Working Paper 57

The Carrot and the Stick:
Evidence on Voluntary Tax Compliance from a Pilot Field Experiment in Rwanda

Giulia Mascagni, Christopher Nell, Nara Monkam and Denis Mukama
October 2016

About ICTD

The International Centre for Tax and Development is a global policy research network dealing with the political economy of taxation policies and practices in relation to the poorer parts of the world.

Our operational objectives are to generate and disseminate relevant knowledge to policymakers and to mobilise knowledge in ways that will widen and deepen public debate about taxation issues within poorer countries. Our ultimate objective is to contribute to development in the poorer parts of the world and help make taxation policies more conducive to pro-poor economic growth and good governance.

The ICTD's research strategy and organisational structures are designed to bring about productive interaction between established experts and new stakeholders.

The ICTD is funded by the UK Government's Department for International Development (DFID) and the Norwegian Agency for Development Cooperation (Norad), a directorate under the Norwegian Ministry of Foreign Affairs (MFA).

For further information, please contact:

International Centre for Tax & Development
at the Institute of Development Studies, Brighton BN1 9RE, United Kingdom
T: +44 (0)1273 606261   E: info@ictd.ac   W: www.ictd.ac

About ATAF

The African Tax Administration Forum (ATAF) is an international membership organisation of African revenue authorities and acts as a platform promoting cooperation, knowledge sharing and capacity building among African revenue authorities. It seeks to ensure greater synergy and cooperation in capacity development amongst all relevant stakeholders to reduce duplication of work and give greater support to African Tax Administrations. From its beginning in 2009, when it was formally launched in Kampala, Uganda, ATAF is growing in stature and in influence. Today ATAF is an important voice in taxation in Africa and the world. It has achieved the status of an international organisation and its membership has grown to 38 African tax administrations.

For further information, please contact:

African Tax Administration Forum, Postnet Suite 430, Private Bag X15, Menlo Park, 0102 South Africa
T: +27 (0)12 451 8835 E: info@ataftax.org   W: www.ataftax.org

ICTD Working Paper 57

The Carrot and the Stick: Evidence on Voluntary Tax Compliance from a Pilot Field Experiment in Rwanda

Giulia Mascagni, Christopher Nell, Nara Monkam and Denis Mukama

October 2016

The Carrot and the Stick: Evidence on Voluntary Tax Compliance from a Pilot Field Experiment in Rwanda

Giulia Mascagni, Christopher Nell, Nara Monkam and Denis Mukama

Summary

Large-scale field experiments on tax compliance have been a thriving field of research in many regions of the world. However, Africa is still lagging behind, as administrative data from anonymised returns is available only in a handful of countries. To the best of our knowledge, there is as yet no published evidence of a tax field experiment from Africa. This paper reports the results of a pilot experiment in Rwanda that served as a stepping stone for a larger experimental study on tax compliance. In this pilot, we test the process of messaging taxpayers to encourage them to comply voluntarily, by providing information on sanctions. The results indicate that communication strategies that aim to inform taxpayers may be effective in increasing tax compliance. However, these results are only indicative. They will be complemented by further evidence from the larger field experiment, where we test different types of messages and delivery methods. Nonetheless, this paper provides some initial insight into the use of tax experiments in Africa, both in terms of initial evidence and lessons learned for future efforts in this field.

Giulia Mascagni is Research Fellow at the Institute of Development Studies (IDS) and Research Director of the International Centre for Tax and Development (ICTD).

Christopher Nell is Research Officer at the Institute of Development Studies (IDS) and the International Centre for Tax and Development (ICTD).

Nara Monkam is Research Director of the African Tax Administration Forum (ATAF).

Denis Mukama is Head of Research at the Rwanda Revenue Authority (RRA).

Contents

Acknowledgements

Acronyms

ATAF	African Tax Administration Forum
ATRN	African Tax Research Network
CIT	Corporate Income Tax
ICTD	International Centre for Tax and Development
ITT	Intention-To-Treat
IV	Instrumental Variable
LATE	Local Average Treatment Effect
OLS	Ordinary Least Squares
PAYE	Pay As You Earn
PIT	Personal income tax
RA	Revenue Authority
RRA	Rwanda Revenue Authority
RWF	Rwandan Franc
VAT	Value Added Tax

Introduction

The literature on tax compliance has recently seen a surge of evidence from field experiments. Field experiments use administrative data from taxpayers' records to evaluate the effectiveness of communication strategies that revenue administrations can adopt to increase compliance. These strategies typically take the form of letters delivered to taxpayers, which aim to provide information on the tax system or to change perceptions on key determinants of tax compliance, such as deterrence or fiscal exchange. Revenue authorities play a key role in this type of study, both in providing data from anonymised taxpayer records and in implementing the intervention. From a theoretical standpoint, most studies of tax compliance are based on the seminal model of Allingham and Sandmo (1972) and its subsequent developments.[1] Generally, tax compliance behaviour is determined by a mix of enforcement measures, such as audits and the effective use of third-party data, and quasi-voluntary compliance, which is motivated by factors often labelled as 'tax morale',[2] such as trust in institutions, social norms, fiscal exchange, and moral factors. One could think about deterrence measures as the 'stick', while measures to encourage voluntary compliance would be the 'carrot'. In practice, these theoretical elements interact in complex ways with individual characteristics (e.g. sector, employment status, gender) and with practical aspects of the taxpaying environment (e.g. corruption of tax collectors, accessibility of information, availability and quality of tax advisers). Field experiments test empirically which factors affect tax compliance in practice (for a detailed review, see Mascagni 2016).

Although field experiments on tax compliance have been thriving in many regions of the world, low-income countries remain under-represented and Africa is still completely absent from this literature. Therefore, we do not know whether the findings of the existing literature are valid also for African countries, or even if carrying out field experiments is feasible at all in the continent. Challenges that are specific to low-income countries may make this type of study particularly difficult in those contexts. From a practical point of view, low administrative capacity is a challenge for studies that require a high degree of commitment and involvement by the local revenue administration (RA). The use of taxpayer records for research, even if anonymised, is still rare in Africa – making many RAs unwilling or unable to grant access (Mascagni, Monkam and Nell 2016a). Moreover, taxpayer registries are often not kept up-to-date, making it potentially very hard to trace taxpayers to physical addresses to deliver messages to them. These challenges may cast doubts even on the appropriateness of the most common delivery method for messages, physical letters, in a low-income context. In other words, the standard research design developed for high-income countries may not be suitable for low-income countries. On a more conceptual level, these countries present some specific features that make the problem of tax compliance somewhat more complex than in high-income countries. Some of the most notable ones are typically large informal sectors, weak enforcement, and low level and quality of service delivery.

In this context, this paper reports the findings of a pilot experiment which has been used primarily to test the feasibility of larger-scale field experiments in Africa, and which has served as a stepping stone for a larger experimental study on tax compliance carried out in Rwanda.[3] The pilot has also enabled us to get initial insights on the effectiveness of nudges to improve tax compliance. As a pilot, this experiment did not aim to address the conceptual issues that make low-income countries different to their high-income counterparts, nor to provide conclusive results on tax compliance. Instead, our main motivation is to explore the practical aspects of carrying out tax experiments in Africa, as this pilot represents the first

---

[1] For example, see Yitzaki 1974; Erard and Feinstein 1994; Myles and Naylor 1996; and Traxler 2010.
[2] For a detailed discussion of the elements composing tax morale, see the review paper by Luttmer and Singhal (2014).
[3] The findings of this larger study are reported in Mascagni, Nell and Monkam 2016b.

publicly available evidence from an African country.[4] In doing so, we are still able to provide some initial results on the effectiveness of communication strategies as nudges to increase tax compliance. The results and lessons learned presented here fed into the design of a larger-scale, and more complex, field experiment testing the effectiveness of different messages (reminder of deadlines, deterrence, fiscal exchange) and delivery methods (letters, emails, SMS) to increase tax compliance in Rwanda (Mascagni, Nell and Monkam 2016b).

Although our intervention is a physical letter, as in other similar studies, we depart from the literature by looking at taxpayers' voluntary revisions of their accounts rather than payments or declarations. Therefore our econometric results are not directly comparable with other studies, but can only speak to the rest of the literature in general terms. More specifically, the content of our letters should nudge taxpayers into making a revision to correct their tax account. The letters do so by providing information on the relevant Rwandan laws, which provide for sanctions as high as 60 per cent if under-reported income is found through an audit, but much lower, about 10 per cent, if the taxpayer voluntarily revises their tax account. Two interrelated issues motivate our focus on revisions. First, taxpayer collaboration is particularly important in low-income countries, where tax authorities face particularly severe administrative and financial constraints. Second, even when taxpayers want to come forward and report previously undeclared income, they may not be able to do so because of high sanctions or lack of information on the legal consequences. In this context our letters may provide a cost effective way to encourage them to revise their account, while benefiting from lower fines that they might not have been aware of.

Our results confirm that in low-income countries, messages can nudge taxpayers into complying more, just as they can in high-income countries. The letters affected small taxpayers' behaviour in particular, whereas large ones did not seem to respond to our treatment. In this pilot we cannot offer more detailed results, for example on revenue gains or on the channel for our effect, because of limitations inherent to the pilot nature of this study – chief amongst these are the need to keep the analysis as simple as possible, and the relatively small sample at hand. Therefore, even if econometric results are reported in this paper, we give more emphasis to the lessons learned on implementing field experiments in Africa.

