JEL classification: C91, H26, K42, L51

The Good, the Bad, and the Regulator: an Experimental Test of Two Conditional Audit Schemes

Jeremy Clark[a]
Lana Friesen[b]
Andrew Muller[c]

March 5th , 2002
reformatted, 16 April 2002

Abstract:

Conditional audit rules are designed to achieve regulatory compliance with fewer inspections than required by random auditing. A regulator places individuals into audit pools that differ in probability of audit or severity of fine, and specifies transition rules between pools. Future pool assignment is conditional on current audit results. We conduct an experiment to compare two specific schemes – Harrington's Past Compliance Targeting and Friesen's Optimal Targeting – against random auditing. We find a production possibility frontier between compliance and minimizing inspections. Optimal Targeting generates the lowest inspection rates as predicted, but random auditing the highest compliance. Past Compliance Targeting is intermediate.

[a] Department of Economics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. E-mail: j.clark@econ.canterbury.ac.nz Phone: (643) 364-2308 Fax: (643) 364-2635

[b] Commerce Division, Lincoln University, PO Box 84, Canterbury, New Zealand
E-mail: friesenl@kea.lincoln.ac.nz Phone: (643) 325-3627 Fax: (643) 325-3847

[c] Department of Economics, McMaster University, Hamilton, Ontario, Canada, L8S 4M4
E-mail: mullera@mcmaster.ca Phone: (905) 525-9140 ext. 23831 Fax: (905) 521-8232

Keywords: compliance; conditional audit rules; experiments

I. INTRODUCTION

How can a tax or environmental agency maximize the effectiveness of a limited enforcement budget? Since Becker (1968) economists have attempted to answer this question using the rational choice framework, predicting when individuals or firms will or will not comply with costly regulations. Based on this framework, economists have employed the tools of game theory to propose audit schemes to maximize compliance per inspection, or minimize inspections per compliance. More recently, economists have designed conditional audit schemes that exploit observable signals about firms or individuals, such as recent inspection record, that may be correlated with their current unobserved behaviour. Unfortunately, it has been difficult to evaluate these various audit schemes empirically because individual violators tend to conceal their actions.

While empirical tests have been rare, conditional audit schemes have been explored in a steady stream of laboratory economics experiments (as surveyed in Alm and McKee (1998)). Here subjects either earn or are given increments of income, which they are then asked to disclose in order that some of the money be deducted. Motivated from the tax compliance literature, these experiments have tested the effects on compliance of alternate fines, inspection probabilities, uncertainty as to taxable income, amnesties, tax rates, and many other variables. Relevant here, they have also tested a series of conditional audit rules, often simple rules of thumb, that make probability of inspection dependent on subject behaviour within the experiment. While always interesting, the conditional audit rules tested have not generally been explicitly derived from theory, and so make no claims to having been optimally designed.

This paper reports on a series of laboratory experiments that compare inspection and compliance rates for two "forward looking" conditional audit rules against a control rule equivalent to simple random auditing. The two mechanisms are Past-Compliance Targeting (PCT), as proposed by Harrington (1988), and Optimal Targeting (OT), as proposed by Friesen (2000). Both schemes exploit information from immediate past audits to assign individuals to one of two audit pools, which we dub the "Green" and the "Red." The Green audit pool is for "good" firms, and imposes a lower probability of audit than the Red audit pool for "bad" firms. The PCT relies on the outcomes of audits to transfer individuals in either direction between audit pools. In contrast, the OT mechanism randomly transfers individuals from the Green pool to the Red, and uses audit outcomes only to enable repentant individuals to escape the Red pool and re-enter the Green.

By using two audit pools, both mechanisms augment the punishment for non-compliance beyond the immediate fine to include the threat of greater future scrutiny. Compared to random auditing both schemes promise to require a lower frequency of inspections to achieve a desired level of compliance. However Harrington's PCT assumes the rules governing transfer between the groups, whereas in OT the transition rules are optimally derived. Thus, for a given target rate of compliance, Harrington's PCT should require fewer inspections in equilibrium than random auditing, but Friesen's OT should require fewer still.

We find that both mechanisms do indeed succeed in lowering overall inspection rates, though only

Friesen's does so significantly. However neither mechanism achieves the overall compliance rate achieved by random auditing, and Friesen's in particular is significantly lower. If the ratio of compliance over inspection rate is taken as an ordinal measure of overall efficiency, Harrington's PCT outperforms Friesen's OT, which outperforms random auditing. Publicizing the results of audits had no significant effects in the neutral setting of the experiment.

