



African Tax Administration Paper 3

Are There Any Reliable Data on Wages in Low-Income Countries? Observations and Lessons from Ethiopia

Andualem Mengistu and Giulia Mascagni

September 2018

ICTD African Tax Administration Paper 3

**Are There Any Reliable Data on Wages
in Low-Income Countries?
Observations and Lessons from
Ethiopia**

Andualem Mengistu and Giulia Mascagni

September 2018

The ICTD's African Tax Administration Papers (ATAPs) are research papers that will be of specific interest to people working in tax administration in Africa. The purpose is to encourage and support useful, policy-oriented research by African professionals directly involved in tax administration.

Are There Any Reliable Data on Wages in Low-Income Countries? Observations and Lessons from Ethiopia
Andualem Mengistu and Giulia Mascagni
ICTD African Tax Administration Paper 3
First published by the Institute of Development Studies in June 2018
© Institute of Development Studies 2018
ISBN: 978-1-78118-477-6



This is an Open Access paper distributed under the terms of the Creative Commons Attribution Non Commercial 4.0 International license, which permits downloading and sharing provided the original authors and source are credited – but the work is not used for commercial purposes. <http://creativecommons.org/licenses/by-nc/4.0/legalcode>

Available from:
The International Centre for Tax and Development at the Institute of Development Studies, Brighton BN1 9RE, UK
Tel: +44 (0) 1273 606261
Email: info@ictd.ac.uk
Web: www.ictd/en/publications

IDS is a charitable company limited by guarantee and registered in England
Charity Registration Number 306371
Charitable Company Number 877338

Are There Any Reliable Data on Wages in low-income countries? Observations and Lessons from Ethiopia

Andualem Mengistu and Giulia Mascagni

Summary

Administrative data from tax returns have recently become available in many African countries thanks to the modernisation of revenue authorities and the digitisation of tax records. The availability of these data has opened new opportunities for policy-relevant analysis of real-life taxpaying behaviour. However, despite the increased availability of these data, a gap remains in many countries in relation to administrative records on employment incomes. Without these records, it is very difficult to obtain a realistic picture of income distribution and its evolution, and to understand how policies and macroeconomic conditions, such as high inflation, affect the welfare of workers. In this study, we explore the current state of administrative wage data in Ethiopia and the main challenges that limit their availability and reliability, and identify three key patterns. First, the administration of employment income taxes and social security still relies heavily on hard copy records. Second, in some cases the hard copy records are summary data at an institution rather than individual level. Third, although there are various initiatives to organise individual-level digital wage data, the digital datasets stored in the various institutions (however incomplete) do not communicate with each other. This is because the datasets do not use a similar identifier for individuals due to the fact that a strikingly large proportion of employees do not have a taxpayer identification number (TIN). In order to facilitate the availability of a digital individual-level wage data, it is paramount to eliminate duplication of effort in which two or more institutions collect similar information. Next, creating an environment in which the TIN is essential for most administrative activities may encourage people to possess one and this in turn will facilitate the compatibility of the various databases. Finally, all databases should be designed in a way that incorporates the TIN as a potential identifier.

Keywords wage data, taxpayer identification, simulation of tax policy.

Giulia Mascagni is Research Director at the International Centre for Tax and Development, and Research Fellow at the Institute of Development Studies.

Andualem T. Mengistu is Director of the Macroeconomics and Trade Centre at the Ethiopian Development Research Institute.

Contents

1	Introduction	5
1.1	Available data sources on employment incomes	5
1.2	Relevance to policy and research: not ‘just’ data	6
1.3	Objectives of this study	7
2	Availability of administrative wage data and key challenges	7
2.1.1	Ministry of Finance and Economic Cooperation (MOFEC)	8
2.1.2	Ethiopian Revenue and Customs Authority (ERCA)	10
2.1.3	Private Organizations Employees’ Social Security Agency (POESSA)	11
2.1.4	Public Servants’ Social Security Agency (PSSSA)	12
3	Lessons learned and next steps	14
3.1	Key results	14
3.2	Recommendations	16
4	References	18
5	Appendix A: Tables	19

Acronyms

ERA	Ethiopian Roads Authority
ERCA	Ethiopian Revenue and Customs Authority
IFMIS	Integrated Financial Management and Information System
LSMS	Living Standards Measurement Survey [World Bank]
LTO	Large Taxpayers Office
MOCIS	Ministry of Civil Service
MOE	Ministry of Education
MOFEC	Ministry of Finance and Economic Cooperation
MOFED	Ministry of Finance and Economic Development
MOH	Ministry of Health
POESSA	Private Organizations Employees’ Social Security
PSSSA	Public Servants’ Social Security Agency
SIGTAS	Standard Integrated Government Tax Administration System
TIN	taxpayer identification number

1 Introduction

Administrative data from tax returns have recently become available in many African countries thanks to the modernisation of revenue authorities and the digitisation of tax records. The availability of these data has opened new opportunities for policy-relevant analysis of real-life taxpaying behaviour. Recently, researchers have pioneered tax research using administrative data in African countries and Ethiopia has been at the forefront of this trend. However, despite the increased availability of these data, a gap remains in many countries in relation to administrative records on employment incomes. Without these records, it is very difficult to obtain a realistic picture of the income distribution and its evolution, and to understand how policies and macroeconomic conditions, such as high inflation, affect the welfare of workers. This study explores the current state of administrative wage data in Ethiopia and the main challenges that limit their availability and reliability, as well as making some reflections on the way forward. Although it focuses on a specific country, other low-income countries face the same data gap and can therefore learn important lessons from this study.

