EVIDENCE REPORT
No 97

Policy Anticipation, Response and Evaluation

A Review of Tax Experiments: from the Lab to the Field

Giulia Mascagni

October 2014

The IDS programme on Strengthening Evidence-based Policy works across seven key themes. Each theme works with partner institutions to co-construct policy-relevant knowledge and engage in policy-influencing processes. This material has been developed under the Policy Anticipation, Response and Evaluation theme.

The material has been funded by UK aid from the UK Government, however the views expressed do not necessarily reflect the UK Government's official policies.

AG Level 2 Output ID: 534

A REVIEW OF TAX EXPERIMENTS: FROM THE LAB TO THE FIELD

Giulia Mascagni

October 2014

First published by the Institute of Development Studies in October 2014

Contents

1 Introduction

Tax revenue mobilisation is attracting increasing attention among researchers and policymakers in developed and developing countries alike. There is widespread recognition that increasing tax revenues should be at the core of development strategies, with the aim of reducing dependence on foreign aid, increasing domestic resources to finance development plans, and stimulating transparency and accountability in government–citizen relations. The recent global financial crisis has certainly given momentum to the issue. On the one hand, it made fiscal matters central to the policy debate in developed countries, which resulted in a higher priority in the development debate as well. On the other hand, it made the possibility of scaled-down aid more concrete. For their part, developing countries – and particularly aid-dependent African countries – have been giving higher priority to domestic revenue mobilisation. This has resulted in widespread reforms on the policy side (most notably with the introduction of value added tax, or VAT) and on the administration side (with the establishment of semi-autonomous revenue authorities, among other initiatives, to modernise tax administration).

This heightened policy interest has partly been fuelled by research on the relationship between taxation and state-building (for example, Brautigam, Fjeldstad and Moore 2008), and it has in turn sparked more research in this area. This increasing body of research and the newfound prominence of tax in the policy debate has led to consensus on a reform agenda that developing countries have followed rather consistently (Fjeldstad and Moore 2008; Fjeldstad 2013; Moore 2013). Despite this consensus, tax compliance in many developing countries is still low and there is limited evidence about the most effective policy options at national level. While revenue agencies often pilot and evaluate possible policy options to increase compliance, this has not always been done in a rigorous way or in collaboration with researchers. In particular, the area of taxation lends itself to experimental evidence, which has recently become the gold standard in development economics.

Methods based on randomisation have attracted praise from some quarters but critics from others. Among the critics, the World Bank's Chief Economist, Kaushik Basu, recently argued that while randomised controlled trials (RCTs) are of immense value in economics and represent the gold standard to *describe* large populations, the claim that they can offer insights into causality is misplaced (Basu 2013). Other critics have raised issues regarding external validity and the usefulness of experimental evidence when interventions are scaled up beyond the village level. For example, concerns have been raised about the external applicability of results from experiments: an RCT may, for example, be informative about the average effect of a drug, but a doctor may be right about not prescribing it to a specific patient. Similarly, a policy intervention may prove effective in an RCT but it may not be the right policy option in the specific time and context where the policymaker operates (Deaton 2009). These are legitimate concerns, as are questions regarding the design of experiments, which needs to be rigorous to ensure internal validity. RCTs may well have been praised more than they merit, but it is beyond doubt that randomisation is the new standard in current research in development economics. Despite its limitations, it has probably contributed to bringing more rigour in applied economics methods and thinking.

This study reviews the application of randomised experiments in the area of taxation, with a twofold objective. First, I review the existing evidence with particular attention to the specific topics explored through tax experiments and their methods, highlighting the main strengths and weaknesses in the literature. Second, I search for evidence (particularly from developing countries), including published work and ongoing projects, with the aim of assessing the feasibility of more research in Africa and other developing regions. The study therefore does not have geographical limitations in terms of the literature covered, as research carried out in developed countries can inform future research in developing countries. However, the motivation and objectives of this study are specifically related to developing countries, to inform future research in those contexts.

Most of the research reviewed here is in the form of published journal articles or working papers. However, the published literature largely under-represents developing countries. Therefore, this review considers both published research and ongoing projects. This approach is a pragmatic response to the fact that tax experiments are relatively new in developing countries, while the research process leading up to publication – including design, data collection and analysis – tends to be rather long. Only considering completed and published work would miss out the increasing number of projects currently being implemented. This study therefore includes ongoing projects in developing countries that have completed the research design phase and, in some cases, published progress reports, including preliminary findings. Given the importance of the design phase in experimental research, ongoing projects are informative and useful to assess the potential for further experimental research on taxation in developing countries.

This study contributes to the literature base by providing a review of existing experimental evidence and ongoing projects in developed and developing countries. A review of the experimental evidence literature on taxation was carried out more than a decade ago by Benno Torgler (2002). This study builds on Torgler's review and expands it by including more recent studies as well as ongoing projects in developing countries. Given the increasing interest and research activity in recent years in both taxation and experimental methods, an updated review is both timely and relevant. By taking stock of existing research with particular attention to developing countries, this review also represents a stepping stone for further experimental research in the area of taxation. On the policy side, it contributes to the growing debate on tax compliance in developing countries, which is one of the big challenges of tax revenue mobilisation.

1.1 A bird's eye view on tax experiments

The review by Torgler (2002) divided tax experiments into two broad categories based on the definition provided by Roth (1995: 22): 'speaking to theorists' and 'searching for facts'. In both cases, the primary issue is compliance, particularly in the case of income taxation. The first category includes experiments aimed primarily at testing theories and whose results feed back into the theoretical debate and literature. For example, several early experiments tested the empirical validity of the economic model of taxpayer behaviour set out in Allingham and Sandmo (1972). These 'speaking to theorists' studies therefore focused primarily on the role of deterrence, sanctions and expected economic returns to explain compliance. In this case, experimenters speak to theorists. The second category includes experiments that analyse variables not explicitly included in theoretical models, but that are motivated instead by the results of other experiments and observations of what actually happens. One of these observations is that the levels of compliance in the real world are much higher than those predicted by traditional models. This has prompted researchers to explore the role of public goods, and moral and social factors in tax compliance, in addition to the traditional economic considerations. In this case, experimenters are talking to each other (Torgler 2002). Tax experiments looking at these issues are 'searching for the facts', although they speak to theorists too since the new observed patterns can feed back into the theory to develop existing models further.

While the categorisation proposed by Torgler (2002) is very useful, it ignores the third category proposed by Roth (1995: 22): 'whispering in the ears of princes'. In this category, experimenters speak to policymakers, with varying degrees of success. The experimental environment is designed to closely resemble the natural environment that is of interest for policymakers, or in some cases the experiment actually takes place in a real-life setting. At the time Roth was published (1995), there was not an extensive literature of this sort, but the author recognised its potential for fostering the role of economists as policy advisers. In fact, even by 2002, when Torgler wrote his review, this category was still rather thin, thus justifying its exclusion from Torgler's work. However, since then, policy-oriented experiments have been thriving, forming a growing set of studies that can no longer be ignored. These tend to be field rather than lab experiments, based on large samples, and displaying a high level of engagement and collaboration with local governments. Of course, 'whispering in the ears of princes' is a difficult thing to do. There are critics of the idea that a direct link can exist between experimental evidence and policy, two of the most prominent critics being Deaton (2009) and Basu (2013). Rather than giving a general assessment on whether experiments can actually be translated into policy, this study aims to include the emerging 'whispering in the ears of princes' literature and to critically assess its potential to inform and influence policy in the specific field of tax research.