1 Context and research design

1.1 The Rwandan context

With a population of about 11 million, Rwanda is a relatively small landlocked country within the East African Community. Its small size is not only relevant as an element of context, but also because of its implications for tax collection. With a smaller number of taxpayers than most other countries in Africa and a smaller geographical area to control, the country may provide an easier environment to enforce effective tax collection. This potential advantage is reflected in a tax to GDP ratio of about 15 per cent,[5] which is in line with the average of low-income countries despite the absence of natural resources, and a relatively high reliance on income taxes (about a third of total tax),[6] which are usually seen as harder to collect than trade or indirect taxes. The total number of taxpayers in Rwanda amounted to fewer than 50,000 in 2014. The value added tax (VAT) and Pay As You Earn (PAYE) are the two most

---

[4] We understand that other experiments are being carried out in other African countries, but to the best of our knowledge the results are not yet publicly available.
[5] Source: ATAF 2016.
[6] For a more detailed overview of the Rwandan tax system, see Mascagni et al. 2016a.

important tax types in terms of revenues, accounting for 34 per cent and 25 per cent, respectively, of the tax take (see Mascagni et al. 2016a for more details). Despite being generally a tax success story, Rwanda still faces many of the typical challenges of low-income countries: a large informal sector, widespread evasion and avoidance, and administrative constraints in government institutions. Schneider and Williams (2013) estimate that the informal economy in Rwanda accounted for roughly 40 per cent of national income between 1999 and 2006, while the informal economy labour force as a share of the official labour force was 75 per cent in 1998 (the latest available figures). From a taxpayer's perspective, the tax system still represents a burden both in terms of tax payments and in terms of administrative procedures. The Rwanda Revenue Authority (RRA) has implemented a number of measures to facilitate tax compliance, such as a system for e-filing, a mobile platform for small businesses, electronic billing machines and, more broadly, outreach activities to sensitise and educate taxpayers.

Recognising the importance of voluntary compliance, the RRA applies different sanctions on under-declared incomes depending on whether they are uncovered through an audit or through a voluntary disclosure by the taxpayer. While in the former case taxpayers can face fines of up to 60 per cent of the undeclared income, in the latter case they would only be liable to a 10 per cent late payment sanction. Taxpayers can only benefit from the lower sanction if they revise their tax account before they receive an audit notification. Seven days before the actual audit begins, RRA staff personally deliver the audit notifications and taxpayers need to confirm receipt by signing a return slip. If a taxpayer makes any changes in the tax declaration after an audit has been notified, they are subject to the higher fines.

Although these sanctions are included in the laws, which are available on the RRA website, it is likely that many taxpayers are either not aware of such details or need to be reminded of the benefits of self-rectification. As explained in more detail in the next section, our experiment exploits this feature to test whether providing this information can have a positive effect on compliance, measured through voluntary revisions.

1.2 Research design

Our intervention is in line with the tax experiments literature, as it is a letter aimed at changing the information set and perceptions, rather than the actual parameters of the tax system (e.g. tax rate, probability of detection). We employ an encouragement design, in which the letter contains information on the legal provisions related to sanctions, as set out in the Rwandan tax code and summarised above. As shown in Appendix A, the information was printed on two sides. The front side is the body of the letter, highlighting the importance of correctly reporting taxes and informing taxpayers that the RRA can apply fines for incorrect declarations. To make the message more salient, the letter includes the following example:

> "If your due tax is 2,000,000 Rwandan Francs (RWF) (US$2,444) but you have only declared 1,000,000 RWF (US$1,222), you will have to pay the outstanding due tax of 1,000,000 RWF (US$1,222). In addition, in case of an audit, you will be subject to a fine of 600,000 RWF (US$733) (plus interest). However, if you correct your due tax voluntarily before you are notified of an audit, the fine will only be 100,000 RWF (US$122) (plus interest)."

The amount used in this example, two million Rwandan Francs (RWF) (US$2,444), was chosen to be close to the average corporate income tax due in Rwanda. One potential issue from mentioning a specific amount is that taxpayers may then feel encouraged to revise by exactly that amount. Although we could have randomised the numbers used in the example, we opted against it to keep the letter preparation process as simple as possible during this pilot (see discussion of implementation challenges in Section 2). In addition to the example, the main body of the letter also includes some information about how to contact the RRA in

case the taxpayer wants to receive more details or ask any questions. The back of the letter includes a table where the sanctions are explained in more detail, each case including an example similar to the one reported in the body of the letter and reflecting the different fine rates.

The behaviour that we seek to nudge with this letter is for taxpayers to voluntarily revise their tax accounts. We capture this with two outcome variables – the probability of revising and the amount of revisions – as discussed in more detail in Section 4.1. As noted earlier, this is a different outcome variable than the rest of the literature that usually looks at tax payments or declarations.

Since voluntary revisions can only benefit from lower sanctions if they are made before an audit notification, we had to time our intervention according to the RRA's normal audit schedule. In particular, we wanted to give taxpayers a reasonable window between the receipt of our letter and the beginning of possible audits. As summarised in Figure 1, the RRA normally plans the audits yearly in an audit plan that usually enters into force in July. Although in principle audits can start any time after July, in practice most of them are carried out in the latter part of the audit period, which lasts until the following June. In particular, the first few months right after the adoption of the audit plan are usually dedicated to wrapping up the audits from the previous audit period, so a very small proportion of new audits are carried out then. Therefore, in our original design we planned to use this window to send out our letters, as shown in Figure 1. Although we did not give a time window to taxpayers to respond, as there is none set out in the law, we expected any reaction to occur relatively quickly after receipt of the letter. Therefore, sending the letters shortly after July would have allowed for a five-month window for reactions before the beginning of the tax filing period on 1 January. Although there is no direct link between revisions and declarations, ideally we wanted to avoid any overlap that could possibly confound the results.

Figure 1: Experiment timeline
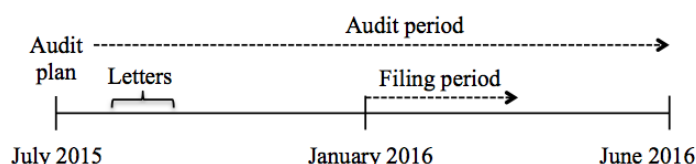
To make sure the message was received as clearly and as easily as possible, each taxpayer received the letter in the three languages spoken in Rwanda: English, French, and Kinyarwanda. Moreover, we tried to keep the text as simple as possible and to include a clear subject line to the letter ("Avoid additional fines by voluntarily correcting your tax declaration before you are notified of an audit"). To make the intervention credible and realistic, each letter was stamped with the official RRA stamp and signed by the RRA's Commissioner General. Since this was an official letter in all respects, the contents were developed in close collaboration with the RRA and approved by the Legal Department to ensure they accurately reflected and respected all relevant laws.

The experiment involved only one treatment, the letter described above and reported in Appendix A, with the counterfactual being a control group that received no letter. Although this design allows us to rigorously evaluate the effect of the letter, we cannot establish whether reactions are due to the content of the letter or to the fact of receiving *any* letter from the revenue authority. To distinguish between the two mechanisms (i.e. specific contents, or any letter) we would need to compare the treatment letter to a control letter, with some general content about RRA activities, in addition to the no-letter control group. However, this

was not possible in our case because of the limited sample available to us, with a maximum of 1,000 letters to be sent and some uncertainty about the success rate of the delivery process. Other studies have found that the very fact of receiving any letter may indeed trigger a reaction by taxpayers (for example, see Del Carpio 2014). So any effect that we detect may occur through either or both of the following two channels: 1) receiving any letter from the revenue authority, which may increase both the perceived probability of being caught and the perception of the effort being put into enforcement; and 2) the content of the letter, which should encourage taxpayers to come forward thanks to lower sanctions. The weakness of our design is that we cannot distinguish between these two channels. However, the reaction that we are trying to identify in our case requires a specific action (i.e. a revision) rather than a general change in behaviour (e.g. more attention being paid in the process of filing the tax declaration). Although this does not fully solve the problem, it makes it somewhat easier to connect taxpayers' reactions to the specific contents of the letter.

1.3 Sample and randomisation

Taxpayers included in our experiment were randomly selected from the total population of taxpayers in Rwanda. We applied three broad criteria to select taxpayers who would be eligible to be part of the experiment. First, we restricted the experiment to taxpayers registered in one of the tax offices of Kigali Province. Still, some of them may operate outside the capital despite being registered there. The reasons for this choice are largely practical. Sending messages outside Kigali was expected to be much more challenging and harder to monitor from the RRA Research Division, our main partner based in the headquarters. Although over half of taxpayers are based outside Kigali, they only contribute 14 per cent to total tax – which supports the hypothesis that it is harder to reach them. Moreover, taxpayers outside of Kigali are much less likely to be audited than those registered in the capital.[7]

Second, we mostly chose taxpayers who we could observe in the baseline year, 2014. The only exception to this general rule is a group of twelve taxpayers from the audited group (see below), for whom we chose 2013 or 2012 as a baseline comparison. The requirement was relaxed for the audited taxpayers, because the low numbers in this group pushed us to retain as many as possible. Since none of these twelve taxpayers made a revision in the period considered, the fact that we use a different baseline year for them does not affect our results.

Third, we only considered taxpayers who pay three main tax types: VAT, corporate income tax (CIT), and personal income tax (PIT). In the context of Rwanda, two clarifications are needed to understand the sample composition. First, PIT refers to the income of individual businesses and the self-employed, rather than employees, who are subject to Pay As You Earn (PAYE). Second, VAT taxpayers here are intended as those who withhold the taxes (i.e. sellers or traders) rather than the downstream buyers or consumers that ultimately pay the tax. By selecting these three tax types, we are considering those taxpayers, both individuals and companies, that need to actively file a declaration to pay taxes and that therefore have a larger margin to under-declare their income.[8]

Based on these eligibility rules, the experiment sample includes 2,000 taxpayers who belonged to the following groups.
1. *1,000 risky taxpayers*, divided in two sub-groups as follows:
    a. 296 audited taxpayers who were audited in 2015-2016, based on the RRA's audit plan. This number was fixed, so we selected all of them.

