in this study. Unintentional non-compliance is the failure of a taxpayer, or of an intermediary acting on behalf of the taxpayer (in this case the employer), to remit the proper amount of tax to the authorities, perhaps on account of the complexity or contradictions in the tax legislation or tax administration procedures (Kesselman, 1994:62-84). It may arise from inadequate effort by the taxpayer or intermediary to ascertain its obligations. It may also stem from the complexity of tax provisions and the difficulty of applying them to the more complex situations of the real world. On the other hand, intentional non-compliance can be divided into two types of activities: Tax evasion and Tax avoidance.

Tax evasion occurs when taxpayers either under-report income or exaggerate deductible expenses when filing their returns. In other words, evasion involves deliberately misreporting the nature of ones' income or undertaking other activities inconsistent with the letter of the law to reduce tax payable. On the other hand, tax avoidance refers to avoiding the payment of taxes by using loopholes in the tax legislation or the arrangement of a person's affairs, correctly reported, so as to reduce tax liabilities within the letter of the law but to a level lower than the limit of the law (Wameryd and Walemd, 1982:187). Therefore, once the manner in which taxes are being avoided has been detected, the problem may be overcome by refining the legal provisions and their interpretation by amending the legislation (Silvani, 1994:274). This kind of underpayment of taxes is sometimes engaged in by large taxpayers who have trained professionals to advise them on the interpretation of legal provisions.

2.1 Brief Summary of some Empirical Studies

One of the most influential papers on the topic of estimating PAYE Gap was conducted by Pissarides and Weber in which they estimated the size of Britain's black economy (defined narrowly as unreported taxable income) by using income and expenditure data drawn from the 1982 Family Expenditure Survey. Pissarides and Weber compared the relationship between food expenditure and income in two groups of workers, self-employed and employees in employment, assuming that employees reported their incomes correctly. For a given level of reported income, the self-employed had higher food expenditure than employees. Pissarides and Weber concluded that the self-employed actual income is 1.55 times reported income, and that this part of the unobserved economy was about 5.5 % of GDP in the United Kingdom in 1982 (Pissarides and Weber, 1989).

In a study by Martinez-Lopez in which he used the Spanish household surveys over the period 2006-2009, and in which he replicated the approach by Pissarides and Weber but extended its interpretation by including the concealment of income by salary workers, it was found that the reported income by the self-employed has to be increased by about 25 % to obtain the level of income which would equal the level of underreporting by employees (Martinez-Lopez, 2012).

In a study conducted by ZIPAR titled Uncovering the Unknown: An Analysis of Tax Evasion in Zambia, it was discovered that applying the tax thresholds to the income of both the self-employed and the paid employees, they estimated the average tax liability for each household. This amount was then multiplied by the average number of people with taxable income in each of the two categories, annualised and compared to the reported PAYE for 2010. The study revealed that the PAYE gap amounted to K5.2 billion, which was 6.7 % of GDP and 40.3% of the total tax revenue**.** The study showed that by just concentrating on the wage earners who were above the tax threshold, the Zambia Revenue Authority would have collected an additional K800 million from those classified as wage earners even before taxing the 10 % of the self-employed. Notwithstanding the difficulty and administrative burden that would result from collecting this money, as pointed out by Phiri and Kabaso (2012), the study showed that about 10 % of the many people who were in self-employment were above the tax threshold for paying PAYE tax in 2010 and would potentially contribute as much taxes as those in wage employment.

2.2 Brief Summary of Selected Countries Which Estimate Tax Gaps

A number of countries undertake tax gap estimates. Prominent examples of countries estimating various types of tax gaps include France, Sweden, the United States of America (USA), and the United Kingdom (UK). Several individual states of the USA also estimate tax gaps, such as Minnesota, Idaho, New York and California. Other countries with lists of publicly available tax gap estimates include New Zealand, the Philippines and Brazil. Sweden has estimated tax gap on a broad range of taxes and social security levies including VAT for the years 1997 and 2000. Sweden’s tax gap estimates are calculated to provide guidance on the magnitude of the gap rather than precise year-to-year trends in tax gap. Tax gap estimates were also routinely prepared by the US Internal Revenue Service (IRS) throughout the 1980s and 1990s. Although the IRS continues to examine available data and use it in its compliance management, data collection supporting these estimates has been suspended.

In addition to the various countries choosing to estimate tax gap for internal management purposes, the International Monetary Fund (IMF) and the World Bank also use tax gap estimates as a performance measure. The IMF requires a tax gap estimate in relation to countries or states that it supports, as a condition of providing assistance. The World Bank also assesses government performance using tax gap estimates amongst other measures.

2.3 Benefits and Importance of Tax Gap Analysis

It can arguably be stated that the increasing number of countries and states investing in the development of tax gap models indicates there is an important value in doing so. The literature on tax administration and tax compliance suggests there are various benefits that flow from estimating tax gap. The following is the summary of the associated benefits of tax gap analysis:

- ✓ Identification of the types and level of non-compliance that contribute to the tax gap;
- ✓ Improved efficiency of resource allocation within a revenue authority to combat non-compliance;
- ✓ A measure of effectiveness of a revenue authority. Tax gap estimates are able to help revenue authorities assess their overall performance by monitoring changes in the estimated gap.
- ✓ Identifying any loopholes that exist in current legislation
- ✓ Highlighting areas for administrative reform (in the collection and accounting of revenues)
- ✓ forcing the authority to focus attention on the need to understand how non-compliance occurs and how the causes can be addressed-whether through tailored assistance, simpler legislation, redesigned processes or targeted interventions.

- ✓ Offering policy choices to the government (by expanding the tax base, introducing new taxes, or altering the rate or coverage of particular taxes)
- ✓ Tax gap analysis provides a long term health check to validate the strategic decisions taken by the Authority and their effectiveness in tackling major risks.

In summary, the generally acknowledged benefits or importance of undertaking tax gap analysis are that the Authority and government will be better informed about: tax system integrity; risks to revenue buoyancy; performance of their tax collection agency and processes; evolving risks to revenue (and potential failures by the tax collection agency); problems with the tax legislation; problems with the national statistics; and the impact of the non-observed economy on revenue. Therefore, these assurances are increasingly important to governments worldwide. For government, increasing demands for the provision of services (such as health and welfare) means it is imperative that taxes due are paid. For the general public, any evidence of tax non-compliance has a direct impact on the equity and economic efficiency of taxes and this can lead to a loss of public confidence in the integrity of the tax and the revenue authority.

CHAPTER THREE

3.0 Research Methodology

3.1 Research Design

This study adopted a quantitative research approach (desk analysis).Descriptive analysis was utilised to describe the data and characteristics making up the PAYE and TOT returns. Using micro-level data from the two returns from October 2013 to May 2017 the study uses matching techniques using the STATA command to establish the potential tax that can be collected from the potential PAYE tax payers established from the TOT return.

Before merging of the two data sets, the monthly data sets are appended in STATA to come up with yearly data set. Note that variables with string characters and duplicates are dropped out from the sample to come up with a clean data set and a unique TPIN for each data set. A more detailed approach of the methods of analysis is explained as the results are presented in Chapters 4 and 5.

3.2 Sources of Data

The data used in this study consists of only secondary data obtained from the ZRA TaxOnline system on PAYE and TOT returns,

Chapter Four

4.0 Descriptive Characteristics of PAYE and TOT database

In this section, the results from a descriptive exercise that characterises the PAYE database and TOT database are presented. These descriptive results may be of interest per se, and trigger further analysis into specific topics. But of equal importance, these descriptive results help to introduce the methodology in the analysis of Chapter Four. These general, descriptive results also provide a sense of context for the magnitudes of the potential employee and PAYE that will be uncovered amongst potential non-compliers in Chapter Five.

4.1 Descriptive Characteristics of the PAYE database

The exercise begins with a descriptive analysis of the PAYE database. The full PAYE dataset contains monthly PAYE returns from October 2013 to May 2017. The dataset is structured such that each row corresponds to a unique PAYE return filed by a taxpayer in a specific month of a specific year.

4.1.2 Distribution of basic salary at the level of the monthly PAYE return

The first exercise conducted was to describe the distribution of basic salary at the level of the monthly PAYE return. For this exercise, we pooled all monthly individual returns that were contained inside either of the three calendar years: 2014, 2015, and 2016. The description of the distribution of basic salary can in principle also be done at other time-levels, such as quarters of a given year, or across months of a given year. Here, we chose to focus on the yearly distributions. In order to characterise the distribution, we calculate the percentiles of basic salary in a given calendar year. Creating the percentiles of a distribution arranges the individual observations into groups (in this case, 100 groups), and each group contains the same number of observations. Observations are classified into groups based on their value. As an example, the percentiles of the basic salary distribution in 2014 will group all the monthly returns into groups with an equal number of returns in each, and which are arranged by the value of basic salary indicated on the monthly return. So the 1^{st} percentile of the 2014 distribution will contain the 1 percent of lowest value of monthly returns. The 99^{th} percentile will contain 1 percent of the full sample of monthly returns which correspond to the highest values of basic salary in the full sample of the monthly returns. Within each percentile group, we then calculated the average of the basic salary. Here, several other options are possible: the minimum, the maximum, the median. Since in practice our percentiles are based on a large number of observations, these alternative computations will reveal very similar results.

Table 3: Summary of basic salary distribution and unique number of employees by year

2014		2015		2016	
Percentiles	**Basic Salary**	**Percentiles**	**Basic Salary**	**Percentiles**	**Basic Salary**
1%	-	1%	-	1%	-
5%	33	5%	-	5%	-
10%	457	10%	487	10%	513
25%	1,013	25%	1,048	25%	1,088
50%	2,152	50%	2,323	50%	2,489
Mean	**7,539**	**Mean**	**8,086**	**Mean**	**8,576**
75%	6,422	75%	6,914	75%	7,422
90%	13,750	90%	14,747	90%	16,660
95%	24,745	95%	26,410	95%	29,813
99%	86,845	99%	91,834	99%	99,547
Obs	**4,143,567**	**Obs**	**4,343,786**	**Obs**	**4,206,616**

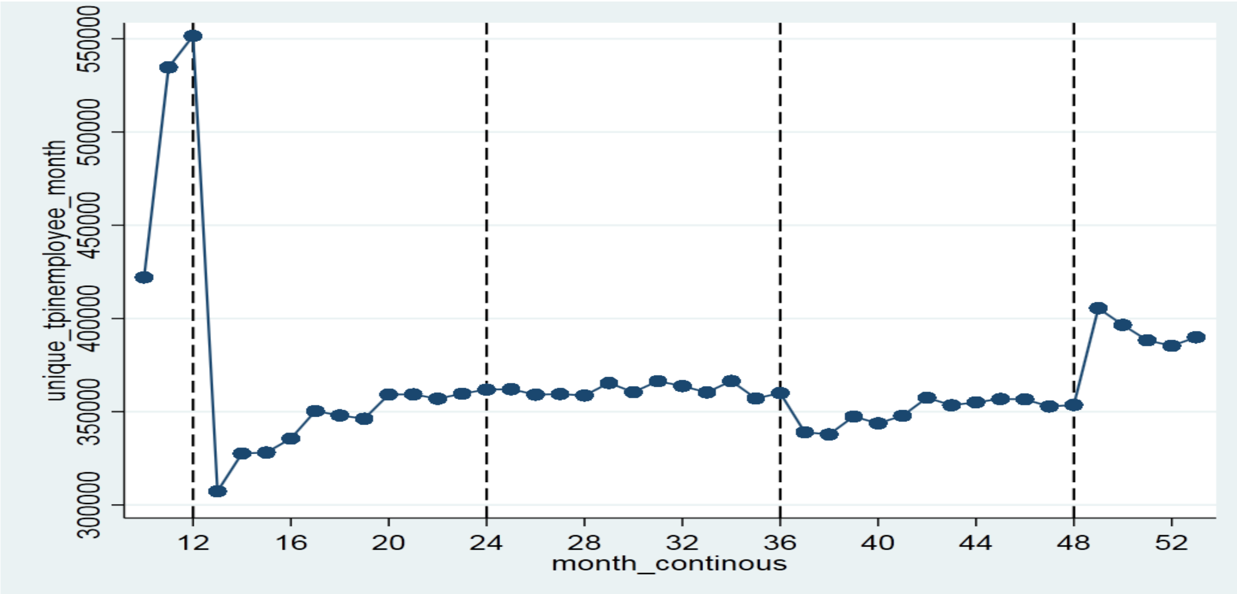
Table 3 displays the results from computing the percentile (average) values, separately for 2014, 2015, and 2016. To interpret the table for e.g. 2014, let us focus on the value of the 10th percentile – 457.12. This number tells us that the average value of monthly returns in the 10th percentile in 2014 is K 457.12. A complementary interpretation of the 10th percentile is to say that 10 percent of monthly returns in 2014 had a reported amount of basic salary that was below K 457.12. Note, as the bottom percentile indicates, our calculation of percentiles does not deal with cases where the reported basic salary on the monthly return is 0. Reporting 0 as basic salary could arise for several reasons that we do not address in the main text. In order to remedy the issue of nil returns, one could construct the percentiles of the *average* monthly individual basic salary distribution.

The usefulness of constructing percentile values is that this allows a comparison of how basic salary has evolved over time, *for a specific part of the distribution*. Per example, by comparison across years, we can see that the value of the 10th percentile grew from K 457.12 in 2014 to K 486.81 in 2015 to K 513 in 2016. Between 2014 and 2016, this represents a nominal growth in monthly basic salary of 12.2 percent for the 10th percentile. In comparison, the 90th percentile grew from K 13,750 in 2014 to K 14,747.15 in 2015 to K 16,660.48 in 2016. This amounts to a nominal growth of 21.16 percent between 2014 and 2016 for the 90th percentile. The nominal growth, in percentage terms, for the 90th percentile of basic salary was therefore almost double as large as the nominal growth for the 10th percentile. The differential nominal growth patterns across different percentiles of the salary distribution have important impacts on the degree of salary inequality over time. On the other hand, it is important to compare nominal growth trends for *all* parts of the income distribution – per example, the 2014-2016 nominal growth of the 99th percentile (14.62 percent) was much lower than the nominal growth of the 90th percentile and much more in line with the nominal growth of the 10th percentile.