The three categories of experiments described so far also represent the evolution of this field of research over time, which has seen tax experiments progressively moving away from theory and closer to real-life experiences. This 'movement towards reality' has involved both the introduction of new questions and changes in the design of experiments. For example, questions on moral factors and political legitimacy have become more common. As far as design of experiments is concerned, researchers have increased sample sizes, tried to include real taxpayers, and have tried to approximate real-life conditions by using tax-specific language and simulations of exchange of resources. However, it is important to note that these are still artificial settings, and the 'movement to reality' can be considered complete only with the large-scale field experiments described in Section 3.

It is possible to make another distinction in relation to tax experiments based on the type of experiment, namely: lab experiment, field experiment, and quasi-experimental study. In a way, introducing this further distinction closely overlaps with the three categories described by Roth, while it also reflects the evolution of tax experiments over time. Lab experiments are typically speaking to theorists and they involve a simulation (more or less realistic) of real-life tax payments. Most experiments that are 'searching for the facts' are also lab-based. Field experiments typically aim to evaluate concrete policy options and they mostly belong to the third category, 'whispering in the ears of princes'. Finally, quasi-experimental methods can be used to evaluate real-life interventions – for example, in cases where groups that are sufficiently similar to make comparison sensible have received different treatments under existing policies. Comparing compliance outcomes between the group that has been affected by the policy and the one that has not can offer insights into the effectiveness of the intervention. Quasi-experimental studies[1] would fall into the third category of tax experiments, but they are excluded from this review.

This study builds on these overlapping categorisations and reclassifies tax experiments into two main groups, which correspond to the structure of this report. Tax experiments based on 'speaking to theorists' and 'searching for the facts' are reviewed in Section 2. It includes traditional tax experiments and shows their progressive move towards more realistic experimental settings. These tax experiments are mostly not concerned with developing countries in particular, but they set the conceptual framework for more recent research. Even when they are 'searching for facts', they have a strong theoretical component and are not primarily concerned with policy. For this reason, the first two categories are grouped together. Tax experiments based on 'whispering in the ears of princes' are reviewed in Section 3, covering the emerging stream of experiments that are more policy oriented, with a particular but not exclusive focus on those specific to developing countries.

[1] Two examples of quasi-experimental studies of taxation are Waseem (2013) and Gadenne (2013).

2 Speaking to theorists and searching for the facts

The earliest tax experiments were largely concerned with testing the model of taxpayer behaviour developed by Allingham and Sandmo (1972), which focused on income taxes. This model is embedded in the tradition of the economics of crime and it adopts the same logic to explain tax evasion. It holds that a taxpayer's decision on whether to evade (and to what extent) is determined by economic factors, and particularly by the expected returns to evasion, which in turn are affected by the tax rate, taxable income, the probability of being detected, and the size of the penalty to be paid if found guilty. The early literature on tax experiments (in the 1970s and 1980s) distinguishes itself from the previous empirical tax literature by explicitly trying to capture causality rather than simple correlations. The authors engaging with tax experiments at this time argued that a lab setting allowed greater control of the environment and independent variables than real-life or survey-based research would allow. They also argued that experiments allowed them to circumvent the problem of getting honest answers about dishonest behaviour by working within hypothetical settings (Spicer and Thomas 1982; Friedland, Maital and Rutenberg 1978).

These early experiments tested the empirical validity of the economic factors set out in the Allingham and Sandmo (1972) model on two variables: the probability of evading and the amount of taxes evaded in the form of under-reported income. The pioneering simulation study by Friedland *et al*. (1978) analysed changes in the tax-paying behaviour of 15 Israeli psychology students when these economic factors vary. They found that reported income was very elastic to tax rates after a certain threshold. In other words, when the tax rate rose above a certain level, taxpayers (or, more accurately, the participants in the experiment) responded by sharply decreasing their reported income. Moreover, the study found that large sanctions are more effective than frequent audits in improving compliance.

The authors also found that decisions about whether to evade (and by how much) were determined by different factors. The rate of tax was found to be the most important determinant of the probability of evasion, whereas other socioeconomic factors determined the extent of evasion. Although pioneering experiments in this field, Friedland *et al*. (1978) presented a number of typical drawbacks of the literature: a very small sample size (only 15 individuals observed), all of whom were students rather than taxpayers. Moreover, the experimental environment in this study was artificial, thus raising concerns about the possibility that actual decisions on real tax returns may be very different from those made in the simulation. However, the authors fully acknowledged these drawbacks and suggested directions for future research, based on increased realism, consideration of cultural differences, and the inclusion of additional factors such as uncertainty and public spending.

Uncertainty is indeed one of the factors that was picked up rather early in this literature, recognising that in the real world, taxpayers may not have access to full information, particularly about the probability of audit. Spicer and Thomas (1982) used a sample of 54 students from the University of Colorado to find that the predictions of the economic model on the relation between evasion and the probability of audits is only empirically valid when taxpayers have full information about these probabilities. When information is available, an increase in the probability of audit results in a reduction in the propensity to evade, as predicted by the model – but does not result in changes in the extent of evasion. Beck, Davis and Jung (1991) reinforce these results[2] by finding empirical support for the

[2] Their sample was composed of 112 graduate and undergraduate students from the University of Illinois.

theoretical predictions that evasion is lower when audit probabilities and sanctions are higher, and by confirming that uncertainty weakens these effects. In other words, in the presence of uncertainty, taxpayers tend to evade less when audit probabilities and sanctions are low. Alm, Jackson and McKee (1992b) challenged the view that taxpayer uncertainty increases compliance. While taxpayers did indeed evade less under conditions of uncertainty, this relation was reversed when the authors introduced public goods into the experiment. If participants were told that tax revenues would be used to benefit everyone, to simulate a public good, then increased uncertainty led to less compliance. With this finding, the authors highlighted the important role played by the institutional environment in which the compliance decision is made.

In addition to uncertainty about the probability of audit, individuals may differ in their perceptions about this probability. Spicer and Hero (1985) tested two hypotheses using a group of 36 psychology students from the University of Colorado: first, that the taxpayer will evade more when he or she perceives that the level of evasion in general is higher; second, that taxpayers who have been audited in the past will perceive a higher probability of audit and will therefore tend to evade less. The authors found no support for the former but did find evidence for the latter. Along similar lines, Spicer and Becker (1980) looked at how perceptions of fiscal equity influence tax evasion. They gave 57 students from the University of Colorado different information regarding the average tax rate in the population, while they all faced the same actual tax rate. The authors found that perceptions of fiscal equity matter: evasion was higher among those who were told their rates were higher than the population average, and lower among those who were told their tax rates were lower than average.

2.1 Institutions, social norms and emotions

The results from the early 'speaking to theorists' tax experiments still failed to explain the actual levels of compliance observed in reality, which are consistently higher than predicted by economists, even in contexts of relatively low enforcement. Taxpayers apparently do not behave as pure gamblers who maximise income (utility) under uncertainty. So the puzzle is really about why there is so little cheating, rather than why there is so much (Alm, Sanchez and de Juan 1995). This observation prompted researchers to go beyond deterrence and risk, and to search for the facts that explained this difference. They focused on two elements: social norms and institutions.