---

[7] See Mascagni et al. 2016a.
[8] It is well known that those who only receive income from wages have hardly any margin to evade taxes, since they are withheld from their salaries by employers (see for example Kelven et al. 2011).

b. 704 risky taxpayers who were not audited in 2015-16, but are still considered risky based on RRA's criteria for risk management and audit selection (more details below).
2. ***1,000 non-risky taxpayers*** randomly selected from the general population.

The only group that included a fixed, and relatively limited, number of individuals is the one of audited taxpayers. The group includes both taxpayers who are subject to a *comprehensive audit* of all their operations and to an *issue audit* including only a specific tax type or fiscal year.

Risky taxpayers were included based on a list provided by the RRA, complemented with additional taxpayers selected based on RRA's criteria to define riskiness, which fed into the first step of the audit selection process. More specifically, we selected taxpayers based on the following criteria, which we weighted using the same weighting used by the RRA in the audit selection process: size; additional tax obtained from previous audits; difference between turnover as declared for CIT and for VAT; profitability; and frequency of audits in the past.

The 2,000 taxpayers in our sample were then randomly assigned to the control (no letter) and treatment group (who received the letter). Assignment was based on stratification to make sure that the two groups were balanced on key variables that were expected to matter for the results. More specifically, we used four variables for stratification. The first one is a measure of riskiness, indicating whether a taxpayer is audited, risky, or non-risky. The second one is the size of the taxpayer, based on whether the RRA categorises the taxpayer as small or large.[9] This categorisation is likely to include both considerations on size and on enforcement capacity. Thirdly, we used a geographical variable indicating whether the taxpayer is registered in the city of Kigali or elsewhere in Kigali Province. Finally, we considered previous revision behaviour, seeking a balance between those who made revisions in the previous year and those who did not.

Given some challenges in implementation, described in more detail in Section 2, we had to make two small changes in our sample after our original design but before implementation – i.e. before any letter was sent to taxpayers. The first one is the replacement of eight non-risky taxpayers, due to unavailability of contact information in the RRA database. They were replaced with taxpayers from the same strata as the ones who were dropped, thus preserving balance. The second one is a reorganisation of the treatment and control groups for the audited taxpayers. This was necessary because delays in implementation meant that some taxpayers in the audited group had already received the audit notification, therefore invalidating our letter (i.e. those taxpayers would not have a chance to revise at a lower sanction any more). Twenty-eight taxpayers were in this situation and were excluded from the sample. To keep balanced control and treatment groups, we randomly reallocated taxpayers from the same strata to the two groups. The final allocation was still balanced both regarding the key variables and sample size, as confirmed by the statistical tests reported in Appendix C.

---

[9] Throughout this paper, we will refer to taxpayers as 'large' if they are classified in the top-medium or large taxpayer office at the RRA, and as 'small' otherwise.

2 Implementation and lessons learned on tax experiments in Africa

Based on the contents described in Section 1.2, all letters were prepared by RRA staff. The process of letter preparation was in practice more burdensome and time consuming than expected, because until now the RRA has not used letters extensively as a mean to communicate with taxpayers. The preparation process involved the following tasks, which were performed according to RRA standard procedures: letter translation in three languages, legal check of contents, signature of the Commissioner General on each letter, printing on RRA letterhead paper, stamping with the official RRA stamp, and manual filling in of taxpayers' names and addresses on each letter and envelope. Although we contacted 1,000 taxpayers, making letters available in three languages meant having to prepare 3,000 letters. The process was coordinated by the RRA's Research Department, which includes a limited number of staff members who were aided in this task by interns and students hired specifically for this purpose.

All letters were eventually ready in early September 2015 and were transferred to auditors for delivery, which happened within three days between 9 and 11 September.[10] Auditors asked for a confirmation of receipt from the taxpayers, in the form of a signed copy of the letter – one of the three identical copies that each received in three languages (as shown in Appendix A). Delivering letters was a challenging task, mainly due to the fact that the RRA taxpayer registry is not fully up-to-date, so it was impossible to reach all taxpayers successfully just by relying on the addresses available to the RRA. This meant that all taxpayers had to be contacted by phone on the day of delivery to confirm the address and, in some cases, to make sure someone would be available to receive the letter and sign a confirmation. The most common reasons for failed delivery were related to inactive phone numbers, wrong numbers, or phones being switched off. There were also some cases where the taxpayer refused to receive the letter.

The choice of delivering letters through auditors was based both on our desire to follow standard RRA procedures to make our letters fully credible, and on practical considerations. Relying on a private mail company would have posed two challenges. First, it would have risked breaching taxpayer confidentiality since phone numbers would need to be disclosed to a third party. Second, no private company would have the skills, information, and experience that auditors have in tracing individual taxpayers, as they deal with them on a day-to-day basis. This specific expertise was crucial in making sure most letters were delivered successfully. Clearly the drawback of this feature, as anticipated in Section 1.2, is that in this pilot experiment we cannot identify whether the observed effects are due to the letter's contents or to the contact with RRA officials, which makes the probability of detection more salient. In other words, our design does not allow us to distinguish whether revisions are driven by the perceived threat through the delivery of letters by RRA auditors, the information about the level of the fines, or a combination of the two effects. In order to disentangle these effects, we would have required a control message without information on fines. The larger-scale experiment (Mascagni et al. 2016b), for which this study is a pilot, explores these issues in full detail.

Eventually, 589 letters were successfully delivered, representing about 60 per cent of the original treatment group. Compared to similar studies in middle-income countries, this compliance rate is actually relatively high. For example, a recent experiment in Colombia obtained a compliance rate of 38 per cent for similar letters (Ortega and Scartascini 2016). Thanks to the delivery confirmation reports, in the majority of cases we know exactly who

---

[10] In a few rare cases, letter delivery required a follow up the following week.

received the letter and who did not. Not surprisingly, the group of taxpayers who received the letters ('compliers') differs significantly from the group of taxpayers who did not receive the letters (the control group and taxpayers in the treatment group for whom letter delivery was not successful – 'non-compliers'). Although the original allocation into treatment and control groups was random, the actual group of compliers is a selected sample. For example, non-risky taxpayers were less likely to be compliant with the original randomisation (i.e. less likely to have actually received the letters); while taxpayers subject to VAT and CIT were more likely to have received the letters. Since we can accurately identify the compliers, this endogeneity issue can be partly addressed in the regression analysis using instrumental variables (IV) (see Section 3).

On a more general level, the implementation of this pilot experiment provided at least four broad lessons learned. The first one is related to the great constraints in terms of staff capacity that the RRA faces, like most other revenue authorities. For example, our main partner, the research unit, has only five staff members, although they work closely with an additional team of about five in the planning and statistics unit. This pilot was possible thanks to the high degree of commitment and invaluable support from our partners at the RRA. However, the implementation of the experiment did represent a substantial burden both on the research division and on the auditors' teams. In some cases, the revenue gains from experiments like the one reported here would counterbalance this burden. Still, researchers need to be aware that this type of study represents a challenging task on the revenue administration's side, which may not be feasible in countries where there is a lower level of commitment or organisational capacity than in Rwanda. The first step of any field experiment should therefore be a thorough assessment of the level of buy-in and commitment from the local partner at all levels – from the Commissioner General to the officers who are involved in the smallest details of implementation. When these conditions are not in place, researchers should expect a high possibility of failure.

Secondly, many revenue authorities in Africa, including the RRA, do not use letters to communicate with taxpayers as commonly as in high-income countries. Therefore, processes to do so are not efficient and the necessary infrastructure is often not in place – starting with reliable addresses for taxpayers. An additional implication is that we had to rely on auditors for letter delivery, generating a burden for the RRA and potentially confounding our results, for example due to uncertainty on the interaction between taxpayers and auditors.[11] Moreover, this means a potentially high rate of non-compliance with the treatment, and delays in implementation, which can crucially affect the study when timing is fixed due to specific deadlines for taxpaying. In our case, the timing constraint came from the rolling out of the audit plan, which meant that our sample could potentially become smaller with time. In facing these challenges, our pilot study triggered some changes in RRA's internal processes. For example, the procedure to personalise letters, which was manual in this pilot experiment, will be done electronically in the future, including in our main experiment. This study also sensitised the revenue administration about the importance of personalising letters with taxpayers' names, to make sure the message is more salient and effective (in line with the behavioural economics literature, for example see BIT 2012).

Thirdly, one of the biggest challenges we encountered was related to shortcomings in the taxpayers' registry. This is an issue that affects the RRA more widely than just for this research project, as well as affecting many other tax administrations in low-income countries. A process to update the registry is ongoing but is proving very challenging due to the massive work needed initially to clear the backlog, as well as the need for constant updating. In this context, our experiment may support these efforts thanks to the information collected in the process of delivering letters, which can feed back to the registry. The main lesson

---

[11] However, auditors were instructed to only deliver the letter and did not know anything about the content or about the existence of the study.

learned for researchers is to avoid assuming that taxpayers can be reached just because some information is available in the revenue administration's database. In our case, phone contact with taxpayers was almost always necessary to deliver the letters. In practice, this means that researchers may need to carefully select a sample of taxpayers that are more likely to be reached with the information available. For example, restricting the analysis to a specific geographical area, usually the capital city or province, may help. Going forward, our larger-scale experiment limits the sample to those taxpayers who registered recently, to make sure the information on the database is not too outdated. Although this is not ideal from a research design perspective, as it may introduce selection bias, it is a pragmatic response to the specific constraints present in many low-income countries.