4.1.3 Unique number of employees TPINs in the monthly returns

In the next exercise, we calculate the unique number of employees TPINs that we observe in the monthly returns. Note, the employee TPIN technically refers to the NRC number for most cases. We choose to calculate the unique number of employee TPINs in every month of every year. Note that in order to calculate the unique number of employee TPINs at a more aggregated level, such as the yearly level, it will not be correct to simply add up the unique TPINs across all months of the year – indeed, in this case, the duplicate TPINs across months would not be accounted for, and the number of unique employee TPINs would be artificially too large. The statistical command in Stata required to construct the unique count of number of employee TPINs is straightforward. The results are displayed in figure 1 below.

Figure 1 Unique number of TPINS by month



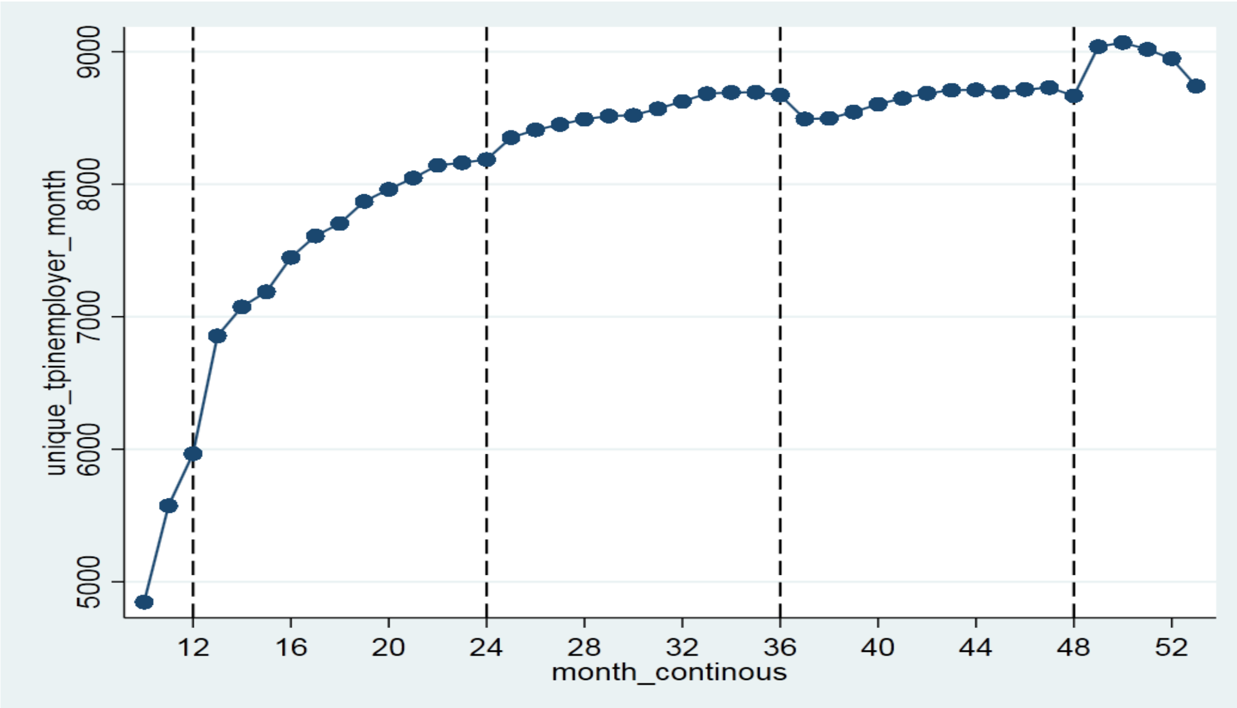
Note: 12 = December 2013, 24 = December 2014, 36=December 2015, 48= December 2016, and 52 =April 2017

Each dot corresponds to the total unique number of employee TPINs in a given month of the year. Several things stand out from this graph. First, the number of unique employee TPINs is much larger in the months of 2013 than in the months of any other year. Second, the year 2014 saw a strong steady increase on a monthly basis in the number of unique employee TPINs. This growth came to a halt in 2015. Third, there are important, large discrete changes in the number of unique employee TPINs around the change of year – both between 2015 and 2016, and between 2016 and 2017. Such immediate and large jumps could be explained by 'real' factors, such as changes in tax policy and the tax code; sudden but large changes in the underlying economic structure; and, firm behaviour (including large lay-offs before the turn of calendar year). On the other hand, further investigation should clarify if (at least parts of) these large jumps are not also caused by transitions across IT software, or yearly transitions in the way that employee TPINs are reported on monthly returns.

4.1.4 Unique employer TPINs across months and years

To accompany the time-series of the unique number of employee TPINs, we also compute the number of unique employer TPINs across months and years. The results of this exercise are displayed in figure 2

Figure 2 Unique number of employers by month



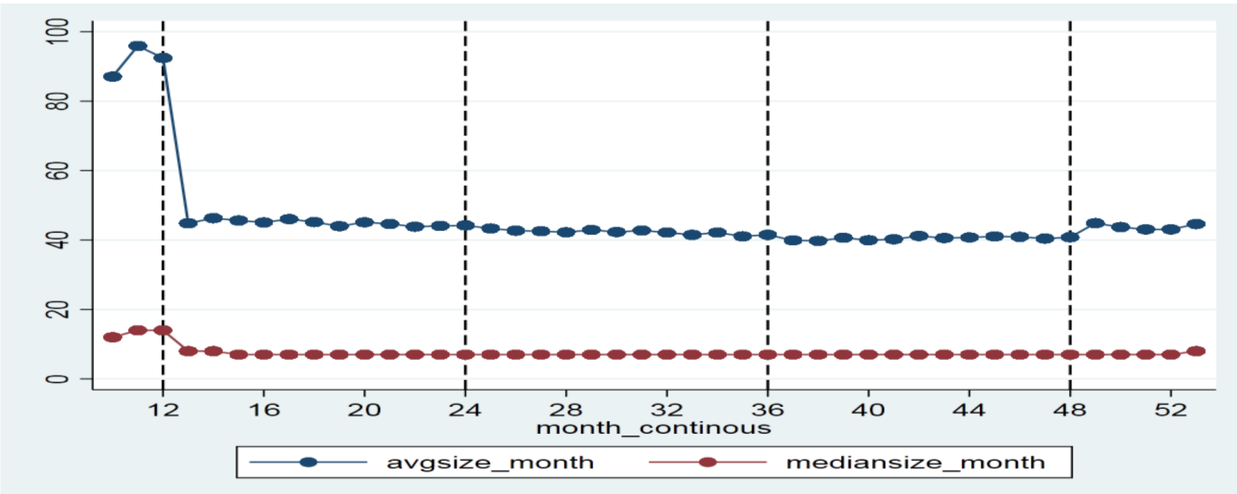
Note: 12 = December 2013, 24 = December 2014, 36=December 2015, 48= December 2016, and 52 =April 2017

Several items are worth noting. First, while the unique number of employee TPINs saw a large drop between end of 2013 and beginning of 2014, the unique number of employer TPINs saw an equally large increase in the same period. Second, the strong growth in unique employee TPINs during 2014 is mirrored by an equally strong growth in number of unique employer TPINS over the same period. In contrast, 2015 witnessed a continued growth in the number of unique employer TPINs while the unique number of employee TPINs came to a halt. Finally, the end-of-year discontinuous jumps in employee TPINs (down between 2015 and 2016; up between 2016 and 2017) are also reflected in jumps of the number of unique employer TPINs.

4.1.5 Average firm size

To complete this section, we construct a measure of the average firm size in the PAYE monthly return dataset. We construct this size-measure by counting the number of unique employee TPINs by unique employer TPIN in every month of every year. We then take the average over this ratio for all unique employer TPINs in every month of every year, and define this as the monthly average firm-size. We similarly calculate the median monthly firm-size. We calculate and report the median as well as the average, since the average value may be driven by a few extreme outliers and thus mask the reality of the situation for the vast majority of the number of firms in Zambia. The average monthly firm-size thus corresponds to the monthly average number of unique employee TPINs reported by unique employer TPIN. Results are shown in figure 3.

Figure 3 Average firm size by month



Note: 12 = December 2013, 24 = December 2014, 36=December 2015, 48= December 2016, and 52 =April 2017

First, the large drop in average (and median) size is noticeable between end of 2013 and beginning of 2014. Second, there is a very gradual decrease in average size between 2014 (average size of 43) and 2016 (average size of 40). On the other hand, the median size stays *constant* throughout this entire period, with a value of 9. Finally, there is a large increase in the average firm size immediately upon entering 2017, which could be driven in part by tax policy changes.

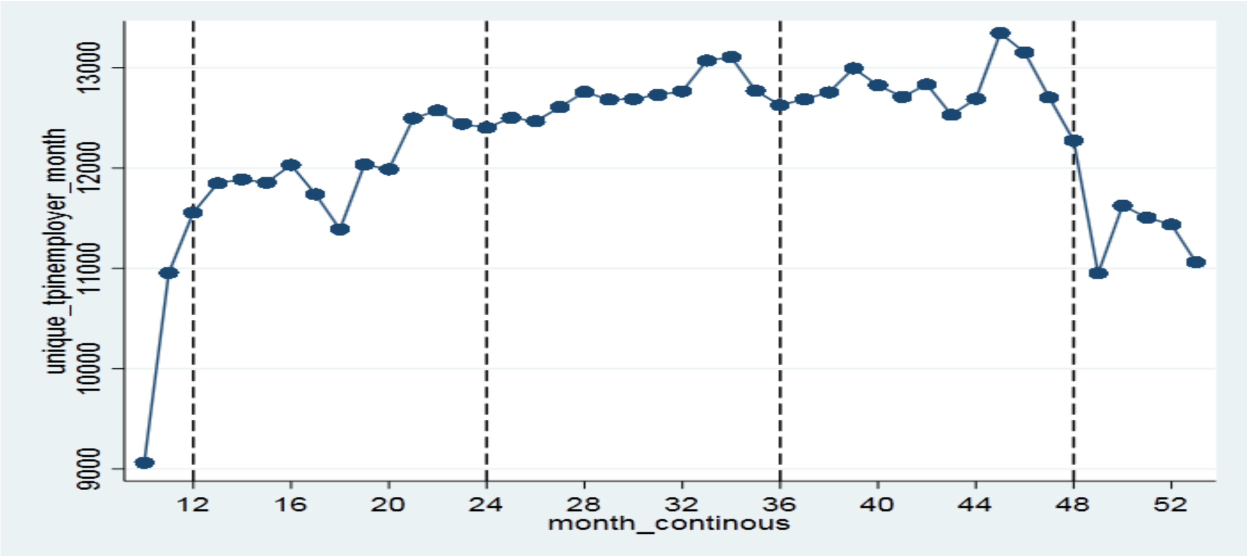
4.2 Descriptive Characteristics of the TOT database

In this subsection, we report the results from the descriptive exercise to characterise the turnover tax database.

4.2.1 Unique number of TPINs across months and years

The first exercise we undertake is to count the unique number of TPINs in every month of every year. The data is organized at the level of firm-level monthly returns. Note that, in our calculations, an increase over a given month in the unique number of TPINs can be driven by two factors. First, it can be driven by new TPIN registrations. Second, it can be driven by turnover tax firms that do not file on a regular basis – e.g. if a turnover firm has filed in the past, but not in the previous month, and then files again in the current month. The results are displayed in figure 4.

Figure 4 Number of unique taxpayers by year and month



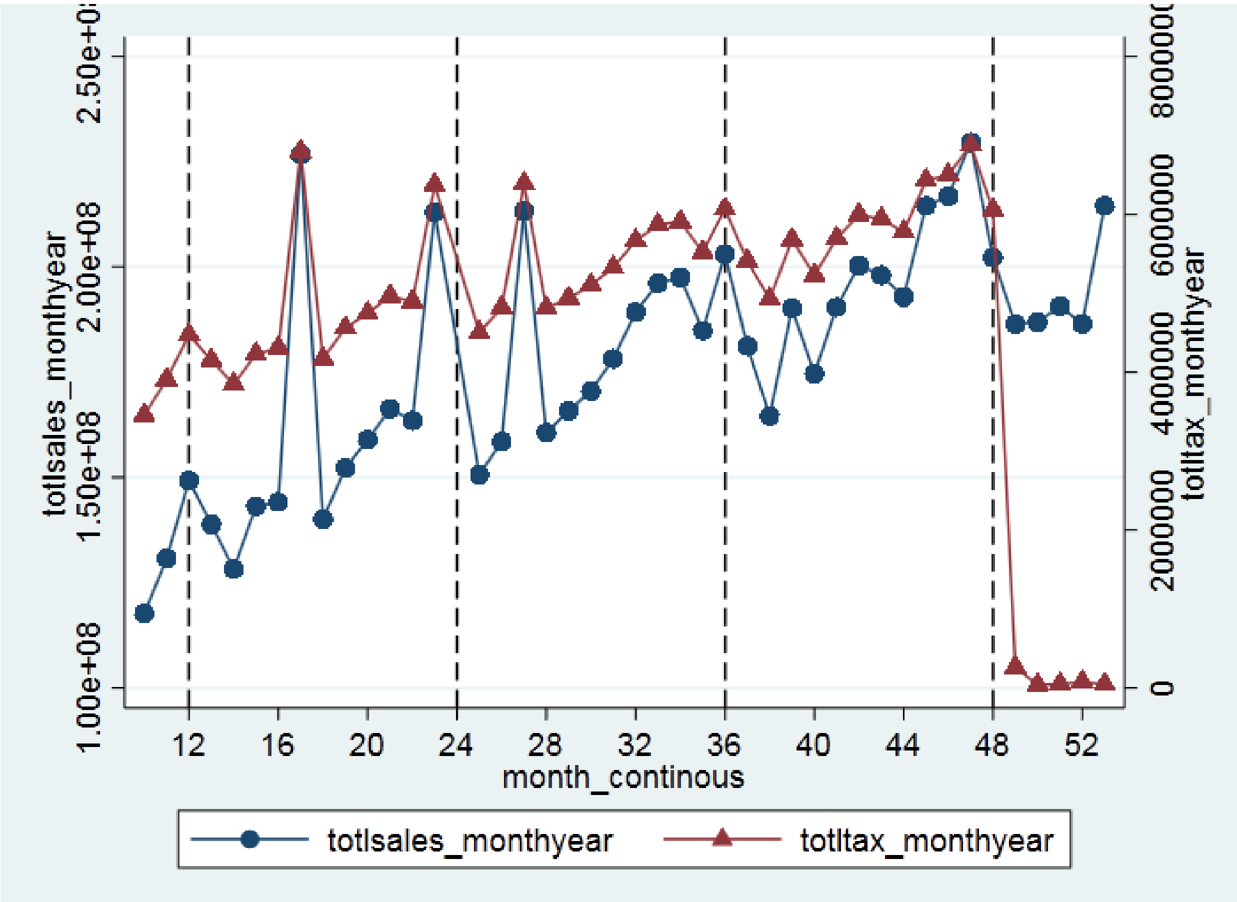
Note: 12 = December 2013, 24 = December 2014, 36=December 2015, 48= December 2016, and 52 =April 2017

There is an overall trend over most of the period, indicating a steady growth on a monthly basis in the number of unique TPINs that file a turnover tax return. However, towards the last quarter of 2016, the number of unique TPINs starts to significantly drop, and this trend continues into 2017. At the time when the data series stopped (May 2017), the number of TOT TPINs was very close to its lowest level since October 2013.