Following up on the seminal work of Alm *et al*. (1992b), some studies have focused on public spending and institutional factors. Becker, Buchner and Sleeking (1987) argued that taxpayers take into account government transfers when assessing their tax burden and making decisions on evasion. They therefore test the hypothesis that transfer payments by the government are negatively related to evasion – the more they benefit from public spending, the less likely taxpayers are to evade paying taxes. The authors found that not only do transfer payments matter for compliance, but that the perceived equity in such transfers helps fight evasion. The authors attempted to increase the level of realism in the experiment by, for example, allowing for uncertainty in audit probabilities and a simulation of earned income. However, they recognised that the setting was still somewhat artificial[3] and therefore that the results might not have reflected real-life dynamics. Similarly, Mittone (1997) considers that rational utility-maximising taxpaying behaviour would take into account both tax payments and transfers derived from the redistribution of the tax yield. The author found[4] that tax yield redistribution and audits have the most robust effect in reducing tax evasion. Alm and Jackson (1993) analysed fiscal exchange, the use of tax revenue, and the decision processes for deciding such uses. They found that individuals increase compliance when they know that the tax yield is spent on programmes they approve of and when they are an active part of the decision on spending. Moreover, they found that compliance is higher when

[3] Participants comprised 85 students from the University of Bonn and 31 from the University of Cologne.
[4] The author uses 60 subjects for the one-shot experiments and 30 for the dynamic games.

the outcome of the vote reflects widespread support rather than a close win. On the contrary, compliance is reduced when unpopular expenditure programmes are imposed top-down on participants.

In a related study to that carried out by Mittone (1997), and using the same sample,[5] Bosco and Mittone (1997) found that tax evasion is reduced more by a feeling of collective blame, than by reflections on the redistribution of the tax yield and on the reduced welfare for society resulting from evasion. Alm, McClelland and Schulze (1999) looked at how social norms influencing compliance can be affected by voting and communication on the basic parameters of the tax system. When asked to vote on the desired level of enforcement, participants[6] voted it down to very low levels. This resulted in lower levels of compliance in the next rounds of the experiment, as the social norm around evasion may have been changed in the voting process to justify lower compliance. However, when participants were allowed to talk to each other, they voted for higher levels of enforcement, and evasion levels dropped dramatically. So communication changed the social norm in favour of higher compliance.

One of the first studies to look explicitly at social norms was Alm and Jackson (1993). The authors carried out experiments with students from Spain and the US[7] – where attitudes about tax evasion differ substantially – to show that social norms matter not only for the level of compliance but also in the responses to policy interventions designed to increase compliance. Generally, they found support for the traditional deterrence and punishment factors, but argued that compliance is also significantly improved with positive rewards, a wise use of public revenue and social obligations. Alm, Jackson and McKee (1992a) reinforced the finding on positive rewards by testing four different options of incentives, or treatments: (1) a lottery that could be accessed by those who were tax compliant, where the chance of winning was 1 in 25; (2) a fixed reward for compliant subjects; (3) an audit reduction; and (4) a public good. The lottery had the largest effect on compliance, although it had the same expected value of the fixed reward. Audit reductions and public goods improved compliance but less so than other treatments. The authors therefore concluded that positive rewards need to be immediate and salient to have a significant effect.

All the studies reviewed so far present some common characteristics. First, they all used students from various universities as experimental subjects. While students will respond to the incentives set out in the experiment, they are not real taxpayers and they may not have any experience at all with taxpaying in real life. Second, these studies usually relied on rather small samples, often involving only a small number students in repeated rounds. Finally, they still used largely artificial experimental settings, despite efforts to move to more realistic settings. In a lab setting, the individuals involved obviously know that they are part of a study and they may behave differently than they would under the pressures of real life. So there remains a strong possibility that behaviour in the artificial environment of the lab differs from actual decisions when faced with a real tax return in a real social context. This problem has largely prevented experiments from having strong policy impact; indeed, all authors involved recognise that results must be taken with caution.

Despite these problems, current research still uses traditional lab experiments. The most recent lab tax experiments contribute to the literature by explicitly linking the traditional economic factors with psychological theories of shame and emotions. In doing so, they also borrow from psychology the methods used to measure these factors. Coricelli *et al.* (2010), for example, argued that while the literature has focused extensively on the monetary factors

---

[5] They only carry out one-shot games, so only that sample of 60 subjects is the same.
[6] The study is based on observations of the behaviour of 11 undergraduate students for each round from the University of Colorado.
[7] About ten students were involved in each round, from Universidad Carlos III de Madrid and the University of Colorado in the USA.

in the decision to cheat, psychological costs and benefits have largely been ignored. They therefore used data from self-reports and skin conductance responses to investigate the emotional dimension of tax evasion. The experiment, involving eight subjects in each repeated round, showed that taxpayers are not pure gamblers since their decisions are not dictated by monetary calculations alone, but also by moral and emotional implications. The emotional intensity preceding the decision to evade is positively associated with actual evasion. Moreover, the emotional intensity at the time of audit is increased by the prospect of both monetary and non-monetary sanctions, including public disclosure of deceptive behaviour. The fact that the picture of cheaters will be displayed, if they are detected, increases emotions and deters evasion, even in the artificial lab environment.

The authors also found that audits increase emotional intensity and decrease the extent of cheating in the next period, but larger fines increase the proportion of evaded income. Realism is increased by using non-neutral language, by replicating the real structure of income reporting, and by making audits endogenous to the median report in the group. Coricelli, Rusconi and Villeval (2014) investigated two types of shaming for cheating: one that results in stigmatisation and exclusion from the community, and another that leads to reintegration. In both cases, the shaming ritual involved displaying, in public, a picture of the participant who chose to evade, as well as monetary sanctions. The authors found[8] that the first type of cheating, resulting in exclusion from the community, significantly increased evasion compared with the second type of shaming, which leads to reintegration. In particular, the shaming and exclusion treatment was associated with more intense negative emotions related to tax evasion. Coricelli *et al*. (2014) concluded that while shaming can potentially be a powerful tool to increase compliance, sensitivities should be taken into account. Calvet Christian and Alm (2014) provide further support for the inclusion of non-monetary factors in the analysis of tax compliance. They investigated the role of empathy and sympathy in tax compliance games, which are measured using methods derived from psychology such as the Davis Empathic Concern Scale. They found that the presence and promotion of empathy and sympathy increased compliance in most cases.

2.2 Lab in the field

In an attempt to achieve greater realism in tax experiments, researchers have moved the lab to the field. Cummings, Martinez-Vazquez and McKee (2001) argued that tax experiments that focus exclusively on social norms fail to fully understand cross-country differences in compliance behaviour. Instead, such differences result from the way in which tax administration is managed and depend on the relation and fiscal exchange between citizens and governments. To capture these issues, Cummings *et al*. conducted laboratory experiments in three countries: the USA, South Africa and Botswana. The three countries present many similarities in their tax systems but public perceptions of government and of the equity of the tax system in each country vary. However, the precise criteria for case selection are not clear. To achieve greater realism, the experiments replicated the real tax structures of the three countries. Moreover, they used individuals who had previously filed their own tax returns to ensure that they had some real-life experience of taxpaying. Contrary to some of the previous experimental designs, the authors used and emphasised tax-specific language rather than neutral language. This would arguably help individuals in bringing their own tax experiences to the lab, but it would also highlight national or local cultural factors that may influence perceptions about taxation. As far as economic factors were concerned, they found that compliance was more affected by audit probabilities than by the size of sanctions. They also found that differences in compliance behaviour are, indeed, closely related to differences in tax institutions and government behaviour.