1.1 Available data sources on employment incomes

In many low-income countries, including Ethiopia, administrative data on wages are currently not available in a format that can be used for research. For example, the Ethiopian Revenue and Customs Authority (ERCA) collects this information mostly in hard copy form and in an incomplete format, so there is no detailed digital record at the individual employee level. The information currently available in digital format can only be disaggregated at the employer-level, not at the employee level (see Section 2). Hence, these data cannot be of much help in analysing inequality, income distribution, and the effect of tax policy. Even the aggregate data that is currently available is of very poor quality. As a result, at the moment it is very hard to obtain even basic information on employment incomes, such as the exact number of employees who currently pay tax in Ethiopia. It seems clear that the potential of these data for both tax administration and research remains untapped.

Due to the lack of administrative data on wages, researchers have often relied on survey data.¹ However, it is typically very difficult to get reliable income information from surveys. In some cases, income data is considered so unreliable that it is not even collected. One such case is the main household survey of Ethiopia, the HICES (Household Income, Consumption and Expenditure Survey)² that stopped including income information since 2010/11. As a result, many survey-based studies on inequality use consumption to proxy income (e.g. Hill *et al.* 2017), with the obvious omission of important information on savings and wealth accumulation. The main problem with using consumption as a proxy for income is that the proportion of income going to consumption (especially on basic goods) decreases as people get richer. In addition, the richest often fail to take part in surveys, partly because they may not be willing to reveal their true amount of consumption and/or income. One consequence of such a flaw is that studies based on survey data may end up underestimating inequality.

¹ A summary of surveys available in Ethiopia can be found in Mengistu, Molla and Woldeyes (2015).

² It became HCES (Household Consumption Expenditure Survey) after income data stopped being collected in 2010/11.

The difficulty of capturing income accurately with surveys is one of the main reasons why there has been an increased interest in collecting data on incomes from tax returns. Perhaps the most prominent of such efforts is the work on inequality that culminated in the bestselling book by French economist Thomas Piketty (2014), *Capital in the Twenty-First Century*. In a similar vein, a group of economists has been putting together a 'World Wealth and Income Database'. This dataset aims to include a more representative group of countries than previous ones, which typically excluded most countries in Africa. Despite these efforts, low-income countries still remain under-represented in existing data and knowledge. However, it is important to note that some research has been done to analyse inequality and the impact of policies on the income distribution using surveys. In addition to efforts under the Commitment to Equity initiative (for Ethiopia, see Hill *et al.* 2017), other studies in Ethiopia applied microsimulation techniques to investigate the impact of tax reforms on income distribution and government revenue (e.g. Mengistu, Molla and Woldeyes 2015, and Hirvonen, Mascagni and Roelen 2018). Most existing studies rely on the World Bank's Living Standards Measurement Survey (LSMS) that, contrary to HICES, does collect data on household incomes. The LSMS, however, does not contain information on whether the household pays taxes, which makes it in practice difficult to use for analysing the impact of tax changes.

1.2 Relevance to policy and research: not 'just' data

As far as policy is concerned, there are three main areas where administrative data on wages would be particularly relevant. First, this information is crucial for tax collection and for fighting tax evasion (Kleven *et al.* 2011; Pomeranz 2015). Taxes on wages are typically subject to third-party reporting, in the sense that the employer withholds and pays taxes on behalf of employees. These withholding procedures generally resulted in lower evasion for employees than, for example, the self-employed (Kleven *et al.* 2011). However, incomplete datasets severely limit the effectiveness of such third-party reports, as the information necessary to prevent and detect evasion is not available in a format that is readily usable by revenue administrations. As we show in Section 2, the current system for collecting wage data leaves a large margin for misreporting and under-reporting that can go undetected even if withholding procedures are in place.

Second, these data can play a key role in supporting reform. For example, Ethiopia recently adopted a new income tax law that, amongst other things, changed the thresholds defining the average rate of personal income tax. In principle, the reform resulted in a decrease in the tax burden for the poorest taxpayers. However, without data on employees' wages, it is very hard to carry out the necessary analysis to evaluate both the current reform and further policy changes. Research using administrative tax data for other tax types (corporate income tax) has shown that the tax system often has different implications in practice than it was designed to have (Mascagni and Mengistu 2018). It is therefore imperative to use taxpayer data to evaluate the effects of any reform in practice, as well as feeding those results into further efforts to reform policy.

Third, keeping this data in a digital and organised format would almost certainly simplify bureaucratic processes – with obvious advantages both for the revenue authority and for taxpayers who face high compliance costs in dealing with a complex bureaucracy. We give a few examples of this in Section 2.

As far as research is concerned, making administrative data accessible for researchers can open new opportunities for policy-relevant analysis. The list below

provides a few examples, specifically for administrative data on wages, though it is not exhaustive:

1. *Inequality*: Administrative wage data would allow the computation of top income shares and a more complete evaluation of income inequality than what is allowed by survey data alone.³ Using these data, researchers could assess the role of tax policy, particularly income taxation, in tackling inequality and redistributing resources. This would include the evaluation of recent and prospective tax reforms.
2. *Gender and income distribution*: These data would open the possibility to analyse gender pay gaps and differences in the wage distribution. This may be possible either by using the gender information available in tax records, if any, or by inferring gender through first names.
3. *Tax administration analysis*: Researchers could map compliance gaps or evaluate measures to encourage compliance across types of incomes, namely those that are self-reported and third-party reported ones.
4. *Labour dynamics*: These data could be used to explore trends in wages and labour participation across sectors, regions, or time.