4.2.2 Gross sales and net taxes paid

To complement these time-series, we construct on a monthly basis the aggregate total amount of gross sales reported and the aggregate total amount of net tax paid. We construct this measure by summing the relevant line-items over all TOT returns in a given month. These time-series are displayed in figure 5.

Figure 5 Aggregates by year and month: gross sales, net tax



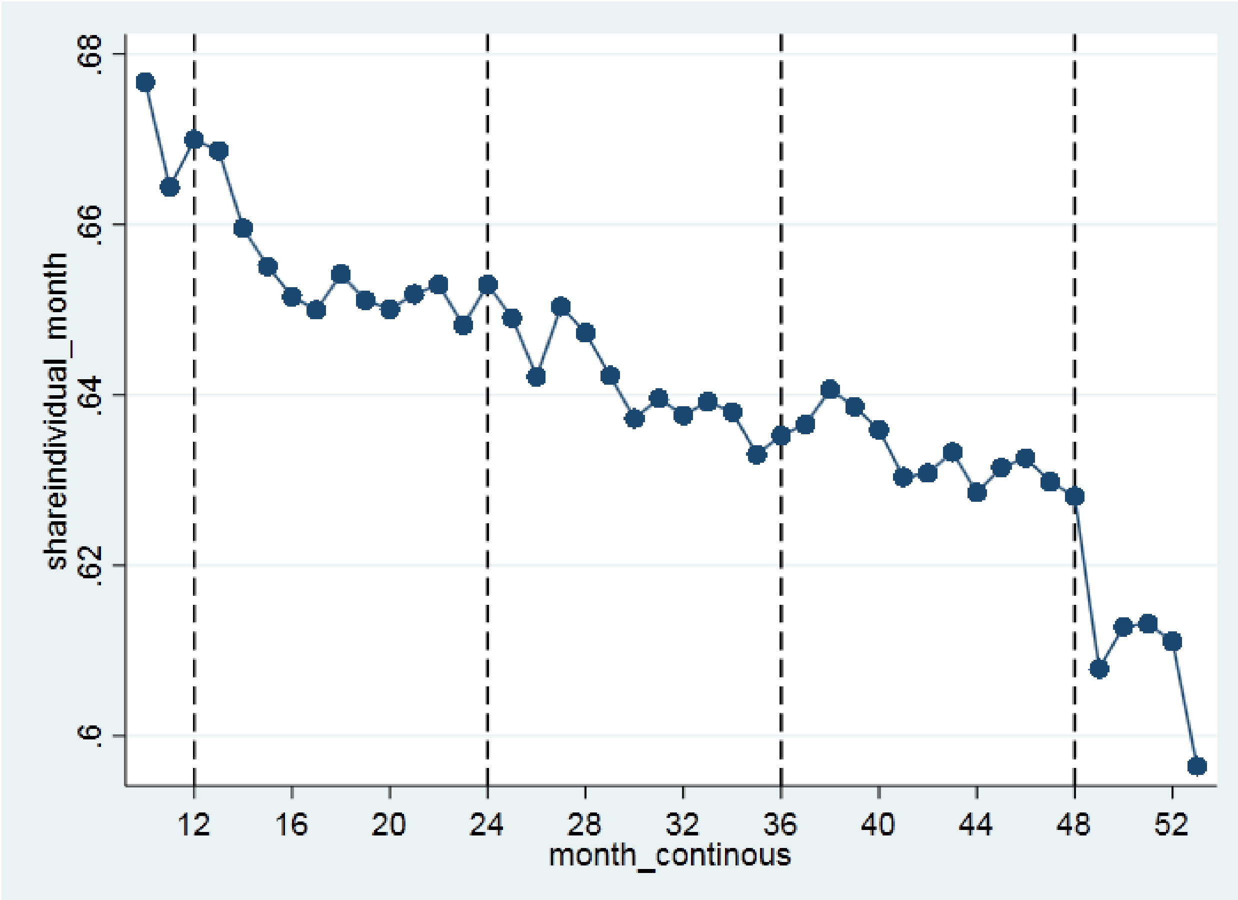
Note: 12 = December 2013, 24 = December 2014, 36=December 2015, 48= December 2016, and 52 =April 2017

The long-run trend is clear: total gross sales and total net tax steadily increase on a monthly basis from end of 2013 until end of 2016. In addition, the short-run fluctuations are equally clear: the beginning of each year marks a downturn, both in sales and in net tax deducted. Both the gross sales series and the net tax series are marked by monthly spikes (June and December 2014, March 2015), which appear worthy of further investigation.

4.2.3 Monthly returns filed by Individual taxpayers

As an additional time-series exercise, we compute the share of monthly returns that are filed by agents that declare their taxpayer type as 'Individual'. This category makes up 64 percent on average in the full sample, while 'Companies' make up almost the entire remaining 36 percent.

Figure 6 Share of taxpayer types (share of individuals) by year and month



Note: 12 = December 2013, 24 = December 2014, 36=December 2015, 48= December 2016, and 52 =April 20

Figure 6 displays a robust long-run trend: the share of returns that are filed under ‘Individual’ taxpayer type are on the decline. Note that this decline could be driven by an increase in the number of returns filed as ‘Company’; a decrease in the number of returns filed as ‘Individual’; and, a combination of these two factors.

4.2.4 Distribution of gross sales at the level of the monthly TOT return

We then turn the focus to characterizing the distribution of the monthly amount of reported gross sales. For each year, we construct the percentile values of the monthly gross sales distribution in a very similar fashion to the construction of the basic salary distribution. The results are displayed in Table 2.

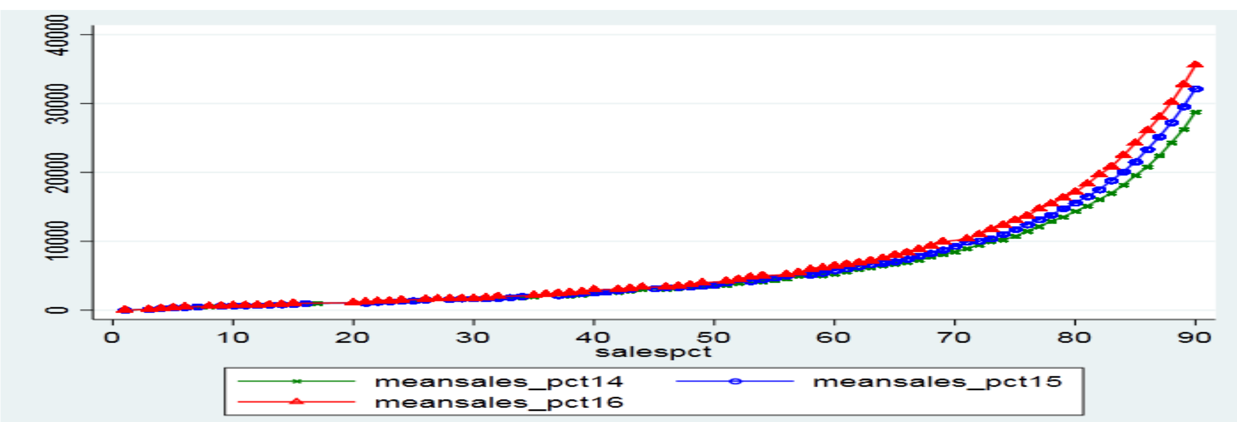
Table 4: Percentiles of gross sales distribution

2014		2015		2016	
Percentiles	**Gross Sales**	**Percentiles**	**Gross Sales**	**Percentiles**	**Gross Sales**
1%	-	1%	-	1%	-
5%	300	5%	350	5%	367
10%	600	10%	667	10%	667
25%	1,350	25%	1,400	25%	1,500
50%	3,500	50%	3,800	50%	4,000
75%	11,027	75%	12,000	75%	13,333
90%	30,000	90%	33,350	90%	37,187
95%	50,000	95%	55,240	95%	60,625
99%	120,000	99%	130,146	99%	149,926
Mean	**38,063**	**Mean**	**14,195**	**Mean**	**15,345**
Obs	**144,691**	**Obs**	**152,778**	**Obs**	**153,506**

As an example of interpretation, the value of the 50th percentile in 2016 indicates that 50 percent of TOT returns reported monthly gross sales that were below K 4,000. The comparison of 2014-2016 nominal growth rates in monthly gross sales across percentiles of the distribution reveals that: the 10th percentile grew by 11.1 percent; the 50th percentile grew by 14.28 percent; the 90th percentile grew by 23.95 percent; and, the top 1 percent grew by 24.9 percent. In contrast to the basic salary distribution, in the case of monthly gross sales, the largest nominal growth rates do appear to systematically be concentrated towards the top.

To complement the table, figure 7 displays the percentile values side by side for the three years 2014, 2015, and 2016.

Figure 7 Average values of percentiles of gross sales



The graph makes clear that in absolute Kwacha amounts, the pronounced growth is occurring towards the top of the percentile distribution.

4.2.5 What values do most firms choose to declare their monthly gross sales

In the final part of the TOT descriptive exercise, we aim to answer the question: at what values do most firms choose to declare their monthly gross sales? In order to investigate this question, we construct the frequency distribution of monthly gross sales. The distribution is constructed as follows. First, we define a set of groups of monthly gross sales. For our purposes, we chose to create groups in increments of K100: the first group goes from K1 to K100, the second group goes from K100.01 to K200, and so on. Second, we count the number of monthly returns that fall within each of the groups – this is called the frequency count. Note, for our purposes we chose to count the number of monthly returns filed *over the full calendar year*, but a similar exercise could meaningfully be repeated on a month by month basis. Third, we plot the frequency count against the groups of monthly gross sales.

Figure 8 Distribution of gross sales by year 2014

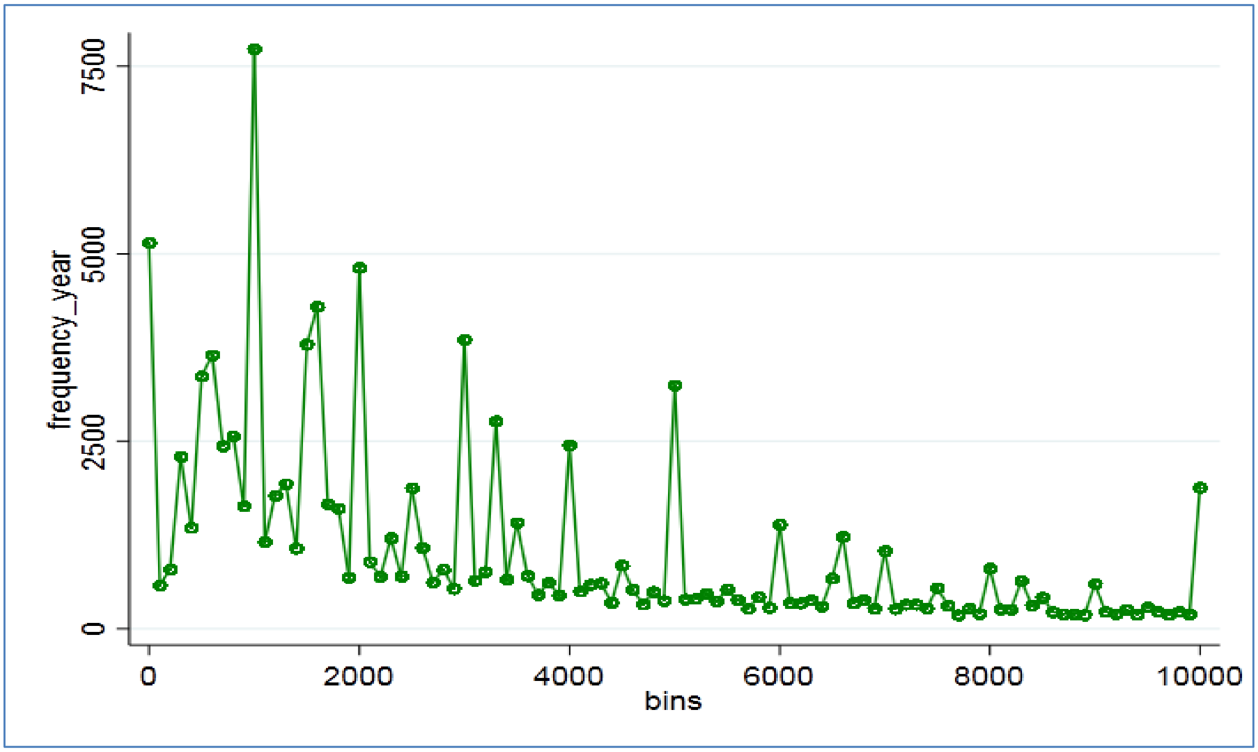


Figure 8 displays the results for the year 2014. At least two features stand out. First, there is a clear trend when considering the full X-axis: the frequency count gradually decreases as the value of monthly gross sales increases. In other words, there are many more firms filing returns with low monthly gross sales than with high monthly gross sales. Second, there are important irregularities at specific points on the X-axis. Upon closer inspection, it becomes apparent that most (but not all) of these spikes in the frequency count occur *exactly* at round-number values of monthly turnover: 1000, 2000, 3000, 4000, 5000, 6000, etc. The spikes are very large and very sharp, and they can be driven by three factors. First, economic conditions such as dynamics of the sales process and/or competition amongst firms may lead a particular large number of firms to genuinely earn an amount that corresponds (or is extremely close to) one of the round numbers. Second, the spikes could suggest that some firms file a round-number on a monthly basis, either due to compliance costs of having to account for the exact monthly sales and/or due to incentives to mis-report the true amount of gross sales. Third, certain elements of the tax code may introduce strong incentives for gross sales to remain below/above specific amounts (e.g. under an exemption threshold), and firms may either respond in real terms by reducing/increasing production, or may respond in reported terms, by reporting a smaller/larger amount.