Finally, we know from informal conversations with the RRA that several taxpayers reacted to the letter by getting in contact with RRA staff. Although the RRA's call centre was explicitly briefed to collect information from those who enquired about the letter, only a relatively low number of calls were received through this official channel. This is most likely a reflection of broader challenges faced by the call centre, such as few staff members, long waiting times, and incentives to keep calls short to improve performance. As a result, several taxpayers enquired about the letters in other, informal ways. Some of them visited the RRA headquarters in person, while two approached the Office of the Commissioner General directly. From a researcher perspective, these interactions introduce some element of uncertainty as they may affect revisions in unpredictable ways. We tried to limit this issue by keeping the information about the research project strictly confidential to a small group of people within the RRA. Therefore, many RRA officials who might have been contacted by taxpayers did not know that the letters were part of a study – a fact that may have discouraged taxpayers to react. However, these responses confirm that the letters did generate a reaction, even if it did not necessarily result in a revision. The econometric analysis is not suitable to capture these various reactions, which however provide a much more nuanced understanding of the effect of our intervention. For this reason, we commissioned a companion paper from the RRA's research staff to collect such responses,[12] adding to the picture emerging from the econometric analysis on the effect of the letters.

3 Empirical strategy

Based on the design and sample described in Section 1, we estimate the following equation, where *i* indicates the taxpayer:

$$Revisions_i = b_0 + b_1 Treatment_i + b_2(Treatment_i \ x \ Large_i) + \ X_i \ b_3$$

Revisions are captured using the two variables for revisions, the binary one, indicating whether the taxpayer revised, and the amount of revision (see Table B.1 and more details below). The equation specified above includes an interaction term between the treatment and the dummy identifying large taxpayers, in addition to the treatment variable. This choice is based on our initial consultations with the RRA, where we developed the hypothesis that there may be fundamental differences in the reactions by small and large taxpayers. While the latter are already well informed about tax laws, the former may not be. Moreover, it is more likely that a small taxpayer would react to an RRA communication out of fear of being caught, while this would not be the case for large taxpayers who are more knowledgeable about their tax affairs. Finally, we include a set of controls *X*, namely: whether a taxpayer is classified as large; the degree of riskiness (0 = non-risky, 1 = risky, 2 = risky and audited); the lagged dependent variable (referring to the same five-month period a year earlier); tax

---

payable at the baseline; and the amount of under-reported tax due if the taxpayer was subject to an audit.

Based on this specification, the equation is estimated using three empirical models. An Ordinary Least Squares (OLS) model is used when the dependent variable is equal to the level of revision. By using the amount in level, the resulting coefficient on the treatment variables (treatment and the interaction with large) can be directly interpreted as additional tax resulting from the revisions by the treated. A probit model is estimated when revisions are measured using a binary variable that is equal to one if the taxpayer revised the declaration at least once positively, so that the new tax due is higher than the originally declared tax. The dependent variable takes the value zero if the taxpayers did not revise or revised negatively. In our third model, we define the dependent variable as the logarithm of the level of revision, because it is less affected by outliers than the level variable. As the logarithm is only defined for strictly positive values, this estimation does not consider negative revisions.[13] Since the majority of taxpayers do not revise their declarations, we add 1 RWF to the level of revision. Thus, the dependent variable takes the value ln(1) if the level of revision is zero. Since a large number of taxpayers do not revise their declarations, the dependent variable revisions are also censored at zero and consequently, we use tobit estimation for the logarithmic specification. For the probit and tobit models, we report both the latent variable coefficients and marginal effects.

We present both an intention-to-treat (ITT) analysis and estimation of the local average treatment effect (LATE). The former analysis is based on the original treatment assignment and does not take into account whether a taxpayer actually received our letter or not. On the other hand, the LATE estimation considers that some taxpayers in the treatment group have not received the letter by using an IV technique. In particular, it uses the original treatment assignment as an instrumental variable for the actual treatment and thus provides an estimate of the impact of the treatment on compliers (see Bloom 2008; Angrist and Pischke 2009). Intuitively, the LATE estimation reflects the potential impact of the intervention in case the RRA is able to fully implement the experimental design (i.e. successfully deliver letters to all selected taxpayers), while the ITT analysis shows the effectiveness of the treatment given the current circumstances. As the letter receipt cannot be mandated, the ITT estimation may be more policy relevant than the LATE analysis (Bloom 2008). Based on these considerations, we focus on the ITT results, while we briefly discuss LATE in the text and report results in the appendices.

4 Data and description of revisions

The analysis is based on administrative data from anonymised taxpayer records obtained from the RRA. On the one hand, our dataset presents some advantages over other sources of data, which are described in more detail in Mascagni et al. 2016a. Perhaps the most important one is that it allows us to capture 'real-life' taxpaying behaviour, rather than attitudes. This feature is enhanced by the fact that all communications with taxpayers, most notably the letter, were coming from the RRA rather than from researchers. Moreover, taxpayers' behaviour has real monetary consequences. On the other hand, taxpayer records do not include all the variables that we would ideally include in the analysis. For example, we do not have information on the gender of business owners, or on whether they are foreign nationals or locals. We also cannot establish with precision where the company is operating in practice, although we know where it is registered for tax purposes. Notwithstanding these

---

[13] Our estimation results are very similar if we define the dependent variable in such a way that it takes the value zero in case of negative revisions (compare results in Table 2 and Table D.3 to Table E.1).

drawbacks, our data, coupled with experimental methods, allow us to explore quasi-voluntary tax compliance in a more rigorous way than previously possible in Africa.

Our dataset includes the variables available in tax declarations, such as various definitions of income (i.e. turnover, gross profit, taxable income), whether the taxpayer is small or large (according to the registration at the taxpayer office), and the location of the tax centre where the taxpayer is registered. A description of all the variables used in this paper is reported in Appendix B. Crucially, our data allows us to track taxpayers in time through unique identifiers that still preserve the confidentiality and privacy of taxpayers – i.e. researchers who are external to the RRA cannot trace any specific observation in the dataset to a specific individual or company. Our dataset spans five years, from 2011 to 2015, with 2014 used as a baseline for randomisation and 2015 used to capture revisions.

4.1 Revisions

While other variables follow fairly standard and straightforward definitions, it is important to report here more details on our key outcome variable: revisions. First of all, we only consider revisions that happened in the five-month period after our experimental intervention, from 8 September 2015 to 7 February 2016. The cut-off date was decided based on three considerations. First, given the delays in implementing the intervention (see Section 2), we had to shift our window for revisions by one month to leave enough time for taxpayers to respond. Second, there is no natural cut-off date that is common to all the tax types considered here, as they follow different timelines. While PIT and CIT declarations are filed once a year between 1 January and 31 March, VAT declarations are filed monthly or quarterly depending on firm size. Still, most PIT and CIT declarations are usually filed towards the end of the filing period, therefore minimising the overlap with our five-month window. Third, we wanted to close the window before starting to implement our next and larger-scale experiment, which also involved sending messages to taxpayers (see Introduction). We are confident that our choice is sensible based on these considerations and on the observation that most responses happen right after the intervention. However, we also check for the robustness of our results by using a four-month window as an alternative.

Based on these considerations, we constructed two variables related to revisions. The first one captures the amount of revisions, calculated as the difference between the tax due from the original declaration and any revisions made by the taxpayer in the five-month period. To avoid simply capturing mistakes in the process of filing, we do not consider revisions that happen on the same day as the declaration. In other words, when there are multiple entries on the day of the declaration, we take the tax due from the latest revision on that day as the amount of the original declaration. Subsequent revisions are always measured compared to the original declaration, so that our variable captures the cumulative amount of all revisions that the taxpayer may have made. Multiple revisions within the same *year* are not common for PIT and CIT, but they are more so for VAT where declarations are monthly or quarterly, and so potentially revisions are, too. Finally, we only use revisions that occur at least seven days before an audit, which is the time of the audit notification, after which lower sanctions are no longer applicable. By doing this, we can isolate the effect of our letter from taxpayers' responses to the audit notifications. Our conversations with the RRA staff revealed that taxpayers often respond to the audit notification by revising their accounts, hoping to 'limit the damage' from the audit. However, as described in Section 1.1, any action taken by the taxpayer after the receipt of the audit notification is not valid for the purpose of determining sanctions. The second outcome is a binary variable taking the value of one when a taxpayer has revised at least once in the five-month period and zero otherwise. We disaggregate this binary variable further to separate positive and negative revisions.

Although revisions can be indicative of quasi-voluntary compliance, there are other possible motivations. For example, taxpayers may make changes as a result of new information

becoming available, because they are trying to decrease their tax burden ex-post, or to correct previous mistakes. As a result, revisions can be either positive or negative. Table 1 reports the percentage of taxpayers who made at least one positive revision in the calendar year 2014, the baseline year, and splits that number into those whose revised tax due was higher and lower than the original declaration. Consultations with the RRA revealed the perception that negative revisions are often related to avoidance or evasion. For example, we heard anecdotal evidence of taxpayers negatively revising their CIT account to compensate for unexpected, or unexpectedly high, VAT payments or for failed refunds.