The paper is organized as follows. Section Two provides a review of conditional audit rules in the tax and environmental compliance literatures. Section Three describes the PCT and OT schemes in particular, and the design of the experiment used to compare them. Section Four describes the results of our experiments, and Section Five concludes with a discussion of our findings.

II. SQUEEZING BLOOD FROM A STONE – CHEAPLY

In Theory

Gary Becker (1968) wrote a seminal paper extending the rational choice model of the household to the domain of law enforcement. Individuals will violate a costly law if their analysis of the expected benefit of doing so outweighs the costs (getting caught). Allingham and Sandmo (1972) first applied this framework to tax evasion using a simple random audit rule. Here the probability of audit and fine for evasion became key parameters. Theorists then turned their attention to using observable information supplied by taxpayers to design more cost-efficient auditing rules. In one branch, Reinganum and Wilde (1985) use a principal agent framework to propose that agencies exploit the level of income that taxpayers report to determine whom to audit. Agencies could choose a "cut-off " level of reported income below which all individuals in a given class are audited.

In a second branch, theorists proposed that agencies exploit an individual's audit record when determining whom to audit. Rickard, Russell and Howroyd (1982) proposed that the results of a person's current audit be used to determine his probability of *back* audits. Alternatively, Landsberger and Meilijson (1982) proposed that a person's current audit be used to determine his probability of *future* audits. The latter authors showed that by targeting audits according to current audit outcomes, agencies could increase tax revenue for a given enforcement budget and fine scheme. While Landsberger and Meilijson demonstrated that schemes exist that are more cost-effective than random auditing, they did not seek to identify an optimal rule.

Greenberg (1984) extended Landsberger and Meilijson's analysis using a repeated game-theoretic approach. Greenberg proposed three audit groups, G1, G2, and G3, each with its own probability of inspection, and rules for transition that were conditional on audit status. Individuals caught under-reporting income in G1 would be transferred to G2. Those caught similarly cheating in G2 would be transferred to G3, or if found in compliance, transferred back to G1. The third group serves as the ultimate deterrent, threatening certain audit and no chance of escape once entered. Greenberg showed that tax evasion could be greatly reduced from that predicted under random auditing. For the tax agency need only audit with tiny probability in G2 because of the great future cost to tax filers threatened by transfer to G3. Unfortunately, the spectacular gains in compliance derived partly from the assumptions of zero discounting and unconstrained fine levels.

Harrington (1988) extended the application of Greenberg's forward looking conditional audit rule to the realm of environmental regulation. Forward based rules are particularly relevant for environmental regulation, because back audits of past compliance might not be feasible. Harrington reduced the decision space of the regulated firms to "comply" or "violate," but also incorporated a positive discount rate and constraints on maximum fine size. These changes combined to reduce the potential efficiency gains from conditional auditing. Nonetheless, a given compliance rate could still be achieved with fewer inspections than would be needed with random auditing, even when the number of audit groups was reduced from three to two.

Harrington solved endogenously for the inspection probabilities and fines that would minimize inspections needed to induce a desired overall compliance rate. The rules governing transition between the audit groups, however, were assumed rather than solved.[1] More recently, Friesen (2000) retained Harrington's binary decision approach for two groups, but made the entire structure of transition rules between groups endogenous. This was shown to further reduce the inspection rate needed for a desired level of compliance. While Friesen's "Optimal Targeting" (OT) scheme claims greater cost efficiency than Harrington's "Past-Compliance Targeting" (PCT) scheme in equilibrium, it holds for a narrower range of parameters. In particular, while neither scheme can be used when pursuing 100% compliance, PCT can be used for higher target rates than can OT.[2] To our knowledge, the PCT and OT schemes are the only two forward-looking conditional audit rules to be formally derived.