The following two graphs [Figure 9 and 10] display the frequency distributions for 2015 and 2016. This allows a comparison of where along the gross monthly sales axis, if anywhere; there have been changes to the frequency of filing.

Figure 9 Distribution of gross sales by year 2015

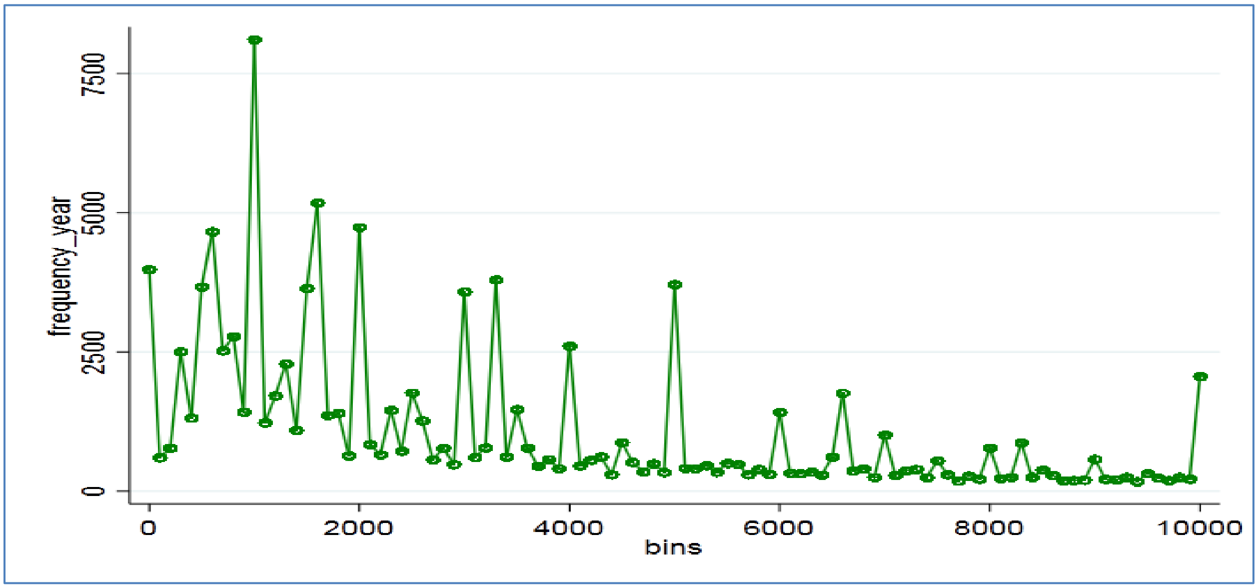
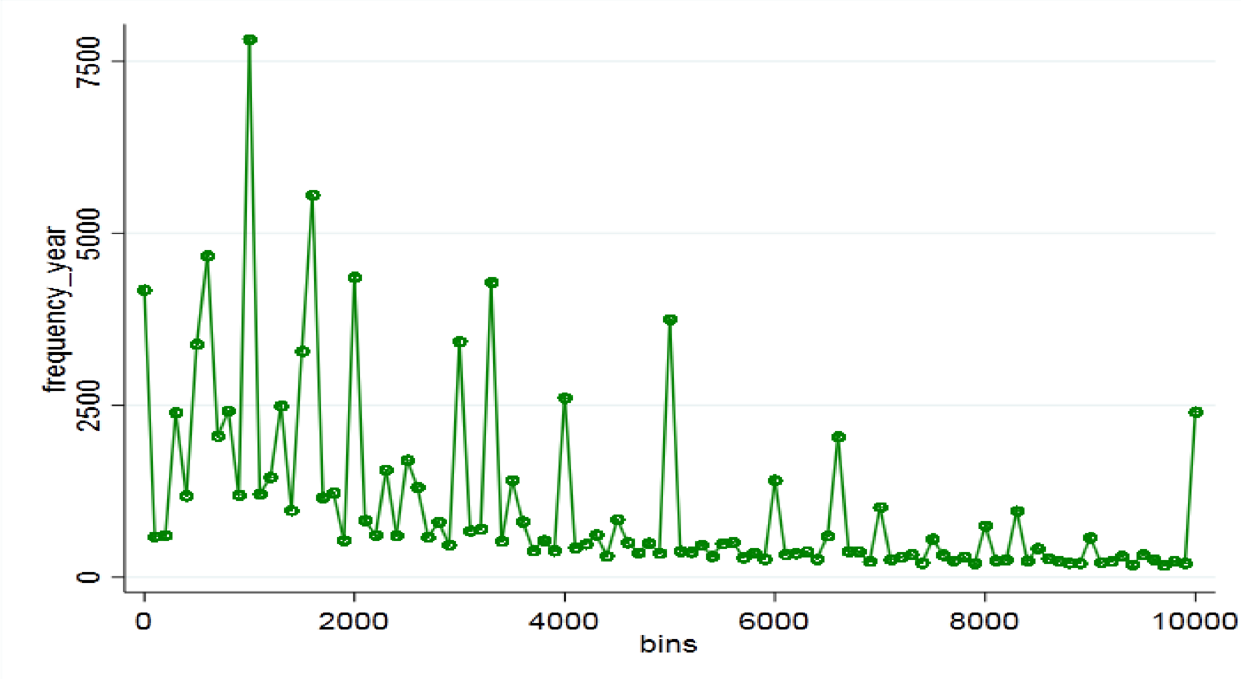


Figure 10 Distribution of gross sales by year 2016



There seem to be no particular changes in the frequency counts across groups of monthly sales, when comparing the year to year distributions.

The frequency distribution can also be separately constructed for sub-groups. In particular, we focus on the distinction between Individual and Company type of taxpayers.

Graphs [figure 11 and 12] suggest that, at almost all parts of monthly sales, the frequency distribution is larger for Individuals than for Companies. On the other hand, the sharp set of spikes around round numbers seems to present both for Companies and for Individuals.

Figure 11 Distribution of gross sales by year 2015 and type – green=individuals, blue = companies

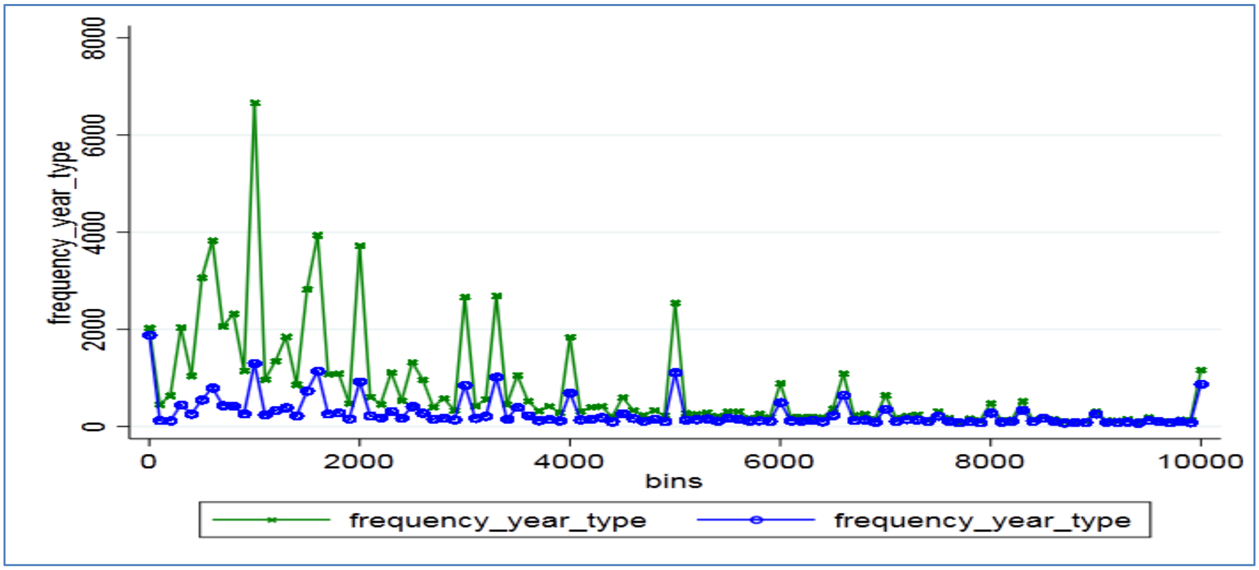
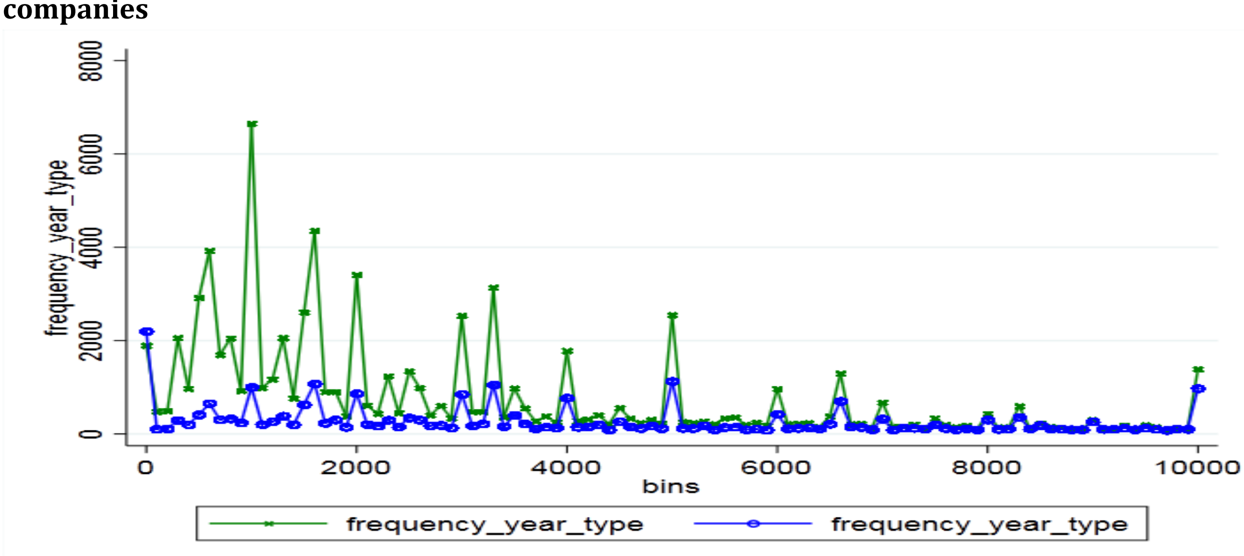


Figure 12 Distribution of gross sales by year 2016 and type – green=individuals, blue = companies



Finally, Figures 12 and 13 overlay the yearly frequency distributions, but control for type of taxpayer.