1.3 Objectives of this study

The overarching objective of this study is to explore what income data are available in Ethiopia and if they are reliable. Given the well-known challenges of survey data, especially concerning the accuracy of income reports and coverage for top incomes (see Section 1.1), we focus particularly on administrative data. More specifically, we aim to explore three main questions:

1. What is the current state of administrative data on wages at the individual level?
2. What are the main constraints (political, administrative, and financial) to keeping these data in an organised and accessible format?
3. What are the ongoing initiatives to improve these data, if any, and what are the opportunities for doing more in the future?

In addressing these questions, we also aim to collect any lessons learned and to assess the future prospects and immediate recommendations for policymakers and the donor community. We highlight simple measures to reform the data management system as well as broader efforts that may be put in place to systematically digitise these data.

While carrying out this study, we uncovered a lot of interesting facts about the ‘plumbing’ of tax administrations and how they work, which are summarised in Section 2. We believe this information can be relevant to other countries in Africa, beyond Ethiopia.

2 Availability of administrative wage data and key challenges

To identify data sources, we started by identifying the institutions that deal with individual-level administrative data, either because of their mandate or because they happen to collect these data through their routine operations. There are four

³ An important caveat is that administrative data only includes declared incomes, thus excluding any unreported or under-reported income.

government organisations that collect individual-level data on wages: Ministry of Finance and Economic Cooperation (MOFEC), Ethiopian Revenue and Customs Authority (ERCA), Private Organizations Employees' Social Security Agency (POESSA), and Public Servants' Social Security Agency (PSSSA). However, none of these organisations provides any publicly available information about the type of data they collect and the way it is stored – let alone its quality. Preliminary consultations with the key stakeholders revealed that there is very little information about this issue, except in the offices that are specifically tasked to deal with these data that, to make things more complicated, are sometimes not located at the headquarters. As a result, we had to visit the relevant office of each institution to gather more details. We designed a brief questionnaire and administered it to selected branches of each of the four organisations mentioned above. The questionnaire was informed by our preliminary consultations, which already revealed some of the key issues we might encounter during the survey. Generally, the questionnaire aimed to gather information about what data is available and in what format, why it has not been digitised, whether digital data would be needed and if there are any ongoing efforts to digitise it, and what the main challenges and bottlenecks are. The full list of questions included in the questionnaires is provided in Tables A1a to A1d in the Appendix.

In addition to the questionnaire, we conducted open interviews to gather information and anecdotes in a more flexible and in-depth way. We also collected a range of evidence including photographs of archives and video interviews that really give a concrete sense of the issue at hand. Overall, the evidence collected is at three different levels: (1) the institution level, i.e. the four organisations mentioned above; (2) the branch level, as more than one branch was interviewed for each organisation; and (3) taxpayers within each branch, as we checked the records for a small number of randomly selected taxpayers during each interview.

As far as ERCA is concerned, we surveyed three branches: the Large Taxpayers Office (LTO), the Addis Ababa West medium taxpayer's office branch, and the Addis Ababa No. 2 medium taxpayer's office branch. For each of them, we obtained detailed information on five firms that were randomly selected. For POESSA, we surveyed 15 firms from the headquarters and five firms each from the North Addis Ababa and West Addis Ababa branches. For PSSSA, we surveyed two government organisations from the head office and three regional government offices from the central branch. Finally, we surveyed three government organisations from MOFEC's headquarters.

All interviews were carried out between November and December 2016, while some follow-up conversations occurred after that window to allow for any clarification or further information emerging from the first elaboration of data. In what follows, we report the key findings from this survey.

2.1.1 Ministry of Finance and Economic Cooperation (MOFEC)

MOFEC oversees the budget preparation in the country and follows up on the implementation and monitoring. There are about 170 federal government organisations (or 'budgetary units') that report to MOFEC. They must submit their budget to MOFEC at the beginning of the fiscal year, as well as monthly reports of payroll expenses. These reports have been submitted in different formats for a long time. To increase the efficiency and transparency of public finance, MOFEC started designing the Integrated Financial Management and Information System (IFMIS) in November 2009. The pilot phase of the IFMIS project lasted two years and was implemented in the following six government entities: Ministry of Finance and

Economic Development (MOFED), Ministry of Health (MOH), Ministry of Education (MOE), Ministry of Civil Service (MOCIS), Ethiopian Revenue and Customs Authority (ERCA), and Ethiopian Roads Authority (ERA). The pilot also included two regions, Oromiya and Southern Nations, Nationalities and People's Region (SNNPR), to assess the effectiveness of the system in the context of regional governments. Starting in November 2012, MOFEC began rolling out the system to more government entities. As of February 2017, the system had been implemented in 12 government organisations and 20 offices (some government organisations may have more than one office). The plan is to further roll out the system to all federal budgetary institutions within the next two years. Once the implementation has been completed at the federal level, regional government institutions will begin to be added to IFMIS. Therefore, it will be some time before all government-level institutions in Ethiopia are fully included in IFMIS.