Table 1 also reveals that most revisions concern VAT, which is true both in percentage and absolute terms (for the latter, see table D.1 in Appendix D).[14] There are at least three motivations to explain why more revisions occur for VAT. First, enforcement efforts have been particularly focussed on VAT in recent years, for example with the introduction of electronic billing machines. This may have increased taxpayers' perceptions of the probability of detection for VAT specifically, therefore pushing them to come forward and reveal those under-declared incomes that they feel are more easily uncovered by audits. Second, the fact that VAT is filed monthly may make it easier for taxpayers to revise it, as the relevant information is more recent and clearer in their memories. In contrast, at the time of the intervention, CIT and PIT declarations were filed five to eight months earlier. Third, taxpayers have only fifteen days to prepare their monthly VAT return, as opposed to three months for PIT and CIT. Since they have less time to prepare and review VAT returns, taxpayers may end up revising them more often to correct mistakes or make amendments.

Table 1 Revisions in the year 2014 (before our intervention)

Tax type	% revised once	% positive	% negative
CIT	3.9	2.6	1.3
PIT	9.9	8.2	1.6
VAT	29.9	19.3	10.6
Total	22.7	15.3	7.3

Note: Authors' calculations based on RRA dataset for the calendar year 2014.

5 Econometric results

We start the analysis by testing whether there are statistically significant differences in revisions between the treatment and control groups, after the intervention. For both our outcome variables, we test whether the control and treatment groups display a statistically different behaviour in terms of revisions. The results of these initial tests are reported in Appendix D. As far as the binary variable is concerned, Fisher's tests reported in Table D.1 (Appendix D) confirm that negative revisions are more common in the control group, while positive revisions are more frequent in the treatment group. Therefore it seems that letters may have a twofold effect: they increase the number of positive revisions (potentially as a result of previous under-reporting) and they make negative revisions less likely (a potential deterrent effect of the treatment message). These differences are statistically significant amongst revisers and borderline significant amongst all taxpayers. However, the number of revisions in our sample is quite small.

---

[14] Note that PIT and CIT taxpayers should not necessarily be registered for VAT (roughly 26 per cent of PIT and 80 per cent of CIT filers in our sample declare VAT), so the number of observations is lower when considering only VAT rather than all tax types.

Looking at the amount of revisions in the two groups, the tests reported in Table D.2 (Appendix D) confirm that the amount of revisions is negative in the control group, but positive in the treatment group. In the five-month period considered, only a small share of taxpayers revised their declaration at least once: 16 in the control group and 17 in the treatment group. Although the average revised tax due amongst all taxpayers is relatively low, due to a large number with zero revisions, the average among *revisers* is substantial: an average negative amount of RWF 27.9 million (US$34,091) in the control group and a positive RWF 1.3 million (US$1,559) in the treatment group.[15] These differences are statistically significant when the distribution is considered, but only borderline significant at the 10 per cent level when we test for equal means. However, given the number of taxpayers in our treatment and control groups, the minimum detectable standardised difference between the treatment and control for a one-sided *t*-test is 2.5 to achieve a power of 80 per cent under the five per cent significance level (which translates to a mean difference of more than RWF 730,000 (US$892)).[16] In other words, the relatively lower level of significance on the equal means test could also be a consequence of the low power of the *t*-test.

Table 2 shows the intention-to-treat (ITT) analysis, which is based on the original randomisation – regardless of whether the treatment was actually received. The ITT analysis identifies the causal effect of the offer of the letter, considering that some taxpayers rejected the treatment or could not be reached (see Angrist and Pischke 2009). An OLS model is estimated in column 1 of Table 2, where the dependent variable is equal to the amount of revision in levels. A probit model is estimated in column 2, where the dependent variable is equal to one if the taxpayer revised the declaration at least once positively, so that the new tax due is higher than the originally declared tax. The dependent variable takes the value zero if the taxpayer did not revise or only revised negatively.[17] Column 3 estimates a tobit model, where the dependent variable is the logarithm of the amount of revision, as discussed in Section 3. All regressions include a constant and control for size, riskiness, and the tax due for the previous fiscal year, as well as lagged revisions and principals, as discussed in Section 3 (also see Appendix B for a description of the variables). The coefficient estimates on these control variables are omitted to improve readability. We estimate the treatment effects both considering all tax types (Panel A) and only VAT (Panel B), where most revisions occur.

The treatment effect based on the OLS regression (column 1 of Table 2) is not statistically significant for taxpayers of any size. As discussed in the previous section, this may be a result of low statistical power, given the low level but high standard deviation of revisions, and the relatively small sample size. In other words, we do not know whether our estimated treatment effect is truly non-significant or whether the lack of significance is caused by a lack of power.[18] Nevertheless, the OLS coefficients would be interpreted as follows: the average revision level for small taxpayers in the treatment group is estimated to be around RWF 36,000 (US$44) higher than in the control group (see coefficient in the first row). For large

---

[15] While the average revision of revisers is close to the RWF 1 million (US$1,222) mentioned in the example of the letter, no taxpayer actually revised their declaration by this amount (the closest revision to RWF 1 million was RWF 1.75 million (US$2,138). In contrast, checking the distribution of revisions, there does not seem to be any form of bunching of revisions.

[16] To calculate the minimum detectable effect size under 80 per cent power, we used the following parameters: the number of taxpayers in our treatment and control groups is 995 and 982, respectively. The mean of the control group is –471,799 in the same five-month period in the previous year (see Table C.1), while the corresponding standard deviations are 8,307,536 and 3,969,788 in the control and treatment groups, respectively.

[17] If we redefine the dependent variable to take the value one if the *aggregate* revisions of the taxpayer in the five-month period are positive and zero otherwise, the results are almost identical and the conclusions remain the same. Results available from the authors upon request.

[18] We do not have geographical information which we could exploit for the estimation of cluster robust standard errors. If we cluster the standard errors on the tax centre (using a *T*-distribution with *G–L* degrees of freedom, where G is the number of clusters and *L* the number of coefficients), we find a weakly significant treatment effect on large taxpayers. However, the relatively low number of tax centres in our sample (10) could lead to severely downward biased cluster robust standard errors (e.g. Bell and McCaffrey 2002 and Cameron and Miller 2015). Therefore, we generally use White-heteroscedasticity corrected standard errors in our regressions. However, the main conclusion regarding the probit and tobit model shown in Table 2 remains the same if we use cluster robust standard errors.

taxpayers the difference between the control and treatment groups would be almost RWF 2.7 million (US$3,299) (i.e. the sum of coefficients in the first and second row). These differences between the revised tax amounts in the treatment and control groups, if significant, could be used to calculate the gross revenue gain resulting from the experiment, which would be RWF 379 million (or about US$464,000) in our case.[19] However, the fact that the coefficients are largely non-significant calls for much caution in interpreting these estimates and prevents us from taking them as anything but indicative.

Table 2: Regression results (ITT)

	(1) OLS	(2) Probit	(3) Tobit
Panel A: All tax types			
Treatment	36,044 (31,010)	0.65** (0.26)	27.0** (10.8)
Treatment x large	2,619,738 (2,234,682)	−0.72 (0.48)	−29.8* (17.9)
Marginal treatment effects, small taxpayers:			
Predicted probability		0.007**	
Effect of censored revisions			0.07***
Effect conditional on positive revision			2.01***
Marginal treatment effects, large taxpayers:			
Predicted probability		−0.001	
Effect of censored revisions			−0.01
Effect conditional on positive revision			−0.22
Other controls	Yes	Yes	Yes
Observations	1,977	1,977	1,939
Panel B: VAT			
Treatment	74,668 (56,356)	0.63** (0.27)	26.0** (11.1)
Treatment x large	2,598,608 (2,377,545)	−0.69 (0.48)	−28.5 (17.8)
Marginal treatment effects, small taxpayers:			
Predicted probability		0.013**	
Effect of censored revisions			0.15***
Effect conditional on positive revision			2.29***
Marginal treatment effects, large taxpayers:			
Predicted probability		−0.001	
Effect of censored revisions			−0.01
Effect conditional on positive revision			−0.22
Other controls	Yes	Yes	Yes
Observations	1,196	1,196	1,162

Notes: Robust standard errors are in parentheses. *** $p<0.001$, ** $p<0.05$, * $p<0.1$. Marginal effects are evaluated at the mean. The dependent variable is the level of revision in column (1), the binary revisions variable in column (2), and the logarithm of the level of revision in column (3). All regressions include a constant and controls for size, riskiness, the latest available tax due, as well as revision and principal of the previous fiscal year.

The probit and tobit estimations show that small firms are more likely to revise positively when they receive the letter (column 2) and, when they do, the amount revised is higher (column 3). We can only detect a significant treatment effect for small taxpayers, while the letters did not have an influence on large ones. One potential explanation is that large taxpayers already know about the benefits of revisions before receiving the letter, as they are

---

[19] The revenue gain for small taxpayers is calculated by multiplying the coefficient on the treatment variable by the number of small taxpayers in the treatment group: 36,044 x 864. Similarly, for large taxpayers the revenue gain would be the sum of the two coefficients multiplied by the number of large taxpayers: (36,044 + 2,619,738) x 131. The total revenue gain is the sum of the two.

more likely to have better knowledge of tax law or access to good tax advisers. In contrast, small taxpayers may have previously not been aware of the financial benefits (i.e. lower penalties) of revising wrong declarations.[20] Therefore, the differences between the control and treatment groups are largely driven by the reactions of smaller taxpayers.