[8] The experiment was conducted at the *Groupe d'Analyse et de Théorie Economique* (GATE), CNRS, France, with 32 students divided into two groups of 16.

A similar experimental design was used by Torgler (2003) in Costa Rica, which included 37 real taxpayers and used real money, but adopted neutral rather than tax-specific language. The focus of the experiment was on fiscal exchange, moral suasion, and positive rewards – thus keeping as constant the traditional economic factors related to deterrence. These three factors correspond to three treatments, which are evaluated against a control group. The results showed that positive rewards and moral suasion were effective in increasing compliance. However, the experiment only included one round (no dynamics), so questions on the sustainability of moral suasion over time and on learning effects necessarily remain unanswered.

Torgler supports the consideration of positive incentives as an alternative policy instrument to deterrence. However, he warns that the results should be taken with caution due to the very small sample size. Indeed, while they may be informative, 'lab in the field' experiments are still rather artificial. The inclusion of real taxpayers certainly helps realism, but the samples are still not representative of the population in the countries considered. Second, using tax language and/or real money does not circumvent the problems related to artificiality, as subjects are still likely to behave differently than in real life. Real taxpayers may not behave as such in the experiment, as they know they are part of a study with no real-life consequences. Therefore, these tax experiments are still subject to criticism regarding their external validity, although they represent a step forward towards greater realism.

3 Whispering in the ears of princes

The biggest drawback of the lab experiments reviewed in the previous section is related to the experimental setting. While the lab allows researchers to control all variables and perhaps get more honest answers about evasion, it is still an artificial environment, with artificial taxpayers, which may lead to different decisions than in real life. The evolution of tax experiments was therefore characterised by the continuous efforts of researchers to bring the lab closer to reality, both by including additional factors in the analysis (for example, uncertainty or social norms) and by changing the experimental design (for example, using real taxpayers, or 'lab in the field'). The previous section has followed this journey from purely theory-based lab experiments to lab-in-the-field studies. However, it is only recently, with real-life and large-scale field experiments, that that journey has led to a sufficiently realistic research design. Despite their drawbacks, lab experiments offer a conceptual framework that underlines the relevant elements in the analysis of tax compliance, such as deterrence factors and social norms. This conceptual framework is the basis for the more realistic and policy-oriented tax experiments reviewed in this section.

In the past 10–15 years, a new category of experiments has emerged that has explicitly tackled the issue of artificiality. The characteristic feature of these experiments is the use of large-scale data based on actual tax returns and audits, collected with the collaboration of revenue authorities. This represents a considerable improvement in research design, as experiments are conducted in a real-life environment where the interface is a real tax authority rather than the researcher. By doing this, and contrary to lab tax experiments, real-life experiments face the challenge of working with variables that may not be perfectly observable (such as unreported income). While in a lab setting, both real and reported income can be observed, in real life, tax authorities will only know how much income is reported in tax returns.

A further improvement is the increased attention paid to obtaining random and representative samples of the population, and the use of adequate econometric techniques to deal with large data sets. Given their increased realism and policy orientation, these tax experiments are concerned not only with estimating the effects of different policy options, but also with understanding the mechanisms. They therefore often stratify the samples according to relevant criteria, to identify how segments of the population respond differently to the treatments. In some cases, different experiments are carried out to test the validity of different possible mechanisms (see, for example, Kleven *et al.* 2011; Paler 2013). By doing this, this category of tax experiments is much better placed to 'whisper in the ears of princes' and advise policymakers on the effectiveness of different policy options.

3.1 Deterrence and moral appeals

Slemrod, Blumenthal and Christian (2001) carried out a pioneering study in this category. It was based on a sample of 22,368 tax returns from taxpayers in Minnesota, US. The authors, in close collaboration with the Department of Revenue, collected income tax data for two subsequent years: 1993 (tax returns filed in 1994) and 1994 (returns filed in 1995). In January 1995, at the beginning of the filing season for the second year, a treatment group of about 1,500 taxpayers[9] received a first-class mail from the Commissioner of Revenue stating that they had been selected to be part of a study and that their state and federal tax returns would be closely examined. The letter also stated that taxpayers would be contacted regarding any discrepancy and that if any irregularities were found, their previous tax returns would be scrutinised as well. The study then assessed the effect of this increase in the

[9] The characteristics of this treatment group were not statistically different from the control group in the baseline year.

perceived probability of audit on the accuracy of income tax reporting. A difference-in-difference methodology was used to analyse the changes in reported taxable income (or tax liability) between 1993 and 1994 in the treatment and control groups. Changes in reported income were therefore used as an indication of changes in compliance, as data on the audit results (or 'close examination') were not available to the authors.

The study further stratified taxpayers along two dimensions: income, and opportunity to evade, defined on the basis of different income tax schedules. The study found that for low- and middle-income taxpayers alike, the audit letter had a significant impact on reported income (corresponding to a 12 per cent increase in tax payments) in a situation of high opportunity to evade. The result was even more significant for low-income taxpayers, although it was less significant than for the middle-income strata. Among the high-opportunity taxpayers, those subject to the schedule for self-reported income represented a major component of the treatment effect. For low-opportunity taxpayers in the low- and middle-income groups, the letter had a much smaller effect, although still a positive one. The result on high-income taxpayers instead is somewhat perverse: they tend to report less income as a result of the treatment letter, indicating lower compliance as a result of increased audit probability. One possible explanation is that high-income taxpayers sought professional advice in response to the letter, which may have helped in finding legitimate ways to reduce their tax returns. The authors also underline a couple of caveats of the experiment design. First, some taxpayers may already have thought, before the letter, that their returns were closely examined by the authority. In this case they would not respond to the treatment. They may also not have believed that such close inspection was actually capable of uncovering possible irregularities. Second, the indication in the letter that previous returns could also be inspected may have induced taxpayers to maintain their reporting pattern to avoid uncovering a history of non-compliance.

Another study by the same authors (Blumenthal, Christian and Slemrod 1998) used a very similar design and method to Slemrod *et al*. (2001), and was also conducted in Minnesota. This time, however, the authors investigated the effect of moral appeals by using two treatments: (1) a letter encouraging taxpayers to comply voluntarily to support the provision of socially valuable goods and activities; and (2) a message indicating that the majority of citizens comply with tax laws and that one should be compliant in order to be part of that majority. The authors found little or no evidence that the letters resulted in any changes in compliance behaviour. However, they confirmed the finding of a perverse effect on high-income taxpayers, whereby an increase in the probability of audit resulted in a decrease in reported income. They also found that high-opportunity taxpayers[10] were less likely to respond to normative appeals.