IFMIS works as follows. Computers in each government entity are linked to the data centre at MOFEC. Each piece of financial information entered by any of the government entities is automatically transmitted and stored in the IFMIS database at MOFEC. A first look at the information contained in the IFMIS database reveals two important facts. First, the database has very detailed individual-level data on income and the associated taxes. Thus, it can be useful to analyse any changes in terms of payroll tax and the pension system. Second, the unique identifier for employees is an IFMIS number that may change when the employee changes the office where he/she works. Thus, tracking individuals over time may be difficult. The ideal identifier for the IFMIS database would have been the employee's taxpayer identification number (TIN). However, the system does not allow for TINs at present. Even if the TIN could be included in the dataset, it would only cover a small number of government employees, as many of them do not have a TIN – despite it being a legal requirement.

As mentioned above, IFMIS so far has only been implemented in a selected number of budgetary institutions. This raises the question: how do the rest of the budgetary institutions report their monthly payroll data? Our understanding is that MOFEC used to receive detailed information for each employee; however, because it was not used for analysis this practice was recently changed. Now MOFEC receives only a summary version from each budgetary unit, including the total payroll, total taxes, and total pensions. This information is used to release funds to the institutions to pay net salaries. This implies that MOFEC centrally retains the tax and pension for all government organisations, which are never disbursed to each budgetary unit. Our understanding is that the recent adoption of summary reports, instead of detailed ones, was at least partly motivated by the difficulty of handling large datasets. Without being properly organised in a dedicated system, they could not be used in a meaningful way for analysis and monitoring. However, we have also learned that all these budgetary institutions keep a digital record of the monthly payroll data in their own database. Therefore, MOFEC in principle could still request the detailed data from any one of them. However, each institution records the data in different and inconsistent formats. Therefore, although it may in principle be feasible to obtain employee-level payroll data for federal government employees, this would require much time and effort because of the inconsistent and scattered nature of the data. This situation also suggests that extending the data to include regional governments may be even more challenging.

Summary

The IFMIS data covers only federal government organisations, does not include TINs, and does not include government-owned enterprises. The format of individual-level data not covered in IFMIS is inconsistent and scattered.

2.1.2 Ethiopian Revenue and Customs Authority (ERCA)

In Ethiopia, revenue collection is organised according to the institutional arrangement set out in the Constitution (Federal Negarit Gazeta 1995, *Proclamation No. 1/1995*). Accordingly, the income tax revenue from employees of all private companies and firms owned by regional governments is collected by regional governments. However, since all taxes for the Addis Ababa city administration are collected by federal ERCA, the related individual-level payroll data are also reported to ERCA. Therefore, ERCA collects payroll data of employees of firms located in Addis Ababa only, while the payroll data of other employees is located in the regional revenue administrations' offices. However, at the moment ERCA does not have any information on employees of government institutions. MOFEC directly retains taxes and other contributions from employees of public institutions, as explained in Section 2.1.1. Therefore all data for the public sector are recorded by MOFEC or by the regional finance bureaus (the regional equivalent of MOFEC). The data available at ERCA represents solely the private sector, which is a major limitation of any potential future dataset based on digitised data from ERCA alone.

Our survey of ERCA branches and firms within them revealed the following results.

First, the examination of firms' records within branches reveals an interesting pattern in terms of the availability of TINs for employees. In both the LTO and the Addis Ababa No. 2 medium taxpayer's office, none of the examined hard copies of firms' reports contain a TIN. In the Addis Ababa West branch, on the other hand, three of the five firms examined have TINs in their workers' report. The rest of the firms do not report TINs for any of their employees. This suggests that a sizeable, and probably very large, proportion of administrative records do not include a TIN, which would be the most obvious and easiest form of unique identifier in this data.

Second, similar to the situation at MOFEC, ERCA digitally stores only summary information on employees' wages. All branches have the summary versions of the payroll data in the Standard Integrated Government Tax Administration System (SIGTAS) database. For each firm, the database records total payroll amount, total income taxes and pension contributions paid, and the number of employees. These summaries are at the employer level and are highly incomplete. All individual-level information is contained in its original hard copy form and has not been entered in the digital database. When we enquired why ERCA does not receive a soft copy version of the payroll data, the response mentioned two main challenges:

- Because many of the firms have a soft copy of their payroll anyway (as confirmed by our inspection of firms' records), ERCA attempted to integrate the data using the e-tax system so that they could declare online their monthly payroll data. However, the system requires each employee to have a TIN for the system to accept the payroll report. Since many employees do not have a TIN, the system did not work as expected.
- SIGTAS is becoming old and unreliable. As a result, keeping a hard copy backup has become necessary and firms have to print out the necessary records and bring them to their branch every month. This creates a clear compliance cost for the taxpayer and an administrative cost for ERCA.

Third, of the firms whose hard copy records we examined, 20 per cent report only summary versions of their payroll data, i.e. number of workers, amount of payroll tax, and pension contribution. In these cases, there is no record at all of each employee, not even in hard copy. For the remaining 80 per cent, some individual-level data is

available, with two possible options. In some cases, individual-level data is available for all employees within a firm but in most cases, firms report individual data only for new employees or for those who changed salary. The information for anyone who remains on the same contract and with the same salary does not get reported monthly. If nothing changes within a firm, they can simply attach a letter to the summary information mentioning that the details are as they reported previously. This follows guidance from ERCA, rather than being due to firms' unwillingness to disclose the full information. An interviewed official confirmed that keeping the individual hard copy records has begun to pose a problem in terms of storage and that it was unclear what use the data could be in this format. That is probably true in the current situation, as hard copy records cannot be readily and systematically used for monitoring taxpayers' compliance and detecting misreporting.

The result is that the current system leaves ample margins for under-reporting, and even an incentive to report that no change has occurred, as well as little scope for detecting evasion. The situation is likely to be similar or even worse at the regional level, where administrative capacity is typically lower than at the federal (ERCA) level.