Figure 13 Distribution of gross sales by year for individuals blue=2014, green 2015, red 2016

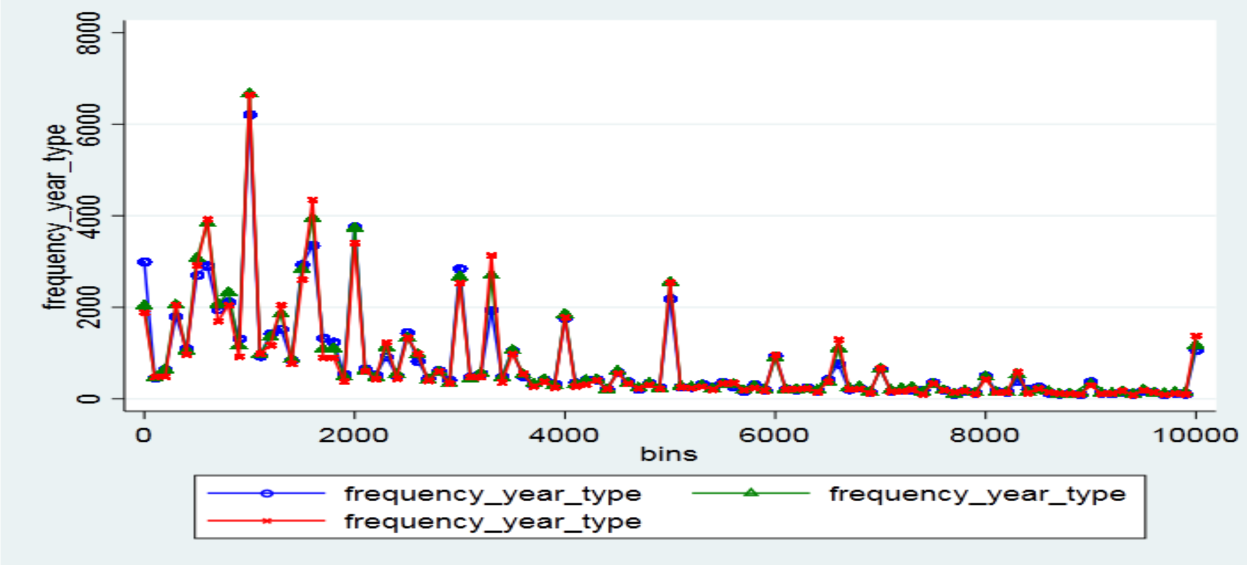


Figure 14 Distribution of gross sales by year for Companies blue=2014, green 2015, red 2016

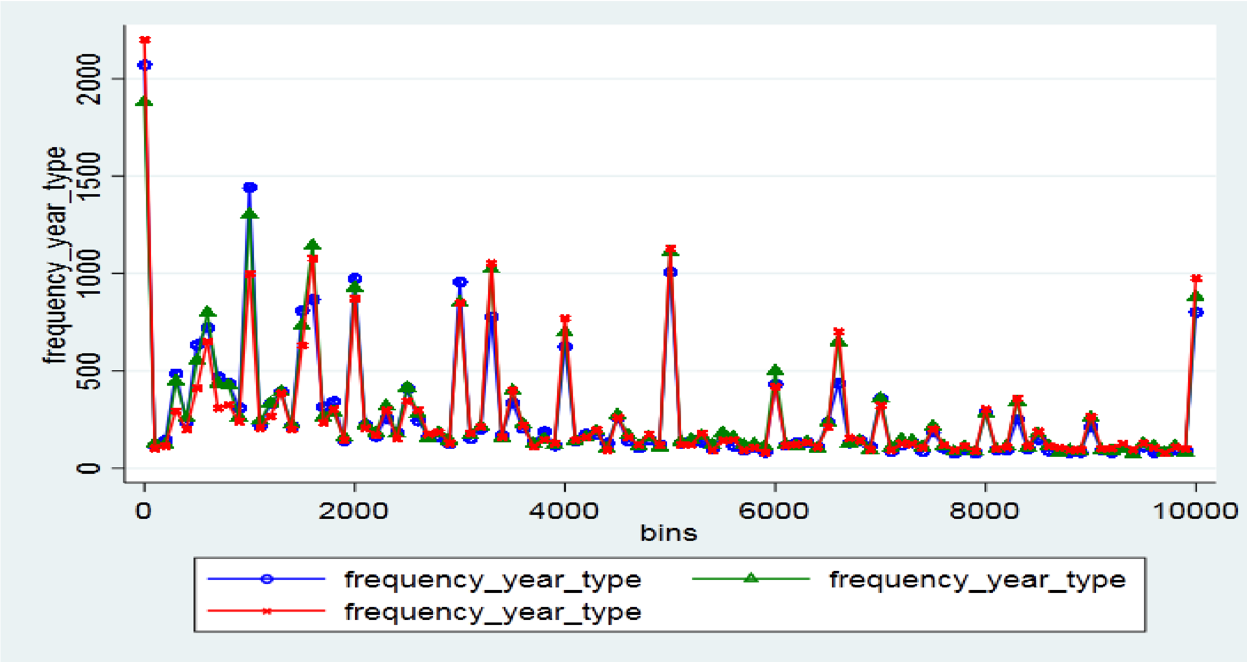


Figure 13 suggests that over the years, there has been very little change in the filing behaviour of Individuals; Figure 14 suggests a similar story for Companies, although there has been a noticeable decrease over time in the sharp filing frequency exactly at 1000, the first large round number.

Chapter Five

5.1 Results from Merging PAYE and TOT databases

5.1.1. Merging strategies

In order to determine the targeted group of potential non-compliers on TOT who should also file PAYE, we rely on a merging strategy between the TOT and PAYE databases. In this strategy, we will first determine the set of firms that are found both in the PAYE and in the TOT datasets: that is, which are the TPINs of firms which file monthly turnover tax and file monthly PAYE in parallel. These firms will be found by merging the PAYE and the TOT on the firm's unique TPIN, and possibly on time variables in addition. The set of TPINs which merge, that is the set of TPINs which are found in both the PAYE and the TOT databases, are defined to be the *complier group.* In turn, we determine the set of TPINs which do not merge, but which have characteristics that most closely resemble those that did merge. This group is the set of TPINs is the group of firms which are most likely not to have complied with the tax code, and are referred to as the *targeted group*.

5.1.2. Results from Merging on TPIN and Month and Year

In order to conduct a merging between two databases, we have to define the variables on which we want the merger to be conducted. In our setting, we report results from three different strategies.

In the first merging strategy, we require a match to occur on the TPIN, the month, and the year. The year and the month refer to the year and month in respect of which the payment was made, and not the actual year and month in which the return was physically filed with the ZRA. In this strategy, only the TPINs that we find in both datasets in the same month of the same year would be considered to have merged. The TPINs which are common to PAYE and TOT but which correspond to returns filed in different months of the same year would not be considered to have merged.

The results from performing this first merger are broken down by year, to be able to study any trends in the merge rate. The underlying sample for this exercise is the set of *monthly returns*. The merge rate is defined as the share of all monthly returns for which a merge exists. So if the same TPIN files two monthly TOT returns and two monthly PAYE returns, and both merge, then the same TPIN will count for two merged returns. On the other hand, if only one instance of PAYE and TOT returns are filed in the same month, then the same TPIN will count for only one merged return. We perform the merge rate at the level of all monthly returns, because this is the level at which the impact on total tax revenue can meaningfully be calculated. Performing the merge at the level of unique TPINs may be informative, but will not be sufficient to estimate impacts on tax revenue. Indeed, suppose that the merge was done at the level of unique TPINs and that firm A did merge, while firm B did not. That would imply a 50 percent match rate. But if firm B, that did not merge, files twelve returns within a year, while firm A only files one return within a year, the match rate at the TPIN level is misleading and grossly underestimates the impact on total tax revenue arising from non-matching.

The results of doing the merge at the monthly return level, and basing the merge on TPIN-month-year is reported in Table 3 below.

Table 5: TOT-PAYE merge rate at monthly level

Year	2013	2014	2015	2016	2017
Not merge	94.73	92.49	91.6	91.55	90.28
Merge	5.27	7.51	8.4	8.45	9.75

In every year, we report the merge rate as the share of all monthly returns filed within a year for which a merge at the TPIN-month-year level occurred. We find that the merge rate is strongly growing over time – the rate of matching goes from 5.27 percent in 2013 to 9.72 percent in 2017.

5.1.3. Results from Merging on TPIN and Year

The merging strategy in the previous subsection defined as a non-merge all instances of monthly returns for which a TPIN was found in the same year, but for which the months on the return did not match. In this subsection, we report the matching rate based on the merge at the TPIN-year level. This merge will include all matched observations from the previous sub-section, and in addition will include all instances where the same TPIN filed returns within the same year but did so in two different months of the year. The results from this merging strategy are reported in Table 6

.

Table 6. TOT-PAYE merge rate at monthly level

Year	2013	2014	2015	2016	2017
Not merge	92.34	88.24	88.49	88.2	87.4
Merge	7.66	11.76	11.51	11.8	12.6

In every year, this strategy leads to approximately an additional 3 percentage points increase in the merging rate, which represents circa a 35 percent increase in the merging rate. This means that in circa 3 percent of all monthly returns, a merge is found at the TPIN and year level, but the returns are declared to relate to payments in different months. Similarly to the previous merging strategy, the matching rate significantly improves over time – it goes from 7.66 percent in 2013 to 12.6 percent in 2017.

5.1.4. Results from Merging on TPIN

The final merging strategy is defined simply at the TPIN level. This merging strategy will therefore include all the merged monthly returns from the previous subsection, and will in addition include all instances of monthly returns where the TPIN is the same but where the reported calendar year differs. The results from this merging strategy are reported in Table 7.

Table 7. TOT-PAYE merge rate at monthly level

	Not merged	Merged
Individual	69.75	16.62
Company	30.25	83.38
Lusaka	32.77	45.20
Kitwe	6.47	15.71
Livingstone	5.83	6.16
Chipata	5.06	1.42
Kasama	4.23	0.69
Mansa	2.07	0.23
Sales: Mean	**12,151**	**46,960**
Sales: st dev	**59,888**	**114,385**

Abstracting from the year 2013, the additional increase in the merging rate is around 3.5 percentage points – when compared to the merge rate in the previous subsection.

5.2. Comparing the merged sample to the non-merged sample

5.2.1. Summary characteristics

In this subsection, we study how the sample of merged firms compare to the sample of the non-merged firms. We choose to focus on the merged and non-merged samples that derive from the matching strategy at the TPIN and year level (subsection 5.1.3). In addition, we focus on the year 2016, as this is likely to have most relevance for policy recommendations in 2018 and beyond.

To compare the merged and non-merged samples, we study differences in key characteristics across the two groups. The results are reported in Table 5.

Table 8. TOT-PAYE merge at monthly level

	Not merged	Merged
Individual	69.75	16.62
Company	30.25	83.38
Lusaka	32.77	45.20
Kitwe	6.47	15.71
Livingstone	5.83	6.16
Chipata	5.06	1.42
Kasama	4.23	0.69
Mansa	2.07	0.23
Sales: Mean	**12,151**	**46,960**
Sales: st dev	**59,888**	**114,385**

The first column describes the characteristics. These are: type of taxpayer [Individual versus Company]; location [we focus on the most commonly reported locations]; amount of monthly sales [as reported in the TOT returns]. The second column summarizes the characteristics, exclusively for the group of non-merged firms; the third column summarizes the characteristics, exclusively for the group of merged firms.

The first finding of this exercise is that while the merged firms are predominantly Companies, the non-merged firms are predominantly Individuals. Indeed, almost 84 percent of the merged taxpayers are companies, while only 30 percent of non-merged taxpayers are companies; in contrast, 70 percent of non-merged taxpayers are individuals while only 16 percent of merged companies are Individuals. This result immediately suggests that amongst the non-merged firms, the most likely group of non-compliers will be Companies.

The second finding is that the merged taxpayers are much more strongly concentrated in the largest cities – Lusaka and Kitwe – than the non-merged taxpayers. Indeed, 45 percent and 15.71 percent of merged firms are in Lusaka and Kitwe respectively, which is only the case for respectively 32.77 and 6.47 percent of non-merged firms. On the other hand, non-merged firms are more evenly spread out across other locations including Kasama and Mansa, which is not the case for the merged taxpayer types.

The third finding is that the merged firms have both much higher average monthly sales than the non-merged firms, and have much higher variance of monthly sales. Indeed, the average monthly reported sales of the merged sample are K46,959, which is 3.86 times larger than the average monthly sales of the non-reported sample (K12,151). The variance of monthly sales is almost twice as large in the merged sample than in the non-merged sample.

One issue when comparing averages is that differences in averages may be driven by a few extremely high values, which would be outliers. To investigate this possibility, we instead compare the *distribution* of monthly reported sales between the merged and the non-merged samples. To do this, we first construct percentiles of monthly sales separately in the two samples, and we then compute the average value of monthly sales in all percentiles (similar to the construction of percentile values in subsection 4.2.4). The results are displayed in figures 13 and 14