<u>In Practice.</u>

A limited number of empirical tests of tax compliance mechanisms have been conducted, using for example, the Taxpayer Compliance Measurement Program of the United States' Internal Revenue Service. The limitations of these studies are discussed by Hessing et al. (1992) and Alm and McKee (1998). They must by necessity combine self-reported surveys and official records, but are hampered by low sample response and attrition, and confidentiality restrictions. To our knowledge no empirical tests of the efficiency of conditional audit mechanisms have been conducted, though the revealed practice of enforcement agencies has been examined by Helland (1998).[3]

Experimental tests of tax compliance began with Friedland, Maital and Rutenberg (1978). Early papers concentrated on the effects of parameters identified by the static tax compliance model. Thus, the size of fines and probability of audit have been widely examined, and found to have some effect on compliance, though less than predicted (Beck, Davis and Jung (1991), Alm, McClelland and Schulze (1992), Alm, Jackson and McKee (1992a)). Other variables have also been considered, such as income uncertainty and risk preference (Beck, Davis and Jung (1991)), the purpose of the money collected (Alm, McClelland and Schulze (1992)), and tax amnesties (Alm, McKee and Beck (1990)). A common finding to emerge from these studies was that individuals tend to comply far more often than would be predicted in a selfish game theoretic sense, though the qualitative effects of treatments variables are usually in the direction predicted by the rational choice model. "Over compliance" in experiments, as in empirical studies, has been attributed to social norms (Alm, Sanchez and de Juan (1995)), and to people's tendencies to overweight small probability events such as tax audits (Alm, McClelland and Schulze (1992)).

Experimental tests of conditional audit rules began with Collins and Plumlee (1991), who test a "cut off" audit scheme loosely based on the principal-agent model of Reinganum and Wilde (1985). A fixed number of audits were conducted on individuals reporting the lowest incomes. Collins and Plumlee also tested a "conditional cut off" audit scheme, where individuals are first sorted into two groups according to earning ability demonstrated in a practice session. Here the individuals reporting the lowest incomes in each group were audited. Thus the authors control the aggregate probability of inspection across regimes, though subjects could not know their individual probability of audit in either cut-off scheme. Collins and Plumlee found that both cut off schemes were equally successful in reducing under-reporting relative to random audits. Risk preferences were measured in a pre-experiment questionnaire, but were not found to be significant in predicting truthful reporting.

Alm, Jackson and McKee (1992b) provided the first test of a conditional audit rule based on audit outcomes. They tested a forward looking "audit reduction" scheme, though without explicit reference to prior theory. Under Alm et al.'s scheme subjects who were audited and found in compliance would have their future probability of audit reduced from .04 to .027, and then again to .013. The audit probability would remain as is in the absence of audit, or revert to .04 if non-compliance was detected. Alm et al. found that the audit reduction scheme significantly raised compliance rates over random auditing, but not as effectively as other positive inducements, such as reward lotteries for individuals found to be in compliance. Alm et al. did not control for the number of inspections across schemes, but almost controlled for the *ex ante* probability of inspection.[4] Risk preferences were not controlled, though subjects' frequent "all or nothing" income reports lead the authors to conclude that risk-neutrality was a plausible assumption. In contrast to Collins and Plumlee, Alm et al.'s subjects always knew their probability of audit.

Finally, Alm, Cronshaw and McKee (1993) compared a cut-off, backward-looking, and forward-looking conditional audit scheme, respectively, against random auditing. Each conditional scheme was based loosely on the corresponding theories of Reinganum and Wilde (1985), Rickard, Russell and Howroyd (1982) and Greenberg (1984). Subjects inspected in a (5%) random audit were back- (or forward-) audited for two periods with certainty if found to be under-reporting income. If found to be in compliance subjects in the forward conditional scheme were spared the 5% chance of random audit for the next two periods. Alm et al. (1993) had to wrestle with the problem of comparing compliance rates across regimes with endogenously determined inspection rates. They cleverly solved this problem by running random-audit control treatments at several different levels of inspection probability. They then the compared the compliance rate observed in a conditional audit treatment against the compliance rate in the control treatment with the closest matching inspection rate.

Alm et al. (1993) found that all three conditional schemes generated compliance significantly greater than corresponding random audit rules, though the back-audit scheme induced more compliance and fewer inspections than the forward-audit. Alm et al. again simply assumed risk neutrality, and did not control for the effective discount rate as specified in the underlying theories. Subjects were in all cases informed of the audit probabilities they faced.

The experimental tests described above provide encouraging evidence that conditional audit rules can increase the cost-effectiveness of tax or regulation enforcement. In each case, however, these experiments have been based only loosely on the mechanisms proposed in theory. We turn now to our test of two formally-derived conditional audit rules.