Summary

Detailed individual-level income data is reported to ERCA in incomplete ways, none of it is digitised, and a sizeable portion of available records do not include TINs. Thus, currently this is not a good quality source of individual-level income and tax data.

2.1.3 Private Organizations Employees' Social Security Agency (POESSA)

Although the social security system has been in operation since 1963, its coverage has been limited to public service workers including the military and police. This changed following the issuance of Proclamation No. 715/2011 which stipulated that starting from July 2011 those (employees and employers) in the private sector will start contributing to the private organisation employees' pension scheme. As this scheme needs to be administered, POESSA was established in March 2011 under the Council of Ministers Regulation No. 202/2011 (Federal Negarit Gazeta 2011).

Since neither ERCA nor the regional tax offices keep complete payroll records of employees of private sector workers (as shown in the preceding discussion), POESSA has prepared its own reporting form for payroll data and has been collecting hard copies from firms. More specifically, we have learned the following from our survey of the agency.

First, all firms in the formal sector are expected to report the pension contributions of both the firm and their employees to POESSA. However, interviews with POESSA officials did not provide the exact proportion of firms that are reporting out of the total that are legally expected to. Since they do not have a good sense of what the universe of firms is, they cannot know what proportion reports and what do not. This is a good example of the basic stuff that 'we' do not know because the data is not there.

Second, all firms that file to branches of POESSA report detailed employee-level income, income tax, and pension contribution data in hard copy format. Our interviews with POESSA officials show that they know most firms have a soft copy of the report. But they keep receiving the hard copy for two reasons: (1) the report must be formally approved, stamped, and documented; and (2) the difference in the type of format each firm uses to keep their soft copy of the data means that receiving the

soft copy may create an information gap. Our interviews also revealed the view that the current hard copy records are not properly organised, as well as being vulnerable to deterioration due to humidity and pests (e.g. rats). There is a clear sense that digitising these hard copy forms is necessary and efforts to do so are underway (see below).

Third, all branches of POESSA keep the hard copies of the filings for all periods – although hard copies are not available for all firms or for all employees within each firm. The process of digitising these records is being implemented in two stages. The first stage is to digitally record all the firms that currently have records available in hard copy (this is still not the universe of firms that should be reporting to POESSA). This includes filling in the firm’s information such as name of the firm and TIN. The second stage is to fill employee-level information for those firms that have been recorded in the first stage. This means that, even when all the hard copy records are digitised, the dataset is still likely to be incomplete since there are some firms that are not reporting at all. We do not have a clear sense at the moment of when this second-stage process will be completed. The current effort is focusing on setting up the infrastructure to be populated with data. Digitising all records is expected to face challenges on various levels: technical (automatically extracting the information from scans may not be possible in some cases because records have deteriorated and/or are handwritten), personnel (related to constraints in capacity), and financial.

Fourth, the detailed employee-level information of a large number of firms is entered into a digital database. However, an interview with a POESSA official could not give us a precise estimate of the proportion of firms whose data is entered as a share of the total number of firms that report their detailed employee-level information.

Finally, the database (both the digital version underway and the hard copies) also includes TINs for those employees that have one. However, using the TIN as a unique identifier is impossible since most workers do not have one. For instance, of the total number of workers of firms whose hard copy file we investigated, only 16 per cent had a TIN. For this reason, the database allocates a unique identifier for each employee based on the first, middle, and last name, and their firm. As a result, it is virtually impossible to track employees when they change their firms. This speaks to the importance of institutional coordination to make a consistent data-gathering effort work.

Summary

Detailed information on employees is available in hard copy but is likely to be incomplete. Ongoing efforts to digitise the data face several challenges. TINs are not available for many employees.

2.1.4 Public Servants’ Social Security Agency (PSSSA)

The PSSSA administers the social security fund for all government workers. These include workers in the federal civil service, the police, the military, and regional governments. In addition, workers employed in state-owned enterprises are included in the scheme. State-owned enterprises are defined as those where the government (federal or regional) has at least a 51 per cent ownership stake.

Whereas the headquarters deals with keeping records and relevant activities of federal government workers, regional offices (branches) handle the business of workers employed by regional governments and regional government-owned enterprises. Our survey of the headquarters and the central branch of PSSSA reveals the following patterns.

First, all government organisations including the civil service, military, police, and government-owned enterprises report to PSSSA. These organisations report only a summary version of the amount of pension contributions paid in a hard copy format. Each branch of PSSSA keeps the hard copies intact. An interview with the relevant officials at the surveyed offices of PSSSA suggests that they are aware that each government institution under their purview keeps a digital version of the employee-level payroll data. However, they mention that since this data is kept in different formats and because PSSSA does not have the capacity to gather this data in one template and store it, it has not received this data so far.

Second, all interviewed officers in the surveyed branches state that having a detailed individual-level database would make a significant positive impact in terms of improving their forecast of the amount of pension contribution to be received and pension payment for upcoming retirees. This necessary digital information for forecasting is currently unavailable.

Third, since the branches do not have individual-level information they lack the means to track individuals over time. When employees retire, they bring their employment history and supporting letter from their last employer to PSSSA and pass through a lot of paperwork to start receiving their social security. This is clearly an administrative cost both for PSSSA and for pension beneficiaries. It is an even higher cost considering that the beneficiaries are elderly and may face difficulties to personally visit the relevant offices, with long travel times often involved.