Moral appeals are also explored by Torgler (2004) in a field experiment conducted in collaboration with the local tax administration in Trimbach, Switzerland. The decentralised tax system of Switzerland was seen as a good case to explore moral motives for compliance, since proximity to the government at the local level may make moral appeals more effective. The experiment involved 580 randomly selected individual taxpayers who were divided into two groups. The treatment group received a letter signed by the commune's fiscal commissioner in February 2002. The letter was sent in a separate envelope and printed on a pink sheet to increase taxpayers' awareness of it. Participants were not informed they were part of a special study, in order to reduce bias. The letter underlined the importance of voluntarily complying, to allow the commune to continue providing public goods for the community. The fact that the community of reference was relatively small and close may have made this message more effective. The effect of the letter was assessed against two compliance indicators – timely filing and timely paying – both of which can be observed. Despite the fact that the author conducted the experiment at the local level, he confirmed the

[10] High-opportunity taxpayers are those with self-reported income or rents and royalties income.

finding of Blumenthal *et al.* (1998) that moral suasion has hardly any effect on compliance. The treatment group did not significantly change the timeliness of their returns or payments in response to the moral appeal.

Fellner, Sausgruber and Traxler (2013) also confirmed this finding in the case of TV licence evasion in Austria. Their experiment involved data on 50,498 individuals, five per cent of whom were randomly selected as a control group who received no letter. The standard letter by the Fee Info Service (Gebuhren Info Service, GIS), the managing agency for the TV licence, was used as a baseline. The authors then added three additional treatments to this letter: a threat, social information, and a moral appeal. Moreover, for some participants, the threat treatment was combined with either social information or a moral appeal, making a total of six different mailing treatments. The sample used was not randomly selected but was composed of potential evaders. The letters clearly indicated that participants were selected because they were suspected of evasion. This has implications for the interpretation of results, as this sample of potential evaders may be less motivated by moral appeals. Compliance, the dependent variable, was estimated as a binary outcome indicating whether or not an evader starts to comply.

Generally, the authors found that addressing evaders with a letter had a net revenue result of 15 euros per letter, as the proportion who started paying after receiving the letter was higher than in the control group (who received no letter). A separate survey designed to investigate the mechanisms suggests that this effect may be due to a signal of surveillance and of increased sanction risk. The threat letters had the strongest effect, while moral suasion and providing social information did not increase compliance. Given the sample chosen, however, the finding that moral appeals are irrelevant may not be generalisable. Evaders are likely to care less about social and moral factors than non-evaders, and they would naturally respond less, or not at all, to such appeals. However, the findings suggest that sending letters is not effective in enforcing compliance among evaders. The authors also found heterogeneous effects of giving social information to evaders. When evasion is believed to be common, the information treatment has a weak positive effect; while it has a weak negative effect when evasion is believed to be rare.

Castro and Scartascini (2013) analysed property taxes in a municipality in Argentina to test the effect of different factors on taxpayer compliance, including deterrence but also equity and fairness. The property tax considered was based on the linear front size of the property. It therefore has the advantage of making both the tax base and compliance observable by the authorities. The tax is linked to the provision of public goods by the same level of government that levies the tax. In total, 23,000 individual taxpayers were randomly divided into four groups, three of which received a treatment letter influencing their perceived probability of detection and their beliefs. The first letter was about enforcement and fines; the second gave information on other taxpayers' behaviour, thus affecting beliefs on equity; and the third gave information about how government resources are used, with the aim of affecting recipients' perceptions of fairness.

Compliance increased by almost five percentage points in the deterrence group who received the first letter. However, the other treatments did not have any effects compared with the control group. The reasons for this lack of support for the importance of moral factors may be varied. Moral suasion may not have a large effect if the elasticity of compliance to perceptions is low. Moreover, people who do not comply may have lower trust in the government, therefore making the moral appeals less effective. It is also possible that the 'equity' letter informing taxpayers of other people's level of compliance may have 'changed their beliefs about the prevalence of evasion (being greater than they thought), therefore encouraging them to evade. An additional finding is related to heterogeneous effects, with own compliance behaviour, the level of provision of public goods, and wealth all affecting the response to the messages. By stratifying the sample, the authors found that past compliance and the level of public goods provision in the area tended to increase compliance. The latter should only be taken as evidence of correlation and not causation, since the fairness message did not trigger a significant response. However, there seems to be an information effect of the third letter, as taxpayers in low public good provision areas responded positively to it. Moreover, wealthier individuals – defined as those who own more properties or larger properties – tended to have lower levels of compliance and were less likely to respond to the deterrence message.

While the field studies reviewed so far suggest that moral and social factors may not play an important role, this result should be taken with some caution. First, the fact that the treatment letters were not effective cannot be generalised to say that all moral appeals do not matter. The studies can only show that specific letters were not effective in a given context, not that moral or social appeals are generally irrelevant to taxpayers' decision to comply. There may be other 'treatments' that stimulate compliance. For example, a recent study (survey-based, non-experimental) shows that a link exists between compliance and public service delivery, but this link depends on specific services that are different in the four African countries analysed (Ali, Fjeldstad and Sjursen, forthcoming). Indeed, moral and social appeals are likely to be more effective in some contexts and less effective in others. This is the case both within countries (notably between the broad taxpayer categories of evaders and compliers) and across countries. Within countries, it is likely that moral factors play an important role in explaining the observed *level* of compliance. However, (positive) *variations* in compliance will rely on evaders starting to comply. This group of people is likely to attach little importance to moral factors, therefore responding less to moral appeals. Across countries, the same appeal is likely to have different effects in different social, institutional and cultural settings.

Supporting the importance of these factors, a recent large-scale field experiment conducted in the United Kingdom (Hallsworth *et al.* 2014) shows that social norms and public goods messages significantly increase compliance. The authors looked at the timely payment of taxes to avoid the issue of unobservability of unreported income. They analysed known outstanding tax payments, thus not having to infer compliance from reported income. The experiment looked at the effect of different letters on whether such outstanding amounts are paid to the revenue authority or not. Five letters were randomised across 100,000 individual taxpayers: three contained messages based on social norms and two on public goods. All letters increased the probability of individuals paying their outstanding taxes, with the greatest effect obtained from a 'minority norm' treatment where taxpayers are warned that 'nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet.'

Following these results, the authors carried out a second experiment where they compared the effect of descriptive norms (i.e. how other taxpayers behave) and injunctive norms (what other people think is the right thing to do). The experiment also considered the interaction between these two norms and it included financial details such as the amount of interest for non-payment and the methods to settle the bill. Descriptive norms were found to have a significantly larger effect than injunctive norms. Moreover, the authors confirmed the findings on the 'minority norm', demonstrating large effects of financial messages. Importantly, these experiments show that large revenue gains can be made by increasing moral costs to non-compliance, with no additional cost to the revenue administration.

3.2 Not just morals: information

Building on the contrasting evidence on moral appeals and on the idea of opportunity to evade (Slemrod *et al.* 2001; Blumenthal *et al*. 1998), Kleven *et al*. (2011) used a tax audit experiment to assess whether taxpayers are unable or unwilling to cheat. They uncovered an additional weakness of the lab experiments, which assume that tax assessments occur through self-reporting only. In the real world, different reporting environments coexist, and they have different implications for the ability to cheat. In particular, taxpayers who are only subject to third-party reporting may not be able to cheat even if they wanted to. For this group, the probability of audit is a poor proxy for the probability of detection, as the authorities could easily uncover any irregularity simply by matching information reports to tax returns. The authors hypothesised that the probability of detection was much higher for self-reported income, and that any tax policy to tackle evasion would affect more self-reported than third-party income.