In particular, those taxpayers in the treatment group have more than a one percentage point higher probability of making a positive revision of their original tax due than taxpayers in the control group (see marginal treatment effects in column 2). If a small taxpayer is in the control group, the predicted probability of making a positive revision is only around 0.3 per cent, based on evaluating the marginal effects for the treatment and control groups separately. In comparison, the same probability of small taxpayers in the treatment group is around 1.6 per cent. Therefore, small taxpayers in the treatment group are more than five times as likely to revise than small taxpayers in the control group.

When small taxpayers revise their declarations, the revisions in the treatment group are more than 200 per cent higher than the revisions in the control group (last row of marginal treatment effects in column 3). Therefore, these effects are relatively sizeable, although positive revisions remain a relatively uncommon action amongst Rwandan taxpayers.

To complement the ITT analysis, we estimate the local average treatment effect (LATE) using the original treatment assignment as an instrument for actual treatment. By doing this, LATE estimates the effect of the treatment on the treated, or compliers, taking into account that some taxpayers in the treatment group have not received the letter. It represents the causal effect of the treatment on the compliers (see Angrist and Pischke 2009, p.136). The LATE estimations in Table D.3 (Appendix D) mirror the results of the ITT analysis, while both the coefficients and the standard errors increase in size. First, as expected, the effect of the intervention on revisions is higher for those taxpayers in the treatment group that have certainly received the letters. Comparing the OLS estimates (Table 2) and the corresponding LATE results (Table D.3) confirms that the treatment-effect size almost doubles: from RWF 36,000 (US$44) to RWF 63,000 (US$77) for small taxpayers and from RWF 2.7 million (US$3,299) to more than RWF 4 million (US$4,888) for large taxpayers. Second, the increase in the standard errors is a result of the IV regressions, which also depends on the fit of the first stage estimation. Still, IV techniques are necessary to estimate the LATE, since standard OLS, probit, and tobit models would lead to inconsistent estimates.

The ITT and LATE analyses are therefore largely consistent, both in terms of sign and statistical significance for small taxpayers, thus also supporting robustness. These results, taken together, suggest that information letters could be an effective way to encourage taxpayers to comply by making revisions more likely and leading to the declaration of previously unreported income. Moreover, conditional on making a revision, the amount of tax revised is higher once taxpayers receive the information letter. While these findings seem to suggest that communication strategies of this type can be effective, particularly for small taxpayers, we cannot determine if this effect is due to the information on sanctions included in the letter or to the fact of receiving a letter. The latter is likely to have increased taxpayers' perception of the probability of detection, which is one of the key determinants of compliance, especially because letters were delivered by auditors and in a context in which this type of communication from the RRA is relatively uncommon. This is particularly true for small taxpayers, who are also very unlikely to be audited and generally have less contact with the RRA (Mascagni et al. 2016a). Such an unusual event is likely to have affected perceptions about enforcement efforts and generally led taxpayers to realise that they are on the RRA's radar. This may have pushed taxpayers to revise their accounts, potentially along with the

---

[20] If we compare revisions of small taxpayers to the revisions in the same period one year before, the number of negative revisions is unchanged (10). However, the number of positive revisions increased from 8 to 11 (which is an almost 40 per cent increase). In comparison, large taxpayers tend tended to revise generally less often than in the year before. Especially, the negative revisions decreased from 13 to 6. (The positive revisions decreased from 9 to 6.)

encouragement provided by information on lower fines for voluntary revisions. The key point, as recognised elsewhere in this paper, is that we cannot distinguish the channel of our observed effect.

5.1 Robustness and caveats

We have checked the robustness of our results to the choice of the five-month time window, since this decision is somewhat arbitrary. We have run again the regression analysis using a four-month window, from 8 September 2015 until 7 January 2016. The main advantage of this alternative window is that it minimises the overlap with the CIT and PIT tax filing period, although this overlap is not expected to matter much since most revisions occur for VAT. The results are reported in Appendix D. Using this window, the results are generally similar, but less robust due to the slightly lower number of revisions in this shorter period. For example, the VAT declaration for December 2015 is only due on 15 January 2016 and therefore any revisions to that declaration can only happen after that date. Notwithstanding these considerations, we are confident that our results are not determined by the choice of the five-month time window.

Although we are confident in our results, three caveats in particular should be noted. The first one is related to our outcome variable, which requires a relatively strong change in behaviour on the part of taxpayers: to voluntarily come forward and disclose previously un-declared income. Although our letter seems to nudge this type of behaviour, only a few taxpayers reacted. Secondly, although we made every effort to make the letter as simple and salient as possible, the explanation of sanctions is necessarily quite technical and complex. Therefore, some taxpayers may have failed to fully understand the contents and, as a result, may have ignored it. Third, as already mentioned throughout the paper, we cannot distinguish the effect of the letter's contents from the very fact of receiving a letter, which in itself may have increased the perceived probability of an audit.

These caveats, together with the fact that taxpayers are not used to receiving letters from the RRA, suggest that the intervention may have created some confusion amongst taxpayers. A companion paper,[21] prepared by the RRA, collected qualitative feedback from taxpayers on their perceptions and their reactions to the letter. The anecdotes collected there confirm that some taxpayers were not sure how to react to the letter, although they did have some unreported income to declare. Therefore, we also checked whether they may have changed their declarations behaviour, in addition to making the revisions. In other words, it may be that a taxpayer does not feel comfortable disclosing previous evasion, but she may be more compliant in the future. Although we do not find any evidence of this behaviour in this pilot experiment, we further explore nudges on declarations in our larger experiment (Mascagni et al. 2016b).

Despite the caveats described above, the analysis still yielded results that are both significant and in line with our knowledge of the Rwandan taxpaying environment. Furthermore, the qualitative taxpayer feedback confirmed that the letters indeed provoked a reaction, even if that did not translate into a revision. Since this experiment is a pilot, the caveat observed here, as well as the lessons learned summarised in the next section, fed into our larger-scale experiment.

---

[21] This paper will be available as an ICTD Working Paper.

6 Conclusions

This paper reports key findings from a pilot field experiment on tax compliance, which to the best of our knowledge is the first publicly available evidence of this type from Africa. As such, a key objective of the paper is to provide lessons learned on implementing this type of study in low-income countries, as well as reporting initial econometric results. As far as lessons learned are concerned, Section 2 summarises some characteristic features of low-income countries that researchers need to factor into their research plans and design. These include administrative constraints; the effectiveness of physical letters, for which processes are not streamlined; and shortcomings in the taxpayer registry. Although the success rate of our letter delivery exercise compared relatively well with other similar studies, sending a large number of letters is a burden for the revenue administration, both in terms of letter preparation and delivery. This makes it particularly important for low-income countries to experiment with other delivery methods. For example, our larger field experiment includes emails and SMS, along with physical letters (Mascagni et al. 2016). Despite these challenges, this pilot confirms that large-scale field experiments are indeed feasible in Africa, especially when the local revenue authority is committed to the project.

In terms of econometric results, the analysis reported in Section 5 suggests that information letters can be an effective way to nudge taxpayers into complying more. We used revisions as an indicator of voluntary compliance, which represents a departure from the literature. Our analysis shows that letters work by both reducing negative revisions, which seem to be a way to avoid or evade taxes, and by increasing the likelihood of positive revisions. However, we cannot determine if this effect occurs through an increase of the perceived probability of detection (the 'stick') or through encouragement provided by lower fines (the 'carrot'). This highlights the importance of including a control letter, in addition to a no-letter control group. Although our key results are generally significant, they rely on a relatively small number of revisions in the period considered. Therefore, they should be taken with some caution, especially when attempting to translate them into precise policy recommendations.

Nonetheless, we can highlight some broad implications for policymakers. On a general level, the results suggest that communication strategies can be an effective way to improve voluntary compliance. This is particularly the case for small taxpayers, who probably have less information about the tax system and less access to good tax advisers. Importantly, such communication strategies are in line with the RRA's vision to go beyond audits in its efforts to increase compliance and tax revenues. However, our results are not sufficient to recommend scaling up the precise intervention tested here to a larger group of taxpayers. Revisions are a very specific aspect of non-compliance, although our evidence suggests that they may indeed be used as a way to avoid or evade taxes. Still, nudging this specific behaviour may not be the most effective strategy to increase compliance. The number of revisions we observe in the period after the letter is still low. Moreover, it is difficult to assess if our intervention would be worthwhile on a larger scale, since the lack of statistical significance in the OLS regression makes it hard to obtain reliable estimates of the total revenue gain. Nudges influencing declaration behaviour and tax payments are likely to be a more effective and more profitable strategy for revenue authorities.

Appendices

Appendix A Experimental letter

Figure A.1 Body of the letter

RWANDA REVENUE AUTHORITY

OFFICE RWANDAIS DES RECETTES

TAXES FOR GROWTH AND DEVELOPMENT

Our Ref : ████████████████

Your Ref : ....

Dear; ████████████████████████

RE: Avoid Additional Fines by Voluntarily Correcting Your Tax Declaration Before You Are Notified of an Audit

This letter is to inform you about the importance of correctly reporting your taxes.

If your tax declaration is not fully correct, **you are subject to a 10% fine for late payment and additional fines for understating your income**. The higher the understatement of the income, the higher the fines you will have to pay.

However, **if you voluntarily correct your tax declaration before you are notified of an audit, the RRA will only apply the 10% fine for late payment** on the due tax.

For example, if your due tax is 2,000,000 RWF but you have only declared 1,000,000 RWF, you will have to pay the outstanding due tax of 1,000,000 RWF. In addition, in case of an audit, you will be subject to a fine of 600,000 RWF (plus interest). However, **if you correct your due tax *voluntarily before* you are notified of an audit, the fine will only be 100,000 RWF** (plus interest).