Fourth, each retiree will be issued a social security ID card with a unique pension number. Retired employees' information is stored in a digital format. This makes it easy to have access to information such as how many social security recipients there are, how much each recipient receives, and the age distribution of existing recipients.

Fifth, and interestingly, PSSSA has started a pilot project to establish a digital database of the pension contribution of existing public sector workers. It covers a small number of government organisations that report to the central branch and involves two stages. In the first stage, which has been completed, software was designed to capture relevant data deemed important for pension administration. The second stage involves receiving the employee-level data from government institutions and entering it into the newly designed database. This task is currently being undertaken in the central branch of PSSSA as a pilot and it also aims to digitise older employee-level hard copy records available in the various government organisations.

Obviously, digitising individual information from the past is advantageous. However, in terms of current information most of the information contained in this database is already contained in the IFMIS database of MOFEC for the same workers. Such an overlap may be a (missed?) opportunity for synergy in the two projects.

Summary

There are no individual-level income and tax data, even in hard copy format. Digitisation efforts are underway but likely to take a long time to complete.

3 Lessons learned and next steps

3.1 Key results

Setting out to explore the current state of administrative wage data for tax-paying employees in Ethiopia, we have unveiled several issues of the current system, related to both the data and administration of the employment income tax. The challenges discussed in detail in Section 2 can be summarised in two broad categories.

Working in silos: At the start of this study, we thought that most of the relevant data would be stored at ERCA. However, the institutional picture is much more complex. There are two broad distinctions to make, which are reflected in the institutional arrangement: (1) taxes vs social security contributions; and (2) public vs private sector. Each combination of these two dimensions is represented in one of the four institutions surveyed in this study and each of these institutions records some data relative to their mandate, in the various ways described in Section 2. For example, ERCA has data on taxes, but only for private employees; while PSSSA has some data on incomes, for the purpose of calculating pensions and only for public employees. Importantly, the datasets stored in the various institutions, however incomplete, do not communicate with each other. The main reason for this is the lack of a unique identifier (i.e. the TIN). This lack of institutional coordination is a common challenge for taxation in many low-income countries. So the example we unveil here about employment income data is illustrative of an issue that is of broader relevance. It also highlights that the solution to this problem has to be centrally coordinated, probably by the Ministry of Finance as it holds ultimate responsibility for the public budget. In addition to creating inefficiencies within the public administration, the lack of institutional coordination also generates compliance costs for taxpayers, as they currently have to report very similar information to different institutions in different formats.

Issues with the 'plumbing' of tax administration: Our survey and interviews shed light on a number of practical issues related to the 'plumbing' of tax administration. We would argue that most of these are relevant beyond employment income data, although very little evidence on such issues is available from African countries. The most important ones emerging from our survey are:

- A strikingly large portion of employees in Ethiopia, from both the public and private sector, do not have a TIN. This is despite the fact that by law they should have one. This poses several problems on data and administration. Most importantly, without the TIN, it is impossible to follow employees over time, to match their records across datasets, such as social security and tax records, or different tax types that accrue to the same individual.
- There is generally very little data at the individual level, while summary data at the employer level is available in most cases. This poses a serious problem for both analysis and administration. On one hand, it prevents any analysis of the income distribution based on administrative data, including on the distributional effects of tax reforms. On the other hand, without individual data on taxpayers, it is very hard for the revenue authority to prevent and detect misreporting. It is likely that this lack of data opens up many opportunities for evasion, be it in the form of not reporting or under-reporting. Importantly, it also severely limits the self-enforcing potential of the third-party reporting system for employment taxes.

- The administration of employment income taxes and social security in Ethiopia still relies heavily on hard copy records. The only information available in soft copy is summary data (even these, not in all cases), which are often incomplete, partly because it relies on scattered hard copy records. There are some potential fixes to this issue, both for the short and longer term, which we suggest in Section 3.2.
- Even when individual-level data are available in hard copy, they are kept in different formats in the four institutions surveyed. As far as the public sector is concerned, we suspect that the various budgetary units record data in a scattered way, both in terms of completeness and format. There does not seem to be a common template for those who are not yet part of IFMIS.
- The distinction between private and public employees is reflected not only in the institutions responsible for keeping the relevant data, but also in the availability and completeness of available data, including the feasibility of organising it in a digitised dataset in the future.
- Even when individual-level data could potentially be collected in relatively easy ways, the relevant institutions often forego this opportunity because of capacity constraints that would prevent them from being able to use the data in a meaningful way. We unveiled examples of this from both ERCA and MOFEC, both of which in some instances (described in Section 2) prefer summary reports rather than large individual datasets. This situation is aggravated by the hard copy format of records, which also presents additional challenges related to storage and deterioration.

One implication of the lack of readily accessible data on individuals' wages is that it makes tax evasion easier. In Ethiopian fiscal year 2012/13, MOFEC reports that the government collected 11.567 billion birr from employment income tax. Our simulation, using data from the 2013 Ethiopian Labour Force Survey (CSA 2015), shows that in the absence of tax evasion, the government could have raised 12.827⁴ billion birr – about 10 per cent more than current collections. Although we cannot claim that this gap can be completely eliminated with the availability of digitised labour income data, it could certainly have the potential to significantly contribute to closing this gap.