III. PAST COMPLIANCE AND OPTIMAL TARGETING

The Mechanisms

We consider first Harrington's (1988) PCT mechanism for two audit groups. In each decision round a firm (or taxpayer) must choose whether or not to comply with a required action that costs c. Each firm is aware of being placed in one of two audit groups, which we refer to as the Green Group, or the more punitive Red Group. The probability of audit in Red, p_R, is set higher than in Green, p_G. Firms found in violation must pay a fine of F_G if audited while in Green, or of F_R if audited while in Red. After each decision round transition rules determine the group a firm will be placed in for the next period. These rules follow a Markov process, and are explained in Figure 1. Firms found in compliance while in Red are admitted back to the Green Group for the subsequent round with probability D.

FIGURE 1
Transition Rules in the PCT

Period t		Period t+1
If in Green and not audited	$\Rightarrow$	Green
If in Green and audited,		
-if found in compliance	$\Rightarrow$	Green
-if found in violation	$\Rightarrow$	Red
If in Red and not audited	$\Rightarrow$	Red
If in Red and audited,		
-if found in compliance	$\Rightarrow$	Green with Pr. D , Red with Pr. 1-D
-if round in violation	$\Rightarrow$	Red

Harrington solves for the values of ρ, p_R, p_G, F_G and F_R that minimize the number of audits that the agency must carry out to achieve a desired overall rate of compliance. This is achieved by making it in firms' interests to pursue a strategy of always complying while in Red, and never complying while in Green. Thus, to achieve an overall compliance rate of, say, 50%, the agency will choose parameters that should lead to firms spending half their decision rounds in Red and half in Green.

Friesen's (2000) OT mechanism uses a structure similar to Harrington's PCT, but with the transition rules chosen along with the previous parameters to minimize audits. As Figure 1 illustrates, Harrington imposes transition probabilities between groups of either 1 or 0 for five out of six possible cases. He solves only for D, the probability of escape to Green after being found in compliance in Red. With a total of ten parameters chosen, OT results in the transition rules described in Figure 2. In OT the optimal transition probability from the Green group is independent of compliance or audit status, but rather a fixed probability 2. With random transfer from Green to Red, there is no need for audits in Green (p_G* = 0).

FIGURE 2:

Transition Rules in the OT

Period t		Period t+1
If in Green	⇒	Green with Pr. 2, Red with Pr. 1-2
If in Red and not audited	⇒	Red

If in Red and audited,

-if found in compliance $\Rightarrow$ Green

-if found in violation $\Rightarrow$ Red

Both mechanisms specify that the optimal fine for violation is $F_G = 0$ in Green and $F_R = F_{MAX}$ in Red. For the rest, the specific audit probabilities p_G (for the PCT), p_R and transfer probabilities ρ and θ will depend on exogenous parameters: the cost of compliance, c, and target compliance rate, Z, maximum fine size, F_{MAX} and discount factor *. The exogenous and endogenous parameters used in our experiment are listed in the second and third columns of Table 1.

<u>Implementation</u>.

The implementation of Harrington's and Friesen's mechanisms requires a number of design decisions which we now describe. First, a challenge of these mechanisms is that inspection and compliance rates are jointly determined by subjects' decisions within the experiment. Thus is it not possible to control *ex ante* for, say, compliance and compare inspections rates across mechanisms. We shall thus compare inspection and compliance rates simultaneously. With the parameters chosen we predict overall compliance rates of 50% in either OT or PCT, but an inspection rate of 21.5% in PCT and 14.5% in OT.

Second, it is common to compare conditional audit rules against a control treatment of simple random auditing. We have chosen a control design that is "equivalent" to simple random auditing, following Friesen (2000) in maximizing parallelism with the twoconditional audit rules. Our control treatment retains the use of the Red and Green audit groups, but transitionbetween groups becomes purely random. Audits are not carried out in the Green group, but are in the Red group just often enough to make compliance there individually rational. As in the PCT and OT mechanisms the fine for detected violation in Red is set at the maximum, $F_R = F_{MAX}$, and the cost of compliance is c. The incentives in the control treatment are such that individuals should pursue the same strategy as before: always comply in Red and never comply in Green. The random probability of placement in the Red group is thus set equal to the overall target compliance rate, Z. We set the overall target compliance rate in the Control, 50%, to match that of the PCT and OT mechanisms, but it should require a higher inspection rate of 25%.