This experiment was conducted in close collaboration with the Danish tax collection agency and involved 42,800 randomly selected taxpayers in the filing years 2007 and 2008. In 2007, a group of these taxpayers was randomly selected for unannounced tax audits. Irregularities were corrected and penalised according to the law. In 2008, both audited and non-audited taxpayers were selected for pre-announced audits. Two treatments were tested against a control group: one letter informed them that their tax return would certainly be audited; and the other that half the people in the group would be audited. So this treatment exogenously changed audit probabilities.

An important advantage of Kleven *et al*.'s (2011) study with respect to Slemrod *et al*. (2001) and Blumenthal *et al*. (1998) is that the authors had access to audit data and could therefore measure variations in compliance more accurately. Moreover, they used randomised audits and not ordinary ones, which would be endogenous. Participants were not aware of being part of a study. Using these data, the authors showed that compliance rates are high in Denmark despite relatively high tax rates. The tax evasion rate in the third-party reporting environment was close to zero, while it was substantial in the self-reporting environment. They observed the same pattern in individuals who earn both types of income, as they fully complied in the case of third-party reported income but tended to evade on self-reported income. This finding suggests that it is information reporting, rather than an emotional or social aversion to cheat, which drives compliance. Second, the authors used a quasi-experimental setting to assess the effect of changes in the tax rate on compliance, by exploiting salient kinks in the income tax schedule. They found that marginal tax rates do not have a large effect on compliance. Finally, they used the randomised letter treatment to study the effect of the probability of audit. The threats of audit had a large effect on compliance, which was entirely driven by self-reported income. The study therefore concluded that information reporting is crucial in the decision to evade, and that compliant taxpayers are unable, rather than unwilling, to cheat.

Pomeranz (2013) looks further into information as a key determinant of compliance, but departs from the literature by looking at indirect taxes rather than income taxation. In particular, VAT is often praised for its built-in incentive structure, triggered by the paper trail that it generates. The literature has shown that taxpayers would generally respond positively to increased threats of audit. However, if the self-enforcement properties of VAT are effective, the taxpayer response would be very low due to the preventive deterrence effect of the paper trail. To evaluate the validity of this hypothesis, the author compared transactions that were covered by the paper trail (namely those between firms) and those that were not (namely those with final consumers). The study was conducted in close collaboration with the Chilean tax authority, which sent letters indicating increased audit probabilities to more than 100,000 randomly selected firms. This letter experiment simply evaluated changes in compliance due to the increased audit probability. Consistent with the hypothesis, the letters generated a strong increase in tax payments, but this was largely driven by sales to final consumers rather than transactions between firms. Moreover, small firms appeared more likely to respond, perhaps because they tend to have higher shares of retail sales.

A second spillover experiment was designed to explicitly test the mechanism of the paper trail, whose existence was supported by the letter experiment. In this case, half the sample received a pre-announcement of an upcoming audit. The full sample was then audited and the information was made available to the authors. In addition, data were collected for the trading partners of the audited firms, before the audit. Given the existence of the paper trail, one should observe a spillover effect of the audit on the suppliers of the audited firms, which would increase compliance as a result. The author found empirical support for this mechanism and thus argued that the effectiveness of the paper trail is multiplied thanks to spillover effects. Not only do firms comply more with the paper trail because it facilitates detection upon audit, but their suppliers are also encouraged to do so. This spillover effect is due to the fact that the paper trail allows the tax authority to easily identify discrepancies in reporting between the supplier and the audited firm by cross-checking records. The direction of the spillover effect is upwards along the value chain; no significant effects occur for client firms. The paper trail therefore confirms that information is crucial to tax enforcement.

3.3 Government accountability

Finally, a recent experiment (Paler 2013) looked at how different sources of revenue – namely tax and windfall – affect government accountability and its relations with citizens. While this study can be categorised as 'lab in the field', it is included in this section because it displays greater realism than the lab in the field experiments reviewed in the previous section. It used a real public awareness campaign conducted in partnership with a local non-government organisation (NGO), as well as a larger and representative sample.

The author of the study conducted an experiment in 93 villages of the Indonesian district of Blora. Participants were prompted to think about their taxes being managed within the district budget. The experiment then investigated different aspects of government revenue such as the tax burden, the perceived share of taxes relative to windfall revenue, and the perceived ratio of taxes to spending. Variations in these variables were the treatments used to test the hypotheses that taxes are more likely than windfall to make citizens monitor government actions, engage in political participation and support the incumbent government. An additional hypothesis investigated whether citizens mobilise more against misuse of public funds when they pay taxes. Importantly, the study was not only concerned with finding the effects of different sources of revenue on accountability, but also shedding light on the mechanisms through which those effects come about. To this aim, two exercises were carried out: a revenue experiment and an information experiment. The former used different revenue treatments (taxes and windfall) while the latter used information treatments, both aiming to evaluate variations in the outcomes of interest. The rationale for this was to uncover whether the effect of taxes on accountability is motivational or informational.

The author found that the tax treatment led to a greater willingness to monitor the budget, suggesting that taxpaying enhances the demand for information on the public budget. Moreover, the tax treatment in the context of a low information environment motivated participants to sanction the incumbent government. However, paying tax was not associated with higher participation in the windfall scenario.

3.4 Ongoing projects

The literature reviewed so far is rather extensive and growing. However, two gaps are clear. First, the issue of positive rewards has been largely ignored. The lab experiments carried out by Alm *et al*. (1992a), Alm and Jackson (1993) and Torgler (2003) pointed out that positive rewards can be important factors in increasing compliance. This issue has yet to be explored with large-scale field experiments, which would reveal whether positive rewards are effective policy options in real life. Second, while some developing countries (mostly in Latin America) have been included in this literature, African countries are notably absent. Two ongoing projects address these gaps to some extent, but more research is still needed in these areas.

Positive rewards have been used in some Asian countries to support tax compliance (for anecdotal evidence, see Feld, Frey and Torgler 2006) and they can, in principle, be an alternative policy option to deterrence. In fact, positive rewards may be much more effective than deterrence, at least in theory. Psychology and behavioural economics have looked at how external surveillance and the threat of punishment tend to crowd out intrinsic motivation. On the one hand, extrinsic motivations such as external control and threats may decrease tax morale (intrinsic motivation), and even make it disappear. On the other hand, positive rewards tend to be seen as supporting and would therefore crowd in intrinsic motivation (Feld *et al*. 2006).

An ongoing project[11] studies a unique randomised policy adopted in Montevideo, Uruguay (Dunning *et al.* 2014). The municipal tax authority randomly selects taxpayers and, conditional on a history of good taxpaying, rewards them with a year free of tax payments. The research team plans to evaluate the impact of this innovative policy option by using two sources of data: administrative data from taxpayers' records and household survey data, both available over time. They also plan to evaluate how the effectiveness of this programme can be increased through information interventions that will be mailed to households together with their tax bill. By doing this, the project has potential to shed light on whether compliance is motivated more by positive or negative incentives.