Figure 15. Average monthly sales for merged and not merged

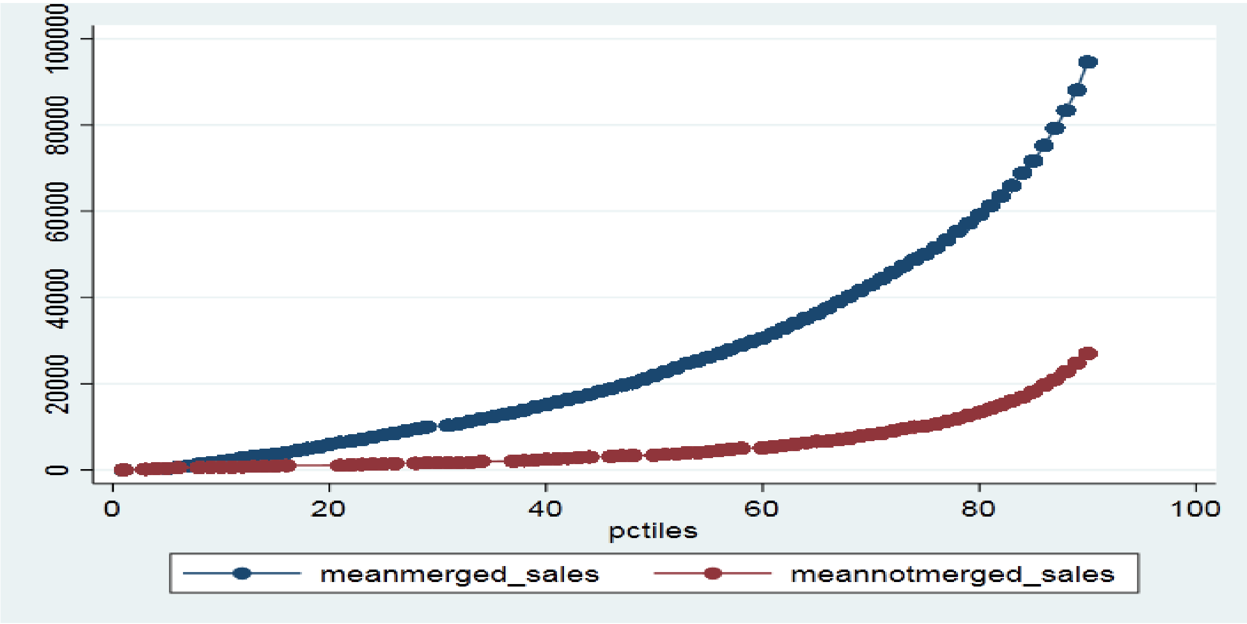
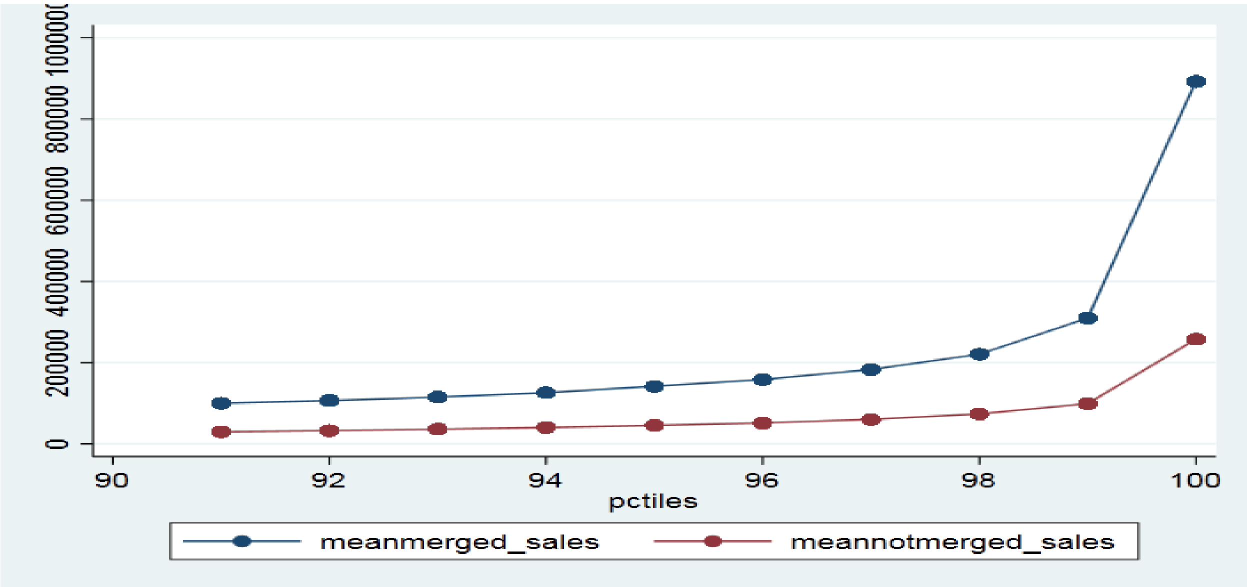


Figure 16. Average monthly sales for merged and not merged



We directly compare the average monthly sales across percentiles of the distribution of the merged sample (in blue) and the non-merged sample (in red). We find that *in all percentiles*, the average value of monthly sales is much higher in the merged sample than in the non-merged sample. This is true both for lower percentiles figure 13 and for the top 10 percentiles in figure 14. This suggests that the entire *distribution* of monthly sales is higher in the merged sample than in the non-merged sample – and thus, that the higher average monthly sales reported in Table 5 is not driven by a few outliers.

5.3. Defining the targeted group of non-compliers

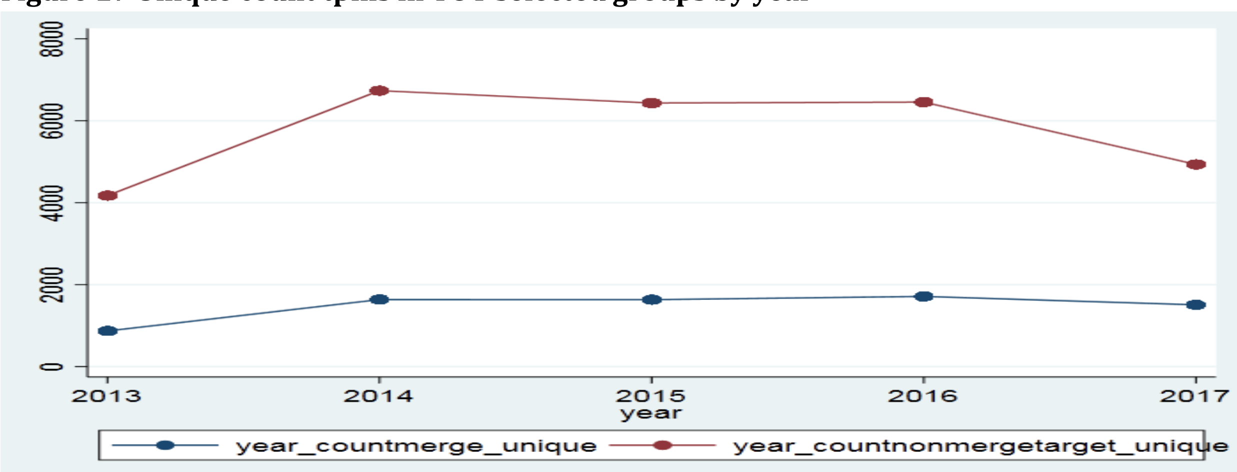
5.3.1. The targeted group of non-compliers is the set of Companies in the non-merged sample

The previous sub-section indicated that by far, the characteristic which was most strongly associated with the compliant group is the taxpayer status of Company. This makes intuitive sense. We therefore choose to define the targeted group of potential non-compliers as the group of Companies within the non-merged sample. From Table 5, we know that the targeted non-complier group represents 30.25 percent of the full set of non-merged taxpayers.

5.3.2. Counting the number of compliers and targeted non-compliers over time

To get a sense of the potential extent of non-compliance, we construct the unique number complier and non-complier TPINs in every year. We define the compliers as the full group of taxpayers that do merge, and the non-compliers as the *Companies* that did not merge. For each group, we count the unique number of TPINs at the yearly level. The results are in figure 15.

Figure 17 Unique count tpins in TOT selected groups by year



The first observation is that the number of unique TPIN non-compliers is much higher than the number of TPIN compliers. In the average year, there are around 6,000 unique non-complier TPINs, while there are only 1,900 complier TPINs. Thus, in the average year, there are roughly 3.15 times as many non-compliers TOT taxpayers as there are compliant TOT firms.

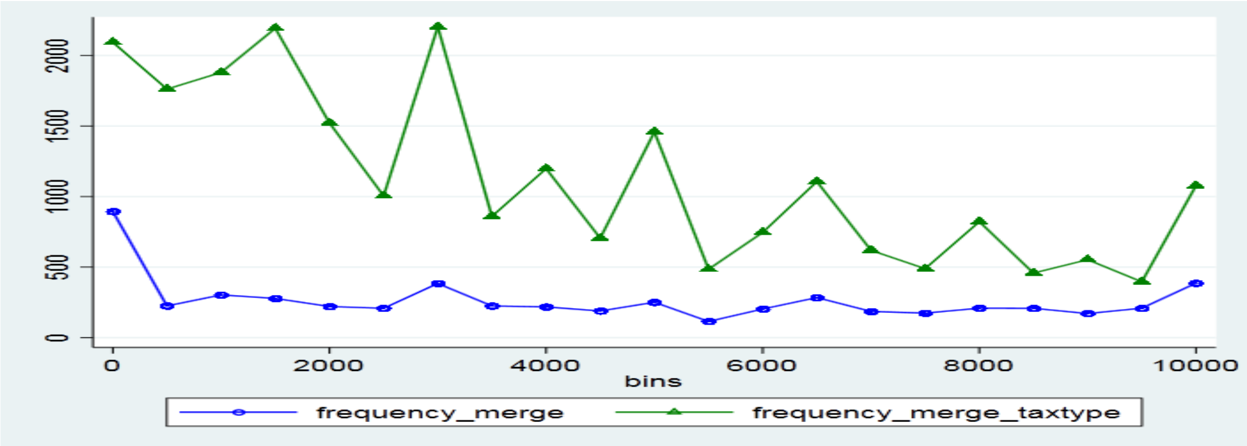
5.4. Estimating the potential number of employees to recover from the non-compliant group

5.4.1 Distribution of filing behaviour for compliant group and targeted non-compliant group

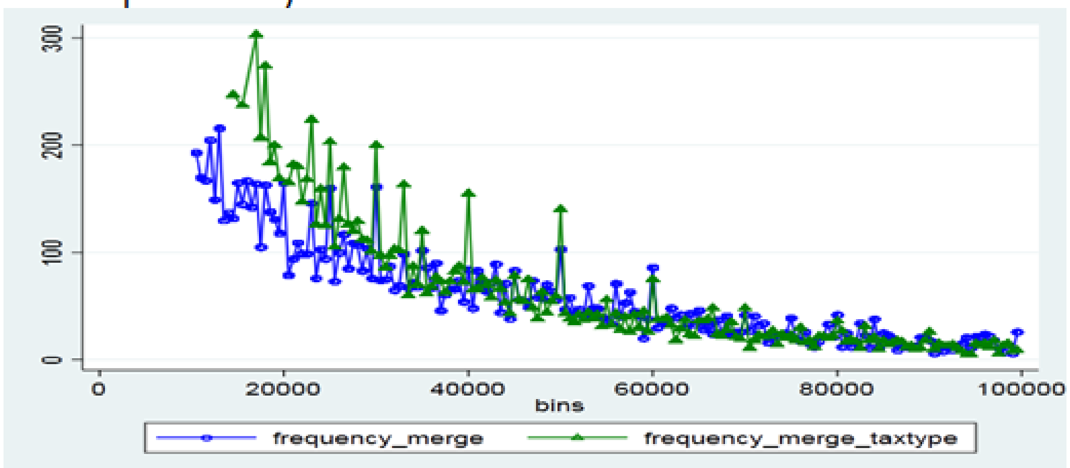
The first element of distributional analysis is to investigate differences in filing behaviour between the compliant and the targeted non-compliant group. To do this, we construct the frequency of filing distributions, separately for the compliant and the non-compliant groups. We again limit the analysis to the year 2016, in order to maximize relevance of findings for future policy decisions.

We construct bins of monthly sales, which increase in increments of K 500. Within each bin, we count the total number of returns over the full year of 2016. We count the filing frequency differently for the compliant group and for the targeted non-compliant group. The filing frequencies are displayed in Figures 16 and 17.

Figure 18. number returns against monthly turnover for merged sample in blue, green (target non-compliance)in 2016



3.3 number returns against monthly turnover for merged sample in blue, green (target non compliance)in 2016



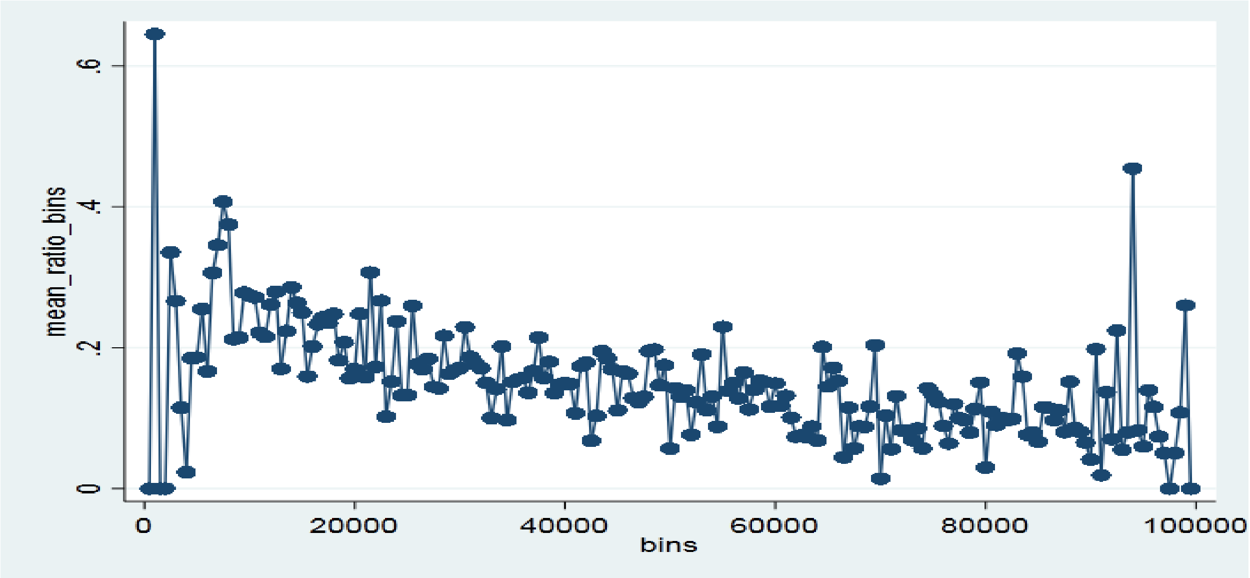
These figures split the frequency of filing to below K10,000 and to above K10,000.

The frequency of filing distribution below K10,000 reveals that both compliant and non-compliant taxpayers file frequently *at all levels of monthly turnover*. This suggests that a targeting strategy which focuses too narrowly on the non-compliant group of taxpayers at a very specific level of monthly turnover may miss out some of the non-compliant taxpayers. The frequency of filing for the non-compliant group is everywhere above the filing frequency for the compliant group. This suggests that at all levels of monthly turnover, the potential base of non-compliers is large relative to the compliers. Interestingly from circa K 40,000 monthly and above, the frequency of filing is similar in the compliant and the non-compliant groups.