Table 1 about here

With these design decisions made implementation is relatively straightforward. We employ a within-subject design to maximize the power of statistical tests, as our predicted inspection rates are not that far apart. Order effects are addressed by running sessions in all possible sequences: ABC, ACB, BAC, BCA, CAB, and CBA. Neutral language is used throughout. In a given decision round of a mechanism, subjects choose between "Option A" (compliance) and "Option B" (violation) , and then face a possible audit described as "entering a random draw." Regarding risk preference, both the PCT and OT mechanisms assume risk neutrality. Rather than presume this risk preference, we attempt to induce risk neutrality by having the compliance decision made over lottery tickets (as in Davis and Holt (1993)). Each subject begins a decision round endowed with 100 "points," and the cost of compliance is the surrender of 40 such points to the experimenter. These points are used to enter a random draw for $1 each round. By surrendering 40 points a subject reduces his probability of winning the $1 draw from 100% to 60%, or by $.40 on average.

To implement a discount factor we use a probabilistic stopping rule (Davis and Holt (1993)). Thus all participants of a given session face a (1-$) probability that each real decision round will be the last for a particular mechanism. With a 90% probability of continuance after the first round subjects can expect to have 10 real rounds with each mechanism, though with a high variance.

The combination of a within-subject design, probabilistic stopping rules and payout over lottery tickets could result in a very complex environment for subjects to understand. This in turn could result in less meaningful compliance decisions. We thus take several steps to aid comprehension. First, we distribute a paper colour coded schematic diagram to each subject for each mechanism as we progress through the experiment. These are reproduced in the Appendix, absent the colour. Second, we give subjects 10 hypothetical practice rounds with each mechanism prior to its first real round. So, for example, the subjects in a session might have, in turn, 10 practice rounds of the PCT, 7 real rounds of the PCT, 10 practice rounds of the Control, 14 real rounds of the Control, 10 practice rounds of the OT, and finally 9 real rounds of the OT.

Finally, we altered a parameter within our overall design so as to make all three mechanism's optimal strategies more transparent. In particular, we calculated endogenous parameters using a cost of compliance, *c*, of 50 points for all three mechanisms, but then reduced this cost to 40 points. Why? All three mechanisms extract maximum efficiency by setting parameters so that individuals in the Red group are only *just* better off complying than violating. By reducing slightly the cost of compliance, we make individuals strictly better off by complying in Red, while still better off not complying in Green. We hope in this way to reduce decision errors, but at the expense of lowering slightly the potential efficiency of all three mechanisms. Put another way, if we had believed that indifferent subjects would always comply when in Red, we could have set inspection, audit and transfer probabilities differently for each regime so as to predict even lower inspection rates for all three mechanisms.

Publicity Effects.

A secondary treatment variable addressed in our experiment is the effect of publicizing audit results on compliance rates and cost efficiency. Real world publicity threatens firms or taxpayers with shame from being exposed in audits, but also better information on the compliance strategies being pursued by others. In the neutral language setting we adopted only the information effect can be captured. We do this by running the control, PCT and OT mechanisms in high and low information variations. In the high information treatment, subjects are informed after each decision round of the number of people who have been in the Green and Red groups, how many of these have been audited, and how many of the audited have complied or not complied. In the low information treatment all of this information is withheld. With two information levels and 6 possible mechanism orders we run 12 sessions in total.

IV. RESULTS

We ran twelve complete and one partial session of the experiment over a three month period in March to May 2001. Overall 141 subjects took part in the complete sessions (where each was exposed to all three mechanisms), and 12 took part in a partial session that unexpectedly crashed after completing Harrington's PCT. Subjects were recruited from large first and second year classes in Economics, Mathematics and Political Science at the University of Canterbury in Christchurch, New Zealand. Each session lasted between 80 and 120 minutes, and subjects earned NZ$ 22.13 on average. (The New Zealand minimum wage was updated in 2000 to $7.55/hour.)