For more information about the fines, please refer to the back of this letter. In case you have any questions, please contact the Call Centre of the RRA on 3004 and cite the reference number above (opening hours: 7am to 5pm, Monday to Friday).

We would like to thank you for your collaboration.

Kind Regards

TUSABE Richard
Commissioner General

Kimihurura, Avenue du lac Muhazi,
P.O.Box 3987 Kigali, Rwanda
HOTLINES 3004/3005

Telephone : +(250) 595 501 , 595712, 595541, 595600, 595665
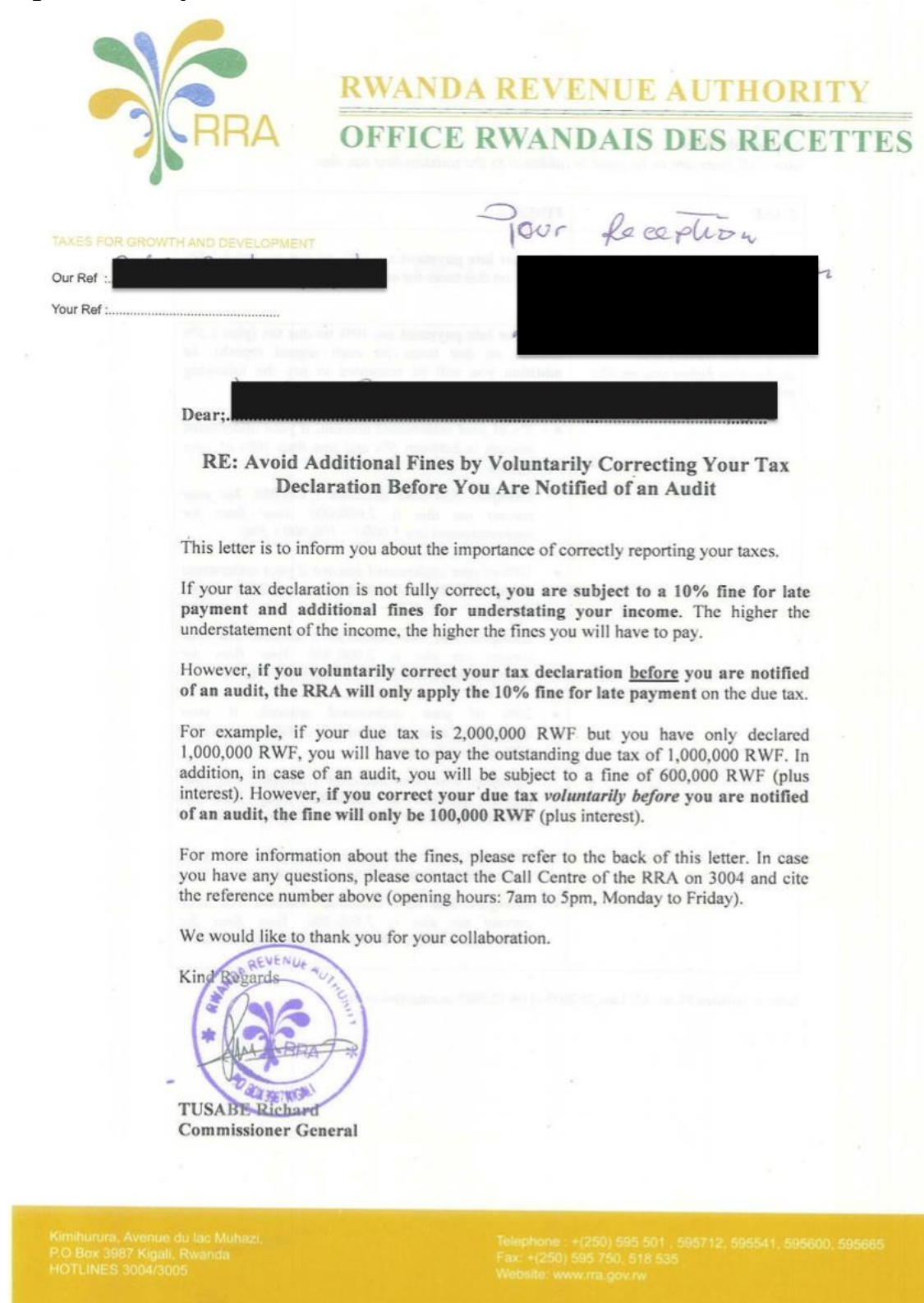Fax: +(250) 595 750, 518 535
Website: www.rra.gov.rw

Figure A.2 Back of the letter

Applicable fines

Note: All fines are to be paid in addition to the outstanding tax due.

CASE	FINES
You *correct your declaration before* you receive an audit notification	**Fines for late payment** are 10% on due tax (plus 1.5% interest on due taxes for each unpaid month)
You do *not correct your declaration before* you receive an audit notification	**Fines for late payment** are 10% on due tax (plus 1.5% interest on due taxes for each unpaid month). **In addition** you will be requested to pay the following **fines for understatement:** • 5% of your understated amount, if your understated amount is between 5% and less than 10% of your due tax. *Example: You have declared 1,900,000, but your correct tax due is 2,000,000. Your fines for understatement are 5 000 (= 100,000 x 5%).* • 10% of your understated amount if your understated amount is between 10% and less than 20% of your due tax. *Example: You have declared 1,700,000, but your correct tax due is 2,000,000. Your fines for understatement are 30,000 (= 300,000 x 10%).* • 20% of your understated amount, if your understated amount is between 20% and less than 50% of your due tax. *Example: You have declared 1,300,000, but your correct tax due is 2,000,000. Your fines for understatement are 140 000 (= 700,000 x 20%).* • 50% of your understated amount, if your understated amount is 50% or more of your due tax. *Example: You have declared 500,000, but your correct tax due is 2,000,000. Your fines for understatement are 750 000 (= 1,500,000 x 50%).*

Refer to Articles 59, 61, 62, Law 25/2005 of 04/12/2005 as amended to date.
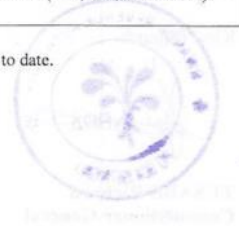
Appendix B Description of variables

Table B.1 Summary of variables used in the regression analysis

Variable	Description
Tax due *(in RWF)*	Total tax liability for CIT, PIT, and VAT for the latest available fiscal year (in RWF)
Revision *(amount)*	Difference between the tax due from the original declaration and any revisions made by the taxpayer in our five-month investigation period (in RWF, before receiving an audit notification)
Revision *(binary)*	Binary variable that takes the value one if the taxpayer positively revised the declaration at least once in our five-month investigation period (before receiving an audit notification)
Treatment *(binary)*	Binary variable indicating whether the taxpayer was in the treatment group or not
Large *(binary)*	Binary variable that takes the value one if the taxpayer is registered at RRA as a large or top-medium firm, and zero otherwise
Riskiness *(categorical)*	Categorical variable that takes the value zero for non-risky firms, one for non-audited but risky firms, and two for audited firms
Principal *(in RWF)*	The aggregated under-reported tax due for CIT, PIT, and VAT, which was discovered through audits in the five-month period one year before our experiment

Appendix C Balance tests of treatment and control groups

Table C.1 Balance tests of treatment and control groups at baseline

(1) Variable	(2) Control *n*=982	(3) Treatment *n*=995	(4) Equal distribution *p*-value
Large	132	131	
Small	850	864	
			0.895
Non-risky	722	505	
Risky (no audit)	349	354	
Risky (audit)	133	136	
			0.998
CIT	563	568	0.928
PIT	332	338	0.962
VAT	629	622	0.484
Real regime	508	506	
Lump sum	261	279	
Flat amount	23	24	
			0.803
Lagged negative revisions	11	10	
Lagged zero revisions	964	978	
Lagged positive revisions	7	7	
			0.964
In RWF:			
Latest tax due	71,453,768	35,121,734	0.945
Lagged principal	121,071	−18,264	0.774
Lagged revisions	−471,799	76,253	0.857

Notes: n indicates the number of observations. For categorical variables, column (4) displays the *p*-value of the two-sided Fisher's exact test. The null hypothesis is that there is no association between the treatment and the category of taxpayers indicated in column (1). For continuous variables, column (4) shows the *p*-value of the Mann-Whitney test. The null hypothesis is that the distribution of the continuous variable is the same in the control and treatment groups.

Appendix D Initial tests and LATE results

Table D.1 Differences in the number of revisers

	(1)	(2)	(3)	(4)
	Control	Treatment	Fisher's test *p*-value (all taxpayers)	Fisher's test *p*-value (revisers)
Panel A: all tax types				
Negative revisions	10	6		
No revisions	968	978		
Positive revisions	4	11	0.122	0.049
Panel B: VAT				
Negative revisions	9	5		
No revisions	616	606		
Positive revisions	4	11	0.108	0.048

Notes: Column (3) displays the *p*-value of the one-sided Fisher's exact test. The null hypothesis is that there is no association between the treatment and the sign of revision. Column (3) considers the frequency of revisers (negative and positive) as well as non-revisers. Column (4) only considers the frequency of revisers (negative and positive). Fisher's test is used to test the null hypothesis that there is no association between the treatment and the sign of revision. Fisher's test is preferable to its parametric equivalent to compare frequencies, the chi-square test, if there are small cell sizes, as in our case. It does not rely on assumptions about the underlying distribution of revisions, unlike the chi-square test.