Interest in terms of digital database: there seems to be much appetite to get this data in order, although this task can certainly be overwhelming. All the institutions surveyed and the people consulted recognised the issue and showed both interest and willingness to solve it. Part of the reason for this enthusiasm is that the current lack of appropriate data means that the government is currently missing very basic pieces of information, such as how many employees pay tax or how many firms are supposed to contribute to pensions. This interest is reflected in a number of initiatives, such as the roll-out of IFMIS and the efforts to digitise data currently ongoing at POESSA (see Section 2). Still, a more general effort to centrally digitise this data would certainly be a more effective and efficient approach. Our broad sense is that it will be feasible to digitise this data in the future, but this requires a streamlining of systems and guidelines that needs to come from the centre (with one obvious possibility being the Ministry of Finance). Our view is that any effort towards improving data should be based on the current efforts, of which IFMIS may have the best chance of succeeding in obtaining an individual-level dataset of employment incomes. Still, some of the challenges we identified will remain, such as the lack of

⁴ Note that the simulation excludes labour that is employed in the agriculture sector. If we include employed labour in the agriculture sector, the tax that could have been collected increases to 13.07 billion birr.

TINs and capacity constraints that may generate dis-incentives to actually collect more individual-level data.

3.2 Recommendations

A few broad recommendations stem from this study, related to both data management and tax administration more broadly. They can be summarised as follows:

- Currently, firms and government institutions are required to report similar information to two separate government institutions (ERCA and POESSA, or MOFEC and PSSSA). This represents a duplication of their compliance costs. Streamlining the information flows can start solving a good part of the problem. Firms could report to only one organisation (e.g. ERCA), while the other (e.g. POESSA) would gather the information they need directly from ERCA's database.
- The limited use of TINs will continue to hamper the usefulness of any future digital database, like those that are currently being developed at MOFEC and POESSA. A number of considerations and suggestions emerged during the project, including some lessons learned from other countries:
 - Catching up with providing TINs to all employees who currently do not have one represents a large investment, both financial (TIN ID cards are quite expensive for the government) and in terms of capacity (processing the necessary information for all employees). ERCA may not have the necessary capacity to carry this out. Therefore, decentralised ways of approaching this registration process should be considered. For example, employers could be involved and play a key role in facilitating the collection of this information and distribution of TINs.
 - Having a TIN could be made compulsory when carrying out (relatively) common activities like opening a bank account, getting a loan, buying a house, etc. This would provide an incentive to taxpayers to get a TIN ID card, which is now not necessary for virtually any of these functions or other public services (instead, the local administration, or *kebele*, ID is normally used).
- Facilitate and encourage the collection of individual-level data instead of summaries.
 - ERCA's e-tax system should be allowed to accept digital information on employees' salaries, filed by firms, even when they do not have a TIN available. Although this is not ideal (i.e. no unique identifier), at least it would represent a record of all taxpaying employees – which is currently missing. This should be a relatively quick, albeit partial, fix.
 - Similarly, for government employees, efforts could be stepped up to collect individual-level information, instead of the summaries, and to ensure this information is collected in the relevant databases (with TIN where available), including the IFMIS database.
- Digitising and integrating processes:
 - Firms should be allowed to submit their employees' data in digital format rather than physically taking a printout to the tax office. This would save time for both taxpayers and the administrators.
 - All relevant organisations should streamline and digitalise their data management processes, including POESSA and PSSSA, as they currently seem to have burdensome procedures based on hard copy documents. This would require the adaptation of current systems (e.g.

allow inputting data in SIGTAS even without a TIN), which would then provide a common template for all firms.

- Importantly, such systems should ideally be harmonised across institutions. This could encourage links between data and information systems between tax administration and social security, ideally through the TIN or another common identifier.
- One way to encourage such links is to introduce a requirement for a social security card/number to be allocated at the beginning of employment, which would then be used to claim a pension upon retirement. This would have obvious benefits in terms of forecasting pension payments, as there would be a clear picture of how many pensioners to expect at various times in the future. Ideally, the TIN would be reported in that card, to allow matching with tax data.

In the context of the substantial work on data management that seems to be necessary in Ethiopia, foreign donors can potentially play a key role. For example, a practical and relatively trivial step would be to support existing initiatives that are underway in the institutions involved. Although a more centralised approach would be ideal, as we argued before, preparing the data in the various institutions would go a long way in improving the administration of tax and social security systems and could potentially feed into a more concerted effort at a later stage.

4 References

- CSA (2015) *National Labour Force Survey 2013 (2005 E.C)*, Central Statistical Agency, Ministry of Finance and Economic Development, <http://catalog.ihsn.org/index.php/catalog/5870/study-description>
- Federal Negarit Gazeta (1995) *Proclamation No. 1/1995; A Proclamation to Pronounce the Coming into Effect of the Constitution of the Federal Democratic Republic of Ethiopia*, Addis Ababa, 21 August, <https://chilot.files.wordpress.com/2011/01/proc-no-1-1995-constitution-of-the-federal-democratic-repu.pdf> (accessed 9 August 2018)
- Federal Negarit Gazeta (2011) *Proclamation No. 715/2011; A Proclamation to Provide for Pension of Private Organization Employees*, Addis Ababa, 24 June, <https://chilot.me/wp-content/uploads/2011/07/7151.pdf> (accessed 6 September 2018)
- Hill, Ruth; Inchauste, Gabriela; Lustig, Nora; Tsehaye, Eyasu and Woldehanna, Tassew (2017) *A Fiscal Incidence Analysis for Ethiopia*, Commitment to Equity (CEQ) Working Paper Series 41, New Orleans LA: Tulane University, Department of Economics
- Hirvonen, K.; Mascagni, G. and Roelen, K. (2018) 'Linking Taxation and Social Protection: Evidence on Redistribution and Poverty Reduction in Ethiopia', *International Social Security Review* 71.1: 3–24
- Kleven, Henrik Jacobsen; Knudsen, Martin B.; Kreiner, Claus Thustrup; Pedersen, Søren and Saez, Emmanuel (2011) 'Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark', *Econometrica* 79.3: 651–92
- Mascagni, G. and Mengistu, A.T. (2018, forthcoming), 'Effective Tax Rate and Firm Sizes in Ethiopia', *Development Policy Review*
- Mengistu, Andualem T.; Molla, Kiflu G. and Woldeyes, Firew B. (2015) *Tax–Benefit Microsimulation Feasibility Study in Ethiopia*, WIDER Working Paper 2015/115, Helsinki: United Nations University World Institute for Development Economics Research (UNU-WIDER)
- Piketty, Thomas (2014) *Capital in the Twenty-First Century*, Cambridge MA: Harvard University Press
- Pomeranz, Dina (2015) 'No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax', *American Economic Review* 105.8: 2539–69