Tables 2 and 3 provide a summary of our results for compliance and inspection rates, respectively. In all cases we take as our unit of observation each person's compliance or inspection outcomes averaged over all decision rounds under a given regime. The Mann-Whitney non-parametric test of differences in frequency distribution indicated no significant difference between high and low information treatments for any regime. Hence the results are pooled in the table and for all subsequent analysis. Note that the sample sizes differ because of the additional partial session for Harrington's PCT, and because some subjects did not experience both audit groups in a given regime.[5]

Insert Table 2 about here

Compliance Rates Observed Under Each Regime

<u>Order Effects</u>

Tests for order effects in compliance and inspections were carried out using the Kruskal-Wallis test for differences between independent samples. No order effects were found for inspection rates under any audit regime. Order effects did emerge, however, for some compliance rates under PCT and RAE. For the PCT, Table 2 shows that subjects were less likely to comply in Green, and more likely to comply in Red, the later they experienced the regime in the sequence of three. This effect was significant in both audit pools (p value$_{\text{Green}}$ = .038, p value$_{\text{Red}}$ = .020). Interestingly, the Red and Green compliance trends under PCT offset each other so that no significant order effect was found for overall compliance (p value$_{\text{Overall}}$ = .357). For the RAE, order effects were found for compliance in Green, (p value$_{\text{Green}}$ = .052), and more ambiguously in Red (p_{Red} = .070). In contrast, no order effects in compliance were observed for OT.

The order effects in compliance in PCT are suggestive of learning that comes specifically from being able to compare features across audit regimes.[6] They create potential difficulties for pooling compliance observations under this regime. Our approach will be to persevere in pooling, and to consider the implications of order effects in the final discussion.

Treatment Effects.

We can evaluate each audit regime's performance 1) absolutely, against its own theoretical prediction, or 2) relatively, against the performance of the other regimes. Absolute performance in compliance and inspection rates is reported in the final column of Tables 2 and 3, respectively, and may be compared with the predictions made in Table 1. Most results line up reasonably well with predictions. Formal comparisons are made with t tests, and significant differences are indicated on Tables 2 and 3 with asterisks.[7] Recall that subjects should never comply when in Green: the average of individual average compliance rates ranged between 6.7% and 11.6% across the three regimes. A greater divergence from theory was observed in compliance rates in the Red Group. While all subjects should comply when in Red actual compliance ranged from only 61.0% under Friesen's OT to 75.4% under Harrington's PCT and 78.2% under the Random Auditing Equivalent.

Turning to inspection rates, those generated for subjects in each audit group by the computer random number generator were generally as expected, though there seemed to be unusually many inspections *ex post* in the Green Group in Harrington's PCT, and in the Red Group in Friesen's OT. Overall inspection rates depend on subjects' compliance decisions, and through them, time spent in each audit pool.[8] The overall rates were not significantly different than predicted for Random Auditing or Harrington's PCT, but were slightly higher than predicted for Friesen's OT (17.8% rather than 14.5%), as under-complying subjects spent "too much" time in Red.

We turn next to relative comparisons across regimes. Does Harrington's PCT achieve the same level of compliance as Random Auditing, with fewer inspections? Does Friesen's OT require still fewer inspections? Descriptive comparisons can be made by moving up and down the final columns of Tables 2 and 3. Formal comparisons are made using the Wilcoxon signed rank test of paired samples, and presented in Table 4. The partial thirteenth session is necessarily omitted for these within-subject tests. Note that sample size again varies for paired observations because not all subjects experienced a given audit pool in every regime.

The comparative results of Table 4 provide arguable evidence that Harrington's PCT mechanism outperforms the control and Friesen's OT, though on different dimensions. Regarding compliance, the PCT does not achieve significantly less than Random Auditing in (1), but achieves significantly greater compliance than OT in (3), both when subjects are in the Red Group and overall. Regarding inspection rates, the PCT generated significantly fewer inspections for subjects in the Red Group than the RAE (see (1)). However, the PCT requires inspections for subjects in the Green Group while the RAE does not, so that the PCT's overall inspection rate is suggestively but not significantly lower (p value = .168, two tailed.). At the same time, the PCT does not require significantly more inspections than Friesen's OT (in (3)). A run-off comparison between OT and RAE is more problematic. As promised in theory, the OT requires significantly fewer inspections than random auditing in the Red Group as well as overall (neither regime requires inspections in the Green Group). However the OT achieved a surprisingly lower rate of compliance than the RAE in the Red Group and overall.