5.4.2 Share of wages in total sales derived from the compliant group

An important element of the analysis to determine the potential employee and tax revenue from the non-compliant group, is to understand what share of monthly sales are in effect paid out as wages and can in turn be subject to PAYE. We investigate this empirically using the sample of compliant taxpayers. In this sample, we construct for each firm the monthly ratio of wages to turnover: We have the total monthly basic salary from the PAYE data, and we have the total monthly sales from the TOT data. In turn, we estimate the average of this ratio in all the groups of incremental monthly sales. The results from this exercise are reported in Figure [slide 43].

Figure 19 Average ratio of wages to turnover against monthly turnover for merged sample in 2016



First, there are outliers – groups of monthly sales where the average ratio is likely driven by a few extreme values. Second, and more importantly, there is an interesting trend in the ratio of wages to sales across monthly sales. The ratio is *highest for firms which report lower monthly sales,* and it steadily declines as the monthly sales increases. This means that firms which report higher monthly sales, on average pay out a smaller share of the monthly sales in wages.

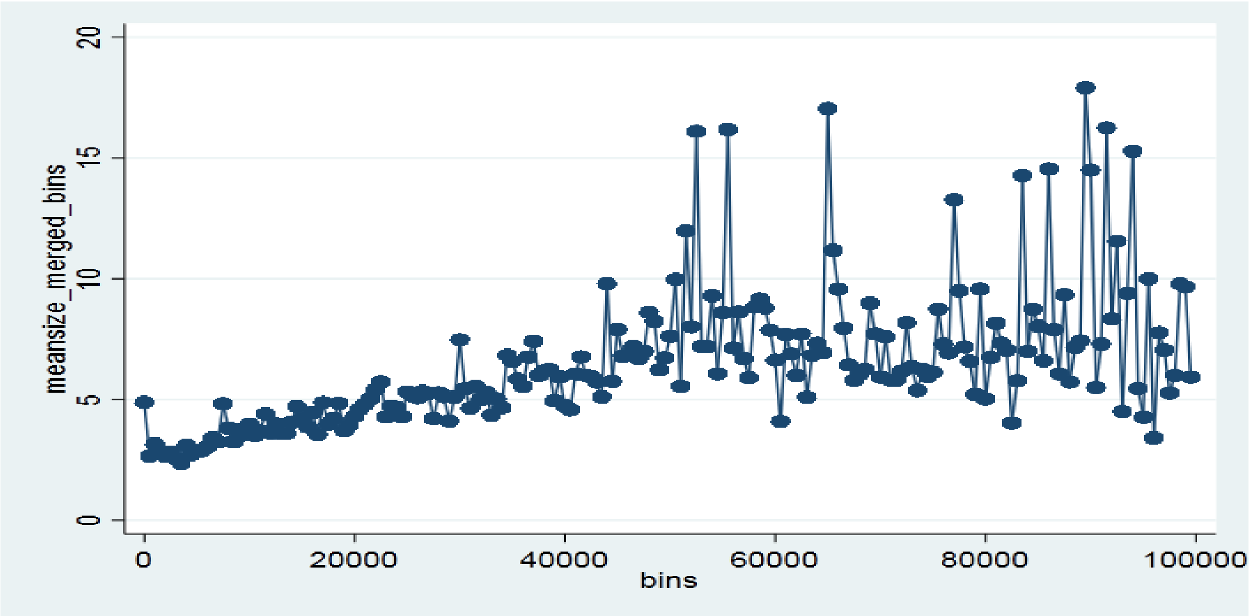
These results are based on the sample of compliant firms. The assumption is that non-compliant firms behave similar to compliant firms. That is, compliant and non-compliant firms which are both of Company type and which report exactly the same amount of monthly turnover are assumed to have similar characteristics. Under this assumption, the result is highly informative for any targeting strategy – it suggests that the potential wage base which can be enforced, as a share of turnover, will be highest for firms with the lowest monthly turnover.

5.4.3 Average number of employees derived from the compliant group

An additional element of analysis required to estimate the aggregate potential employee base that can be enforced upon from the non-compliant group, is to understand what the number of employees is for this group.

In order to understand this aspect, we again turn to the sample of compliant firms. In this sample, we construct the firm size in every month for every compliant TPIN as the number of unique reported employee TPINs (which correspond to unique NRC numbers). In turn, we construct the average firm size in a given bin of monthly sales by averaging over all the individual firm-sizes that declare monthly sales in the given bin. The results are reported in Figure 18 below.

Figure 20 Average number of employees against monthly turnover for merged sample in 2016

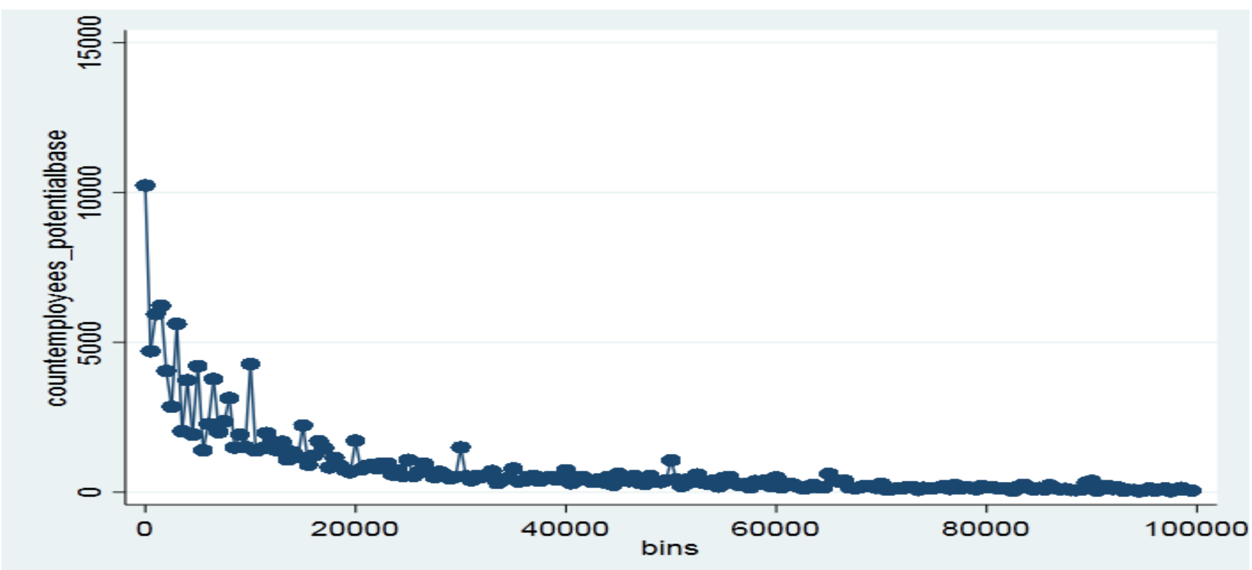


Similarly to the previous sub-section, there are outlier averages in a subset of bins. But there is also a clear trend: the firms with higher monthly turnover on average report a higher firm-size as well. Indeed, while firms with the lowest monthly turnover (between K 0 and K 5,000) report on average 3 employees, the average number of employees for firms with monthly turnover above K 50,000 is between 8 and 9.

5.4.4. Potential total number of employees recovered from the non-compliant group

We use the elements constructed in the previous subsections to estimate the potential total number of employees that can be uncovered from the non-compliant group. This analysis is conducted at the aggregate level for the entire year 2016. From subsection 5.4.3, we know the average number of employees per compliant firm, for all levels of monthly sales. From subsection 5.4.1, we know the total number of returns filed from non-compliant firms, for all levels of monthly sales. For each level of monthly sales, we estimate the total number of employees that can be uncovered across all returns by multiplying the average number of employees [from the compliant firms] with the total number of returns [from the non-compliant firms]. This provides an estimate of the total number of employees that can potentially be declared at the annual level, for all returns from non-compliant firms in a small interval of monthly sales. We repeat this analysis for all levels of monthly sales. The results are reported in figure 19 below.

Figure 21 Total potential employees against monthly turnover for targeted sample in 2016



The graph above displays the estimated total number of potential employees by monthly turnover. The graph suggests that, in absolute terms, the largest number of potential employees are concentrated amongst taxpayers with smallest monthly turnover. We know that the average employee size is increasing with monthly turnover. But since the frequency of filing is so much larger at lower monthly turnover, the total effect is that potential number of employees to be uncovered is largest at these lower levels of monthly turnover. Nonetheless, our analysis is flexible, and the potential number of employees can be studied for any small interval of monthly turnover.

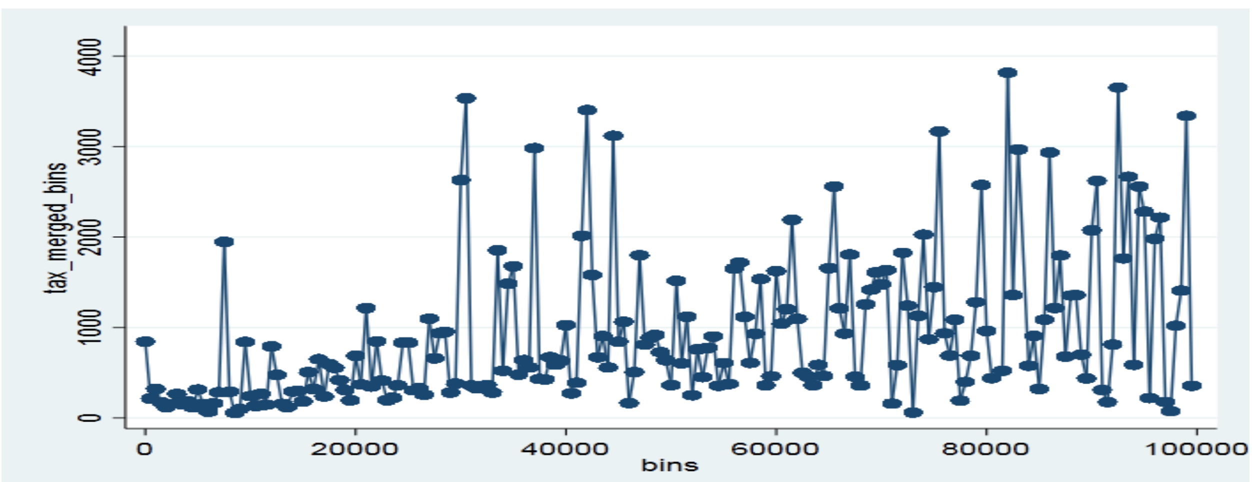
5.5 Estimating the potential PAYE taxes to be recovered from the non-compliant group

5.5.1 Average monthly PAYE taxes derived from the compliant group

To estimate the total potential PAYE to be recovered from the non-compliant group, we first estimate the average monthly PAYE taxes declared by the compliant group.

We proceed by first calculating the total PAYE declared, separately in all months and for all firms in the compliant group. Then, we construct the average monthly PAYE by bins of monthly turnover. Again, the analysis is using the full set of returns in 2016.The results are displayed in Figure 22

Figure 22 Average monthly payee taxes against monthly turnover for merged sample in 2016

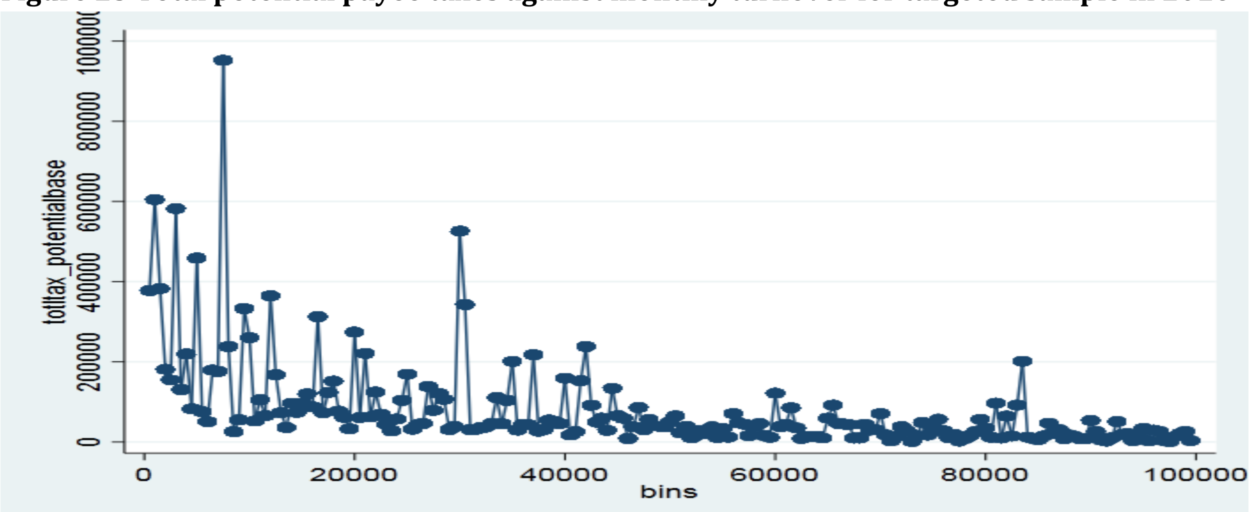


Similar to the previous graphs, there are noticeable outliers in bins of monthly turnover, which are driven by outliers. The Figure suggests that the average monthly PAYE taxes do rise with monthly turnover, but there remains large variance at all levels of turnover. At levels of monthly turnover above K 60,000, the average total monthly PAYE is around K 1000.

5.5.2 Potential PAYE taxes recovered from the non-compliant group

We use the results from subsection 5.5.1 and results from subsection 5.4.1 to calculate the total potential PAYE taxes recovered from the non-compliant group of taxpayers. This analysis is conducted at the aggregate level for the entire year 2016. From subsection 5.5.1, we know the average monthly PAYE taxes per compliant firm, for all levels of monthly sales. From subsection 5.4.1, we know the total number of returns filed from non-compliant firms, for all levels of monthly sales. For each level of monthly sales, and similarly to potential employees, we can therefore estimate the total PAYE that can be uncovered across all returns by multiplying the average monthly PAYE [from the compliant firms] with the total number of monthly returns [from the non-compliant firms]. This provides an estimate of the total PAYE that can potentially be declared at the annual level, for all returns from non-compliant firms in a small interval of monthly sales. We repeat this analysis for all levels of monthly sales. The results are reported in figure 23.