The second gap is addressed by another ongoing project, currently in the design phase, looking at local taxation in Uganda. The study seeks to explore the link between revenue sources of the local government (namely taxation and central government transfers) and governance. The main hypothesis underlying the analysis is that revenue mobilisation efforts can increase citizen engagement, accountability and oversight. This study therefore follows a similar direction as Paler (2013), with the important advantage of being based on a large-scale real intervention rather than being a lab-in-the-field study. The intervention is being carried out by the Governance, Accountability, Participation and Performance (GAPP) programme of the United States Agency for International Development (USAID), which includes a sub-programme to enhance local governance through improved tax collection. In addition to providing technical assistance, this sub-programme convenes a series of meetings and conferences aimed at opening lines of communication and collaboration between the local government and taxpayers, by illustrating the role of taxes in the local budget and thus in the provision of public goods and services. The programme will be implemented in 2014–15 in 90 sub-counties of Uganda, while another 90 will serve as a control group. The research team will focus on three outcome indicators: (1) the level of revenue collected and executed expenditures; (2) 'leakage' resulting from corruption among tax officials; and (3) behaviours and attitudes of taxpayers.

[11] Researchers participating in this project are: Thad Dunning (University of California, Berkeley), Fernando Rosenblatt (*Universidad Diego Portales*, Santiago, Chile), Rafael Piñeiro (*Universidad Católica*, Montevideo, Uruguay), Felipe Monestier (*Universidad de la República*, Montevideo, Uruguay), and Guadalupe Tuñón (University of California, Berkeley).

4 Conclusions

The literature on tax experiments has seen tremendous progress since its early days, both in terms of contents and methods. While the literature remains primarily concerned with compliance, its analysis has become increasingly complex. Tax experiments have progressively moved beyond the standard variables related to deterrence to explain compliance, to assess the impact of other factors such as social norms, institutions, and positive rewards. They have also become increasingly accurate in their understanding of human behaviour, often borrowing theories and methods from psychology, political sciences, and behavioural economics.

While all this has substantially improved the realism of tax experiments, those reviewed in Section 2 still suffer from artificiality and can be of limited use for policy. It has only been with the move of tax experiments to the field that the literature has established a solid connection to real-world dynamics. By involving real taxpayers and revenue authorities, large-scale field experiments have considerable potential to provide policy-relevant evidence. Econometric methods have had to become more sophisticated and creative to face the challenges of real-world data, most notably regarding observability of the relevant compliance variables. Moreover, recent studies have tried to uncover the mechanisms rather than the simple effects of interventions to increase compliance.

This tremendous progress, and particularly the emergence of large-scale field experiments, has tackled the most important drawbacks of the tax experiment literature, making it more relevant for policymaking. Clearly, social, cultural and institutional factors matter for compliance and it is likely that specific policy options do not work in all contexts. The results from an experiment conducted in one context may not be valid in other contexts, where important factors may differ (for example, the existing level of tax compliance, the revenue agencies' approach, historical and political elements of taxpayers' experience). Therefore, external validity remains a concern.

However, modern field tax experiments can offer useful policy indications, at least to policymakers in the country where the experiment is conducted. The focus on taxation at the national level and the use of representative samples of the whole population of taxpayers have also helped tax experiments to overcome the standard limitations of micro-level RCTs, such as external validity (within the country), policy relevance, and open questions on the feasibility of scaling up programmes (see Section 1). Of course, it should be noted that tax experiments still focus on very specific interventions and can only evaluate the treatments included in the experiment. Where there are more systemic issues in the tax system, tax experiments should not be the only source of information; they should be complemented by a broader analysis of the political economy of tax mobilisation.

There are two main policy implications that emerge from the literature reviewed in this report. The first is the potentially important role of positive rewards. Lab-based evidence as well as psychology and behavioural economics support the potential of positive rewards to improve compliance, and the ongoing project in Uruguay is an important step towards testing their validity in real life. However, more research is needed to fully assess the empirical validity of positive rewards, particularly compared with negative incentives and threats, and to make a compelling case for their role in improving tax compliance.

The second policy implication is related to information and its crucial role in ensuring compliance. When taxpayers know that it is easy for the tax authority to check for and detect irregularities, they tend to comply more. While this is surely well known among tax administrators, the experimental literature provides a solid evidence base for such common wisdom. It implies that systems like VAT and third-party reporting should be favoured over systems based on self-reporting. The digitalisation of tax records in developing countries, by making cross-checks easier and quicker, has considerable potential to increase compliance. Capacity should be built and strengthened in tax administrations to ensure that information systems can be used to maximum effect.

As far as moral and social factors are concerned, the literature does not show clear-cut results. While Hallsworth *et al*. (2014) show that moral and social factors can be effective in increasing compliance, other studies have found that they have no effect (Blumenthal *et al*. 1998; Fellner *et al.* 2013; Torgler 2004; Castro and Scartascini 2013). Moral and social factors are likely to have an important role in determining the high levels of compliance observed in many countries, despite low sanctions and audit probabilities. Survey-based evidence shows that compliance levels are indeed associated with these factors, in addition to economic considerations. However, moral appeals may not have large effects in some contexts because non-compliant taxpayers may not be very responsive to moral appeals, as they may have less trust in the state and less regard for social benefits than individual benefits. Rather than representing a conclusive result, the existing literature calls for more research on the impact of moral and social factors on compliance. More research is also needed on the positive relation between accountability and tax revenue, which can generate positive outcomes on both fronts, as shown by Paler (2013).

This study has shown that tax experiments, particularly in their modern form, can contribute not only to academic research but also to policy options. Given the momentum for tax revenue mobilisation in developing countries, there seems to be an opportunity to further expand this area of research, which is still relatively underdeveloped. Field experiments in developing countries are still rare and, indeed, totally absent in the African context. Large-scale field experiments require data availability and close collaboration with governments. In recent years, these conditions have materialised in many African countries, with the widespread modernisation of tax administrations and the high priority afforded by governments to raising domestic tax revenues. The digitalisation of tax records implies that data are, in principle, available.

Of course, challenges remain in obtaining this data, principally for two related reasons. First, tax returns contain highly sensitive information and tax authorities may be reluctant to share them with researchers. This problem can be easily solved by making returns anonymous. Common African standards on how to anonymise returns may significantly improve the availability of tax data. The second challenge is that making data available may represent a burden on already constrained tax administrations. Not only do the data need to be anonymised, but also presented in a way that makes analysis possible. National tax authorities' information systems may not be designed for such analysis. Therefore, making data available to researchers requires a significant commitment from the authority. This illustrates the crucial importance of researchers working in close collaboration with governments, which would need to be fully engaged with the research process to ensure success.

Despite these challenges, there is a considerable opportunity to expand this area of research in the African context, which would not have been possible until a few years ago. The literature offers many inputs of relevant aspects to analyse, including the more traditional deterrence and social factors, but also more innovative interventions such as positive rewards.