Insert Table 4 about here

A Signed-Rank Test Comparison of Regimes

A single cardinal measure of the compliance-inspection tradeoff between regimes is not possible because while the expected compliance cost per round is $.40, the cost to the agency per inspection is unspecified. However an ordinal measure of regime performance, average compliance over average inspection, (C/I), can be calculated for each regime at the session level and then overall.[9] For a given session j,

$$C/I_{\text{Session } j} = (1/n_{\text{Sesssion } j} \Sigma C_i) / (1/n_{\text{Session } j} \Sigma I_i) = \Sigma C_i / \Sigma I_i \qquad (1)$$

where C_i and I_i refer to the average compliance and inspection rates, respectively, of individuals within session j. Then, the overall C/I is the weighted mean of the C/I_{Session} ratios, with the weights reflecting the number of participants per session (11 or 12). This index yields a C/I for Harrington's PCT of 2.21, for

Friesen's OT of 1.94, and for the control RAE of 1.90. If the cost per inspection were equal to the cost of compliance, this would indicate that OT's inspection rate advantage over the control RAE more than compensates for its lower compliance. But Harrington's PCT would dominate both.

Auditing as a Cash Cow**.**

While our criteria for evaluating regimes has been minimising inspections subject to a compliance rate, an auxiliary objective of a regulator might be maximising fine revenues net of inspection costs.[10] Recall that subjects paid an equivalent fine of $1 if caught not complying in Red under all three regimes. Of those inspected, the average "conviction rate" across sessions was 20.2% under RAE, 37.3% under PCT, and 47.2% under OT. Once "no fine" convictions in Green in PCT are netted out, its conviction rate falls to 24.0%, which is still intermediate to the other regimes. Multiplying revenue-yielding conviction rates by overall inspection rates from Table 3, the OT generated $8.42 in revenue per hundred compliance decisions. This is almost double the fine revenue of either the PCT ($4.70) or RAE ($4.91). Since the OT also required the lowest inspection rate, it yielded by far the highest net revenue of the three regimes.[11]

V. DISCUSSION AND CONCLUSION

Conditional audit rules such as Harrington's PCT and Friesen's OT have been designed to reduce the inspections required under random auditing to achieve a given rate of compliance. These rules exploit an observable characteristic of tax payers or firms – their audit record – to assign individuals to differing audit groups. Transition rules between audit groups can augment the stick for present compliance (avoiding fines) with the carrot of future placement in preferable audit groups. By placing fewer restrictions on the optimal design of transition rules, the OT claims to require even fewer inspections than the PCT.

Conditional audit rules have attracted some criticism. The objective posed in both the PCT and OT mechanisms is to minimize inspection costs subject to compliance targets. Harford and Harrington (1991) observe that this conflicts with minimizing the private cost of compliance to firms because marginal compliance costs will differ in equilibrium for otherwise identical firms. Harford (1991) shows, however, that the net *social* gains are likely to be positive in many cases, particularly where the marginal cost of compliance is close to constant.

Our experimental test of Harrington’s PCT and Friesen’s OT against random auditing yielded several unexpected results. Informing subjects of the experiences of other subjects in the audit pools did not seem to affect compliance rates, all else held equal. This held across all three audit mechanisms. It might be interesting to see if audit publicity would have a greater effect in a non-neutral setting where moral value can be attached to compliance and violation.

More importantly, both conditional mechanisms differed from random auditing in failing to induce full compliance among subjects in the punitive Red Group. As a result, both achieved significantly lower overall compliance than predicted. This failure was especially pronounced under OT, where subjects complied only 61% of the time when in Red, as compared with 75% of the time in PCT. In contrast, subjects under random auditing equivalence complied 78% of the time when in Red, causing the overall rate of compliance in RAE to be statistically indistinguishable from that predicted.

The preceding comparison of compliance rates is muddied somewhat by order effects found in the

PCT for Red and Green Groups, and in the RAE for the Green Group. These order effects were suggestive of learning, as subjects in these cases behaved more "rationally" as they accumulated experience across regimes. Fortunately, the lack of order effects on compliance rates in OT provides a convenient benchmark. Average Red Group compliance under pooled OT was lower than that under PCT in whichever order PCT appeared. Thus, the finding that Red Group compliance was lower in OT than in PCT appears robust to order effects. By extrapolation, the more experience subjects gained across regimes, the greater this disparity would grow. Untangling order effects from Green Group compliance results is less straightforward. Average Green compliance in both PCT and RAE among "inexperienced" subjects was markedly higher than in pooled OT. This advantage disappeared and reversed, however, as subjects gained experience across regimes. Combining results, PCT Red compliance rose above that in OT, while PCT Green compliance fell below it. As a result, the overall compliance in PCT remained steadily superior to that in OT. Meanwhile, RAE's compliance advantage in Green over OT disappeared with experience, but not dramatically enough to change RAE's overall compliance advantage over OT.