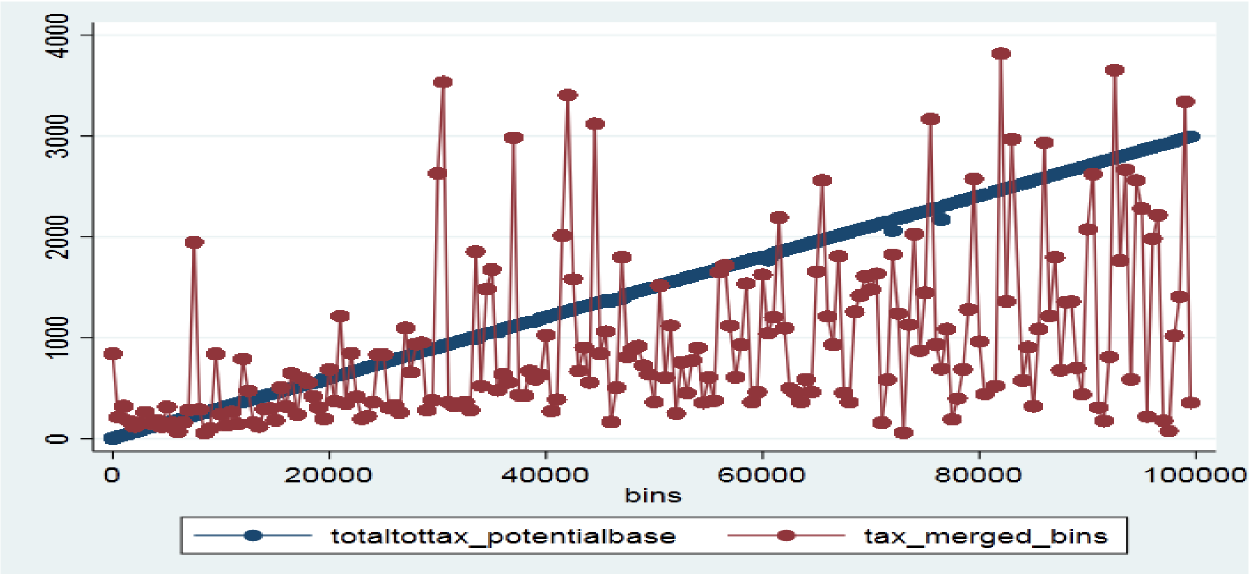
Figure 23 Total potential payee taxes against monthly turnover for targeted sample in 2016



The graph above displays the estimated total PAYE by monthly turnover. The graph suggests that, in absolute terms, the largest amount of PAYE is concentrated amongst taxpayers with smallest monthly turnover. We know that average monthly PAYE is increasing with monthly turnover. Similarly to the potential employees result, since the frequency of filing is so much larger at lower monthly turnover, the total effect is that potential PAYE to be uncovered remains largest at lowest levels of monthly turnover. Our analysis remains flexible, and the potential total PAYE can be studied for any small interval of monthly turnover.

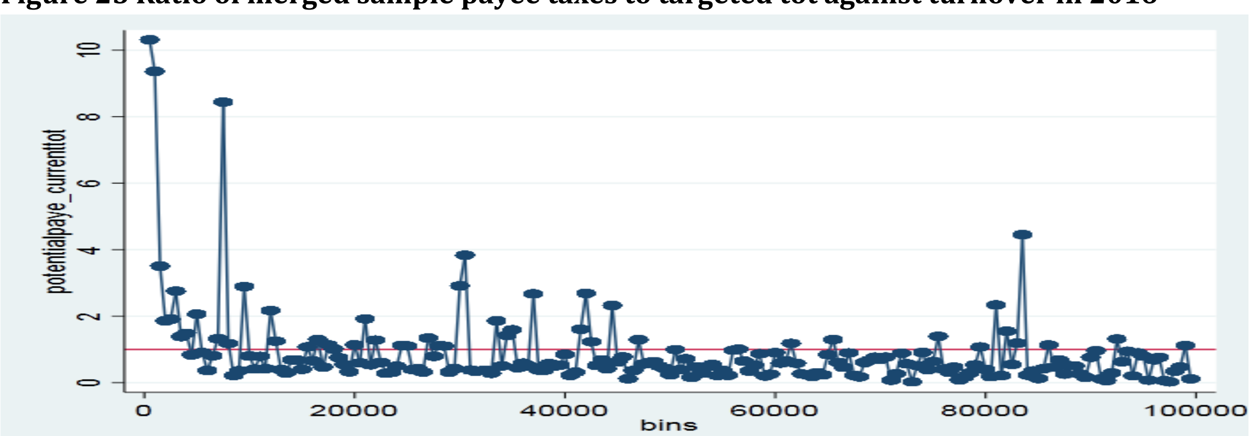
We construct additional graphs to help assess the magnitude of the potential PAYE taxes that can be uncovered.

Figure 24: Average payee taxes by merged sample and average tot by targeted sample against turnover in 2016



In figure 24, we compare, for each level of monthly turnover, the average amount of PAYE taxes declared by the compliant firms with the average amount of TOT taxes declared by the non-compliant firms. While there is a great amount of variance in average PAYE taxes, the striking result is that at most levels of monthly turnover, the amount of TOT taxes by non-compliant firms is almost equal to (or is slightly larger) the amount of PAYE taxes declared by compliant firms. This suggests that, at the level of individual non-compliant firms for a given precise amount of monthly turnover, the potential PAYE to be recovered is equal to (or is slightly lower than) the current amount of TOT taxes paid.

Figure 25 Ratio of merged sample payee taxes to targeted tot against turnover in 2016



The graph above provides an alternative way to see the same outcome as in figure 24: for each monthly turnover, it constructs the ratio of average PAYE taxes from compliant firms to average TOT taxes from non-compliant firms. Apart from outliers, this ratio is equal to (or is slightly lower than) 1 for almost all levels of monthly turnover.

Chapter Six: Conclusions and Recommendations

6.1 Summary of findings

We summarize here the results uncovered in Sections 5.4 and 5.5. In section 5.4., we estimated the total number of potential employees that could be uncovered from the non-compliant firms. We found that while the average number of employees per taxpayer does increase with monthly turnover, the frequency of filing is dominated by firms at the lowest levels of monthly turnover. In Section 5.5., we estimated the total amount of potential PAYE taxes that could be uncovered from the non-compliant firms. We found that while average monthly PAYE is higher for firms with large monthly turnover, the frequency of filing is largest at lower levels of monthly turnover. Again, the second effect dominates: the total amount of PAYE to be uncovered is by far largest at lower levels of monthly turnover.

Considering the aggregate amounts, we find that if all non-compliant firms were to become compliant, we estimate that the total number of uncovered employees is 171,007. Similarly, the annual total amount of PAYE taxes that could be uncovered is estimated to be K 23,430,151. This amount is more than the K 18,003,775 which the compliant firms paid as PAYE in 2016.

6.2. Assumptions

We emphasize that our estimates are based on several assumptions. The first assumption is that all non-compliant firms behave exactly like the compliant firms. In practice, we are assuming that merged and non-merged firms which are both of the Company type, and which file monthly returns in the same small interval of turnover, have the same number of employees and pay out the same amount of total salaries to these employees.

The second assumption is that enforcement on the targeted firms will lead to 100 percent compliance amongst the currently non-compliant firms. If the enforcement does not lead to 100 percent compliance, then our results over-estimate the potential to be recovered. Note that on the other hand, if the compliant firms are not 100 percent compliant, then our results under-estimate the potential to be recovered.

6.3. Recommendations

The targeting strategy will depend on observable firm characteristics and the amount of turnover reported in the TOT returns. Our results suggest that compliant firms have the company taxpayer type – thus, the most likely non-complier observable characteristic is company taxpayer type. In addition, while both average number of employees and average amount of PAYE is higher for firms with larger monthly turnover, the frequency of filing for non-compliant firms is incredibly large at lower levels of monthly turnover.

Based on these findings, we recommend two complementary strategies.

- Under Strategy 1, the targeting takes place for all non-merging firms that are Companies and that file monthly turnover at an exact round number below K 10,000 monthly: K1,000; K2,000; K3,000; K4,000; K5,000; K6,000; K7,000; K8,000; K9,000; K10,000. Under this strategy, the estimated annual potential PAYE taxes are K 8,654,555; and, the annual potential number of employees is 77,392.11.
- Under Strategy 2, the targeting takes place for all non-merging firms that are Companies and that file monthly turnover between K 90,000 and K 150,000. Under this strategy, the estimated annual potential PAYE tax is K 4,854,603; and, the annual potential number of employees is 30,958.54.

One possible intervention would be to select the targeted firms under either of the two strategies. In turn, use the information available on these firms to locate their physical address. This may only have an impact for a sub-set of the targeted firms, where the reported address can be found physically, and where the taxpayers are still physically residing. For this sub-set of targeted firms, the intervention to increase compliance and recover non-declared employees and PAYE could either be in the form of sending deterrence letters, or could be in the form of surprise visits by tax inspectors.

6.0 References

Alm, J. (2012). "*Measuring, Explaining, and Controlling Tax Evasion*: Lessons from Theory, Experiments, and Field Studies." International Tax and Public Finance 19 (1): 54-77.

Andreoni (1998). "*Tax Compliance. Journal of Economic Literature*" Vol. 3 American Economic Association.

Berhan, B., A and G P. Jenkins (2005), "*The high costs of controlling GST and VAT evasion*", Vol. 53, No. 3, available at http://mail.accessed on September 2016.

Bloomquist, K., Emblom, E., Johns, D., and Langetieg, P. (2012). "*Estimates of the Tax Year 2006 Individual Reporting Gap*." Washington, DC: Internal Revenue Service.

Cagan, P. (1958). "*The Demand for Currency Relative to the Total Money Supply*" (1958) 66(3) Journal of Political Economy 302-328;

Chau G., Leung P. (2009). "*A critical review of Fischer tax compliance model*: A research synthesis; Journal of Accounting and Taxation Vol.1 (2),

Damme, L, T. M and S. Orel (2008), "*Taxation Policy in Developing Countries*: available at <http://visar.csustan.edu/aaba/LaurenDamme.pdf>, accessed on September 2016

Erard, B, and Feinstein, J, (2001). "*Models of Household Tax Underreporting and the NRP Examination Process*." Presentation to IRS Research Conference. Washington, DC. June 13.

Eriksen, k., and Fallan, L. (1996). "*Tax knowledge and attitudes towards taxation*", A Report on a quasi-experiment. Journal of Economic Psychology.

Feige, E., L. (1989) "*The Underground Economies*: Tax Evasion and Information Distortion", Cambridge University Press.

Fuest, C and Riedel, N (2009). "*Tax evasion, tax avoidance and tax expenditures in developing countries*: A review of the literature (2009) <http://www.sbs.ox.ac.uk/centres/tax/Documents/reports/TaxEvasionReportDFIDFINAL1906.pdf>]

Gemmell N., Hasseldine J. (2012): *The Tax Gap: A Methodological Review*. "Advances in Taxation", Vol. 20, pp. 203-231.

Internal Revenue Service (2009). "*Income Tax Compliance Research: Net Tax Gap and Remittance Gap Estimates*." (Supplement to Publication 7285). Publication 1415 (4-90). Washington, DC.

Joulfaian, J and Rider, M (1998), '*Differential Taxation and Tax Evasion by Small Business'*. National Tax Journal Vol. 675-687.

Kabaso, P. N., and Phiri, S. C. (2012). *Taxation of the Informal Sector in Zambia*. Lusaka: Zambia Institute for Policy Analysis and Research (ZIPAR).

Kaufmann, D. and Kaliberda, A. (1996) '*Integrating the Unofficial Economy into the Dynamics of Post Socialist Economies*. Journal of Economic Literature Vol. 3

Lewis, A. (1982). *The psychology of taxation*. Oxford: Martin Robertson

Lymer, A., and Oats, L. (2009). *Taxation: Policy and Practice*. 16th ed. Birmin Fiscal Publications. available at, http://www.accenture.com

Martinez-Lopez, D. (2012). *The underreporting of income by self-employed workers in Spain*. Seville: SpringerLink.com.

Mazur, J. and Plumley, A. (2007). "*Understanding the Tax Gap*." Presentation to National Tax Association Spring Meeting. May. Washington, DC.

McManus, Jacqui and Neil Warren (2006). "*The Case for Measuring the Tax Gap*." ejournal of Tax Research <http://www.austlii.edu.au/au/journals/eJTR/2006/3.html#fn1>

O'Doherty M. (2014): *Thinking and Learning in the Tax Evasion Game.* "Fiscal Studies", Vol. 35, No. 3.

Organisation for Economic Cooperation and Development (2002), *Measuring the non-observed economy*: A handbook, OECD 2002. http://www.oecd.org/std/na/1963116.pdf

Pissarides, C. A., and Weber, G. (1989). *An Expenditure Based Estimate of America's Black Economy*. Journal of Public Economics: Journal of Public Economics 39 (1)

Silvani, Carlos and John Brondolo. 1994. "*An Analysis of VAT Compliance*." International Monetary Fund Fiscal Affairs Department mimeo, November

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about our work on our website www.theigc.org

For media or communications enquiries, please contact mail@theigc.org

Subscribe to our newsletter and topic updates www.theigc.org/newsletter

Follow us on Twitter @the_igc

Contact us
International Growth Centre,
London School of Economic and Political Science,
Houghton Street,
London WC2A 2AE

International Growth Centre

DIRECTED BY

LSE

UNIVERSITY OF OXFORD

FUNDED BY

UKaid from the British people

Designed by soapbox.co.uk