References

Ali, M.; Fjeldstad, O.-H. and Sjursen, I.H. (forthcoming) 'To Pay or Not to Pay? Citizens' Attitudes Towards Taxation in Kenya, Tanzania, Uganda and South Africa', *World Development*

Allingham, M.G. and Sandmo, A. (1972) 'Income Tax Evasion: A Theoretical Analysis', *Journal of Public Economics* 1.3-4: 323–38

Alm, J. and Jackson, B. (1993) 'Fiscal Exchange, Collective Decision Institutions and Tax Compliance', *Journal of Economic Behavior & Organization* 22.3: 285–303

Alm, J.; Jackson, B. and McKee, M. (1992a) 'Deterrence and Beyond: Towards a Kinder, Gentler IRS', in J. Slemrod (ed), *Why People Pay Taxes: Tax Compliance and Enforcement*, Ann Arbor MI: University of Michigan Press

Alm, J.; Jackson, B. and McKee, M. (1992b) 'Institutional Uncertainty and Taxpayer Compliance', *American Economic Review* 82.4: 1018–26

Alm, J.; McClelland, G.H. and Schulze, W.D. (1999) 'Changing the Social Norm of Tax Compliance by Voting', *KYKLOS International Review for Social Sciences* 52.2: 141–71

Alm, J.; Sanchez, I. and de Juan, A. (1995) 'Economic and Noneconomic Factors in Tax Compliance', *KYKLOS International Review for Social Sciences* 48.1: 3–18

Basu, K. (2013) *The Method of Randomization and the Role of Reasoned Intuition*, Policy Research Working Paper 6722, Washington DC: World Bank

Beck, P.J.; Davis, J.S. and Jung, W.-O. (1991) 'Experimental Evidence on Taxpayer Reporting under Uncertainty', *The Accounting Review* 66.3: 535–58

Becker, W.; Buchner, H.-J. and Sleeking, S. (1987) 'The Impact of Public Transfer Expenditures on Tax Evasion: An Experimental Approach', *Journal of Public Economics* 34.2: 243–52

Blumenthal, M.; Christian, C. and Slemrod, J. (1998) *The Determinants of Income Tax Compliance: Evidence from a Controlled Experiment in Minnesota*, NBER Working Paper 6575, Cambridge MA: National Bureau of Economic Research

Bosco, L. and Mittone, L. (1997) 'Tax Evasion and Moral Constraints: Some Experimental Evidence', *KYKLOS International Review for Social Sciences* 50.3: 297–324

Brautigam, D.; Fjeldstad, O.-H. and Moore, M. (2008) *Taxation and State-Building in Developing Countries: Capacity and Consent*, New York: Cambridge University Press

Calvet Christian, R. and Alm, J. (2014) 'Empathy, Sympathy, and Tax Compliance', *Journal of Economic Psychology* 40.C: 62–82

Castro, L. and Scartascini, C. (2013) *Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment*, Working Paper 472, Washington DC: Inter-American Development Bank

Coricelli, G.; Rusconi, E. and Villeval, M.C. (2014) ‘Tax Evasion and Emotions: An Empirical Test of Re-integrative Shaming Theory’, *Journal of Economic Psychology* 40.C: 49–61

Coricelli, G.; Joffily, M.; Montmarquette, C. and Villeval, M.C. (2010) ‘Cheating, Emotions, and Rationality: An Experiment on Tax Evasion’, *Experimental Economics* 13.2: 226–47

Cummings, R.G.; Martinez-Vazquez, J. and McKee, M. (2001) *Cross Cultural Comparisons of Tax Compliance Behavior*, Working Paper 01-3, Atlanta GA: Georgia State University, School of Policy Studies

Deaton, A. (2009) *Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development*, NBER Working Paper 14690, Cambridge MA: National Bureau of Economic Research

Dunning, T.; Monestier, F.; Piñeiro, R.; Rosenblatt, F. and Tuñón, G. (2014) *Pre-Analysis Plan, ‘Positive vs. Negative Incentives for Compliance: Evaluating a Randomized Tax Holiday’*, www.thaddunning.com/wp-content/uploads/2014/07/Engendering-Compliance_PreAnalysis-Plan.pdf (accessed 8 October 2014)

Feld, L.P.; Frey, B.S. and Torgler, B. (2006) *Rewarding Honest Taxpayers? Evidence on the Impact of Rewards from Field Experiments*, CREMA Working Paper Series 2006-16, Basel: Center for Research in Economics, Management and the Arts (CREMA)

Fellner, G.; Sausgruber, R. and Traxler, C. (2013) ‘Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information’, *Journal of the European Economic Association* 11.3: 634–60

Fjeldstad, O.-H. (2013) *Taxation and Development: A Review of Donor Support to Strengthen Tax Systems in Developing Countries*, Working Paper 2013/010, Helsinki: United Nations University – World Institute for Development Economics Research (UNU-WIDER)

Fjeldstad, O.-H. and Moore, M. (2008) ‘Tax Reform and State Building in a Globalised World’, in D. Brautigam, O.-H. Fjeldstad and M. Moore (eds), *Taxation and State-Building in Developing Countries: Capacity and Consent*, New York: Cambridge University Press

Friedland, N.; Maital, S. and Rutenberg, A. (1978) ‘A Simulation Study of Income Tax Evasion’, *Journal of Public Economics* 10.1: 107–16

Gadenne, L. (2013) ‘Tax Me, But Spend Wisely: Sources of Public Finance and Government Accountability’, mimeo, University College London and Institute for Fiscal Studies

Hallsworth, M.; List, J.A.; Metcalfe, R.D. and Vlaev, I. (2014) *The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance*, NBER Working Paper 20007, Cambridge MA: National Bureau of Economic Research

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: 651–92

Mittone, L. (1997) ‘Subjective Versus Objective Probability: Results from Seven Experiments on Fiscal Evasion’, mimeo, University of Trento

Moore, M. (2013) *Revenue Reform and Statebuilding in Anglophone Africa*, ICTD Working Paper 10, Brighton: International Centre for Tax and Development

Paler, L. (2013) 'Keeping the Public Purse: An Experiment in Windfalls, Taxes, and the Incentives to Restrain Government', *American Political Science Review* 107.4: 706–25

Pomeranz, D. (2013) *No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax*, Working Paper 13-057, Boston MA: Harvard Business School

Roth, A.E. (1995) 'Introduction to Experimental Economics', in J.H. Kagel and A.E. Roth (eds) *The Handbook of Experimental Economics,* Princeton NJ: Princeton University Press

Slemrod, J.; Blumenthal, M. and Christian, C. (2001) 'Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota', *Journal of Public Economics* 79.3: 455–83

Spicer, M. and Becker, L.A. (1980) 'Fiscal Inequity and Tax Evasion: An Experimental Approach', *National Tax Journal* 33.2: 171–5

Spicer, M.W. and Hero, R.E. (1985) 'Tax Evasion and Heuristics: A Research Note', *Journal of Public Economics* 26.2: 263–7

Spicer, M.W. and Thomas, J.E. (1982) 'Audit Probabilities and the Tax Evasion Decision: An Experimental Approach', *Journal of Economic Psychology* 2.3: 241–5

Torgler, B. (2004) 'Moral Suasion: An Alternative Tax Policy Strategy? Evidence from a Controlled Field Experiment in Switzerland', *Economics of Governance* 5.3: 235–53

Torgler, B. (2003) 'Beyond Punishment: A Tax Compliance Experiment with Taxpayers in Costa Rica', *Revista de Analisis Economico [Economic Analysis Review]* 18.1: 27–56

Torgler, B. (2002) 'Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments', *Journal of Economic Surveys* 16.5: 657–83

Waseem, M. (2013) 'Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform', mimeo, London School of Economics

Institute of
Development Studies

Brighton BN1 9RE

T +44 (0)1273 606261
F +44 (0)1273 621202
E ids@ids.ac.uk
www.ids.ac.uk