5 Appendix A: Tables

Table A1a: The list of information included in the questionnaire for ERCA

General information	Information specific to the firm
How many firms do report their payroll data to this branch of ERCA?	Name of the firm
Total number of employees of the firms reporting to this branch	Annual turnover of the firm
The quality of data management in the branch	Number of employees of the firm
Constraints in terms of keeping accurate employee-level payroll data in digital format	How many employees have a TIN?
How would having access to employee-level data improve the functioning of the organisation?	How does the firm report its data every month to ERCA? (<i>employee level, summary version for all, summary for existing employees and detailed for newer ones</i>)
	How long is the hard copy of the firm's report kept by the ERCA branch? (<i>all hard copy reports are kept, hard copies of older years are discarded, other, specify</i>)
	What does the digital database of ERCA contain for this firm? (<i>detailed payroll data of all employees, summary of all employees and details of new employees, summary of all employees for each office, other, specify</i>)

Table A1b: The list of information included in the questionnaire for MOFEC

General information	
What kinds of government organisations do report payroll data to MOFEC?	Name of the government organisation
How many government organisations do report their payroll data to this branch of MOFEC?	How many employees does the government organisation (office) have?
Do government organisations in the entire country report to the branch? (<i>01=yes, 02=no</i>)	How many of the employees have a TIN?
If only organisations in specific regions report, specify the regional distribution of the reporting government organisations	How does the office report employee-level social security contribution data every month? (<i>01=details of all employees, 02=summary for all employees and details of new employees, 03=only summary for all employees</i>)
Total number of employees of the organisations reporting to this branch	How does MOFEC keep the hard copy data? <i>01=all hard copy the report are kept, 02=hard copies of older years are discarded, 03=other, specify</i>
The quality of data management in the branch	What does the digital database of MOFEC contain? <i>01=detailed payroll data of all employees, 02=summary of all employees and details of new employees, 03=summary of all employees for each office</i>

Constraints in terms of keeping accurate employee-level payroll data in digital format	
How would having access to employee-level data improve the functioning of the organisation?	

Table A1c: The list of information included in the questionnaire for POESSA

General information	Information specific to the firm
Do firms in the entire country report to the branch? (01=yes, 02=no)	Name of the firm
If firms from a specific region report to the branch, specify the regions that reporting firms are located	Annual turnover of the firm
How many firms do report their payroll data to this branch of POESSA?	Number of employees of the firm
Total number of employees of the firms reporting to this branch	How many employees have a TIN?
The quality of data management in the branch	How does the firm report its data every month? (employee level, summary version for all, summary for existing employees and detailed for newer ones)
Constraints in terms of keeping accurate employee-level payroll data in digital format	How long is the hard copy of the firm's report kept by the POESSA branch? (all hard copy reports are kept, hard copies of older years are discarded, other, specify)
How would having access to employee-level data improve the functioning of the organisation?	What does the digital database of POESSA contain for this firm? (detailed payroll data of all employees, summary of all employees and details of new employees, summary of all employees for each office, other, specify)

Table A1d: The list of information included in the questionnaire for PSSSA

General information	Information specific to an organisation
What kinds of government organisations are included in the PSSSA scheme other than civil service, military, and police? (1) government bank employees? (2) telecommunications and electric utility? (3) industrial parks, sugar corporations, etc.? (4) others, specify	Name of the government organisation
How many government organisations do report their payroll data to this branch of PSSSA?	How many employees does the government organisation (office) have?
Do government organisations in the entire country report to the branch? (01=yes, 02=no)	How many of the employees have a TIN?
If only organisations in specific regions report, specify the regional distribution of the reporting government organisations	How does the office report employee-level social security contribution data every month? (01=details of all employees, 02=summary for all employees and details of new employees, 03=only summary for all employees)
Total number of employees of the organisations reporting to this branch	How long does the PSSSA keep the hard copy data? (01=all hard copy the report

	<i>are kept, 02=hard copies of older years are discarded, 03=other, specify)</i>
The quality of data management in the branch	What does the digital database of PSSSA contain? (01=detailed payroll data of all employees, 02=summary of all employees and details of new employees, 03=summary of all employees for each office)
Constraints in terms of keeping accurate employee-level payroll data in digital format	
How would having access to employee-level data improve the functioning of the organisation?	

International Centre for Tax and Development
at the Institute of Development Studies
Brighton BN1 9RE, UK
T: +44 (0)1273 606261
F: +44 (0)1273 621202
E: info@ictd.ac
www.ictd.ac