Table D.2 Differences in the amount of revised tax

	(1)	(2)	(3)	(4)
	Control	Treatment	Equal means *p*-value	Equal distribution *p*-value
Panel A: all tax types				
All taxpayers	−397,787 (337,919) *n*=982	22,358 (24,940) *n*=995	0.106	0.048
Revisers	−27,901,921 (23,353,409) *n*=14	1,308,624 (1,468,516) *n*=17	0.089	0.029
Panel B: VAT				
All VAT payers	−591,579 (526,733) *n*=629	41,604 (39,459) *n*=622	0.117	0.040
Revisers	−28,623,337 (25,212,526) *n*=13	1,617,363 (1,528,414) *n*=16	0.097	0.028

Notes: Standard errors in parentheses. n indicates the number of observations in the control and treatment groups. The null hypothesis of the one-sided *t*-test shown in column (3) is that the level of revisions is the same in the control and treatment groups
. The null hypothesis of the Mann-Whitney test shown in column (4) is that the distribution of revisions is the same. The Mann-Whitney test is a non-parametric correspondent to the independent samples *t*-test. It does not require assumptions about normality, but only that the dependent variable, revisions, is ordinal.

Table D.3 LATE analysis

	(1) IV	(2) IV-probit	(3) IV-tobit
Panel A: All tax types			
Treatment	62,086	1.09**	45.6*
	(53,317)	(0.51)	(23.8)
Treatment x large	3,983,033	−1.24	−60.0
	(3,405,672)	(0.81)	(32.5)
Other controls	Yes	Yes	Yes
Observations	1,977	1,977	1,939
Panel B: VAT			
Treatment	116,837	0.98**	40.2*
	(87,587)	(0.48)	(21.9)
Treatment x large	3,904,613	−1.11	−45.2
	(3,573,086)	(0.79)	(30.8)
Other controls	Yes	Yes	Yes
Observations	1,196	1,196	1,162

Notes: Robust standard errors are in parentheses. ** p<0.05, * p<0.1. An IV regression is estimated in column (1), where the dependent variable is equal to the level of revision. A two-step IV-probit model is estimated in column (2), where the dependent variable is equal to one if the taxpayer positively revised the declaration at least once. Column (3) estimates a two-step IV-tobit model, where the dependent variable is the logarithm of the level of revision. All regressions include a constant and controls for size, riskiness, and the latest available tax due, as well as revision and principal of the previous fiscal year. Note that columns (2) and (3) do not include marginal effects because we used two-step probit and tobit estimators, as the alternative maximum likelihood estimators had difficulties in converging due to the presence of multiple endogenous variables (treatment and treatment x large). This estimator is not directly comparable to the standard probit and tobit based on maximum likelihood, but is rather used to test for statistical significance. We could obtain marginal effects comparable to Table 2 if we used maximum likelihood and ran the regressions IV-probit/tobit only for small taxpayers.

Appendix E Robustness tests

Table E.1 Tobit analysis without missing values (dependent value is zero if negative revisions)

	(1) Tobit	(2) IV-tobit
Panel A: All tax types		
Treatment	19.3**	32.4*
	(8.8)	(18.4)
Treatment x large	−21.9	−37.5
	(16.5)	(28.5)
Other controls	Yes	Yes
Observations	1,977	1,977
Panel B: VAT		
Treatment	17.9*	28.0*
	(9.2)	(17.0)
Treatment x large	−20.2	−32.5
	(16.5)	(27.3)
Other controls	Yes	Yes
Observations	1,196	1,196

Notes: Robust standard errors are in parentheses. ** p<0.05, * p<0.1. A tobit model is estimated in column (1), where the dependent variable is the logarithm of the level of revision. Column (2) estimates a two-step IV-tobit model, where the dependent variable is the logarithm of the level of revision. In case the level of revisions is negative, the dependent variable takes the value zero. All regressions include a constant and controls for size, riskiness, and the latest available tax due, as well as revision and principal of the previous fiscal year.

Table E.2 ITT analysis with four-month window

	(1) OLS	(2) Probit	(3) Tobit
Panel A: All tax types			
Treatment	3,991 (13,803)	0.52** (0.26)	21.9* (11.3)
Treatment x large	2,612,601 (2,231,082)	–0.82 (0.53)	–33.1 (20.3)
Marginal treatment effects, small taxpayers:			
Predicted probability			
Effect of censored revisions		0.005**	
Effect conditional on positive revision			0.06** 1.70**
Significant effects for large taxpayers	No	No	No
Other controls	Yes	Yes	Yes
Observations	1,977	1,977	1,939
Panel B: VAT			
Treatment	15,669 (25,625)	0.47** (0.28)	19.8* (11.8)
Treatment x large	2,625,122 (2,380,975)	–0.78 (0.54)	–30.7 (20.3)
Marginal treatment effects, small taxpayers:			
Predicted probability		0.009*	
Effect of censored revisions			0.10*
Effect conditional on positive revision			1.70*
Significant effects for large taxpayers	No	No	No
Other controls	Yes	Yes	Yes
Observations	1,196	1,196	1,162

Notes: Standard errors are in parentheses. ** $p<0.05$, * $p<0.1$. Marginal effects are evaluated at the mean. An OLS model is estimated in column (1), where the dependent variable is equal to the level of revision. A probit model is estimated in column (2), where the dependent variable is equal to one if the taxpayer positively revised the declaration at least once. Column (3) estimates a tobit model, where the dependent variable is the logarithm of the level of revision. All regressions include a constant and controls for size, riskiness, and the latest available tax due, as well as revision and principal of the previous fiscal year.

Table E.3 LATE analysis with four-month window

	(1) IV	(2) IV-probit	(3) IV-tobit
Panel A: All tax types			
Treatment	4,913 (23,172)	0.77 (0.55)	32.8 (24.0)
Treatment x large	3,960,252 (3,384,534)	−4.53 (216)	−175.4 (9,825)
Other controls	Yes	Yes	Yes
Observations	1,977	1,977	1,939
Panel B: VAT			
Treatment	20,901 (38,222)	0.69 (0.51)	28.6 (22.1)
Treatment x large	3,932,477 (3,564,325)	−4.44 (201)	−170.3 (8,445)
Other controls	Yes	Yes	Yes
Observations	1196	1,196	1,162

Notes: Standard errors are in parentheses. ** p<0.05, * p<0.1. An IV regression is estimated in column (1), where the dependent variable is equal to the level of revision. A two-step IV-probit model is estimated in column (2), where the dependent variable is equal to one if the taxpayer positively revised the declaration at least once. Column 3 estimates a two-step IV-tobit model, where the dependent variable is the logarithm of the level of revision. All regressions include a constant and controls for size, riskiness, and the latest available tax due, as well as revision and principal of the previous fiscal year.

References

Allingham, M.G. and Sandmo, A. (1972) 'Income Tax Evasion: A Theoretical Analysis', *Journal of Public Economics* 1: 323–338

Angrist, J.D and Pischke, J-S. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton, NJ: Princeton University Press

ATAF (2016) *African Tax Outlook*, Hatfield: African Tax Administration Forum Secretariat

Bell, R.M. and McCaffrey D.F. (2002) 'Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples', *Survey Methodology* 28(2): 169–179

BIT (2012) *Applying Behavioural Insights to Reduce Fraud, Error and Debt*, London: Cabinet Office Behavioural Insights Team

Bloom, H.S. (2008) 'The Core Analytics of Randomized Experiments for Social Research. In P. Alasuutari, L. Bickman and J. Brannen (eds.), *The SAGE Handbook of Social Research Methods*, Thousand Oaks, CA: SAGE

Cameron A.C. and Miller D.L. (2015) 'A Practitioner's Guide to Cluster Robust Standard Inference', *Journal of Human Resources* 50(2): 317–373

Del Carpio, L. (2014) 'Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru', mimeo, Princeton University

Erard, B. and Feinstein, J.S. (1994). 'Honesty and Evasion in the Tax Compliance Game', *RAND Journal of Economics*, 25(1), 1–19

Kleven, H.J., Knudsen, M.B., Kreiner, C.T, Pedersen, S. and Saez, E. (2011) 'Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark', *Econometrica* 79 (3), 651–692

Luttmer, E. and Singhal, M. (2014) 'Tax Morale', *Journal of Economic Perspectives* 28(4), 149–68

Mascagni, G. (2016) *From the Lab to the Field: a Review of Tax Experiments*, ICTD Working Paper 46, Brighton: Institute of Development Studies, UK

—— Monkam N. and Nell, C. (2016a) *Unlocking the Potential of Administrative Data: Tax Compliance and Progressivity in Rwanda*, ICTD Working Paper 56, Brighton: Institute of Development Studies

—— Nell C. and Monkam N. (2016b) *One Size Does Not Fit All: A Field Experiment on Drivers of Compliance and Delivery Methods in Rwanda*, ICTD Working Paper, Brighton: Institute of Development Studies (forthcoming)

Myles, G. D. and Naylor, R. A. (1996) 'A Model of Tax Evasion with Group Conformity and Social Customs', *European Journal of Political Economy* 12(1), 49–66

Ortega, D. and C. Scartascini (2016) 'Don't Blame the Messenger: A Field Experiment on Delivery Methods for Increasing Tax Compliance', mimeo, Inter-American Development Bank

Schneider F. and Williams, C. (2013) *The Shadow Economy*, London: The Institute of
    Economic Affairs

Traxler, C. (2010) 'Social Norms and Conditional Ccooperative Taxpayers', *European
Journal of Political Economy* 26(1), 89–103

Yitzhaki S. (1974) 'A Note on Income Tax Evasion: A Theoretical Analysis', *Journal of Public
    Economics* 3, 201–202

ICTD