Moving to inspection rates, the general under-compliance of subjects in Red in both PCT and OT had flow-on effects. While inspection rates *within* a Group were determined by fixed rules or by random draw, *overall* inspection rates were affected by where subjects were spending most of their time. Thus, subjects who under-complied in Red were detained there more often than would be predicted by theory.[12] This in turn necessitated more inspections, which dulled the savings in inspection rates predicted under OT and PCT. In particular, OT did succeed where the PCT failed in achieving significantly lower inspection rates than random auditing. But its actual reduction over PCT was less impressive than predicted and failed to reach statistical significance.

To gain a better understanding of individual decisions in the Red Group, Figure 3(b) below shows the distribution of individual compliance rates across the three mechanisms. A full 60% of subjects chose to comply 100% of the time when in the control, whereas both PCT and OT in particular induced a greater dispersion of compliance rates. In particular, almost 22% of subjects *never* complied in the Red Group under the OT regime. One explanation for the compliance results observed is that subjects were ignoring the future consequences of their compliance decisions, and focussing only on the immediate probability of audit. For Red Group compliance was highest in the regime with highest immediate audit probability (RAE), and lowest in the regime with lowest audit probability (OT). Such behaviour would be consistent with a higher effective discount rate than that induced, or with subject confusion over the conditional audit rules. It is our conjecture, however, that the OT was if anything *less* complex than the PCT, because the consequence of noncompliance in Red involved a simple, rather than compound lottery. The mystery is thus not why subjects learned to optimize in the PCT, but why they did not in the OT.

Insert Figure 3 about here.

The Distribution of Average Compliance Rates

If we compare the two conditional audit rules, they have three theoretical differences that are supposed to combine to keep compliance rates identical. Harrington's PCT threatens inhabitants of the Red Group with a higher probability of immediate audit (36.8 %) and makes the Green Group a more tempting escape destination because of its low audit rate. On the other hand, the PCT offers only a 6.2% chance of escape from Red as a reward for compliance. Friesen's OT has a lower threat of immediate audit in Red (29.0%), and makes the Green Group a less tempting escape destination because of its high rate of random transfer back to Red. But OT also offers a much higher chance of escape from Red as a reward for compliance (29.0%). It seems that at the parameter levels chosen OT's offer of a higher probability of escape from a less punitive prison to a less tempting destination was not as effective as predicted in inducing compliance.

Insert Figure 4 about here

In sum, our results suggest that enforcement agencies may face a production possibility frontier between compliance and minimized inspection, as illustrated in Figure 4. Random auditing seems most effective at achieving compliance, but at a high cost in inspection rates. Optimal Targeting seems most effective in minimizing inspection rates, but at a cost in the compliance obtained. Past Compliance Targeting achieves (almost) as much compliance as random auditing, while requiring almost as few inspections as Optimal Targeting.

ACKNOWLEDGEMENTS

The authors wish to thank Peter Heffernen, John Spraggon, Paul Walker, and two anonymous referees for their helpful comments, as well as session participants at the 2001 Economic Science Associations Meetings in Barcelona. Financial support from the Department of Economics at the University of Canterbury is also gratefully acknowledged.

REFERENCES

Alm, James, Jackson, Betty, and Michael McKee. "Estimating the determinants of taxpayer compliance with experimental data" *National Tax Journal* 45, (1), 1992a, 107-14.

___________. "Deterrence and Beyond: Toward a Kinder, Gentler IRS" in *Why People Pay Taxes*, edited by Joel Slemrod, Ann Arbor: University of Michigan Press, 1992b.

Alm, James, Cronshaw, Mark, and Michael McKee. "Tax compliance with endogenous audit selection rules" *Kyklos* 46, (1), 1993, 27-45.

Alm, James, Gary H. McClelland and William Schulze. "Why do people pay taxes?" *Journal of Public Economics* 48, 1992, 21-38.

Alm, James and Michael McKee. "Extending the lessons of laboratory experiments on tax compliance to managerial and decision economics" *Managerial and Decision Economics* 19, 1998, 259-275.

Alm, James, McKee, Michael, and Paul Beck. "Amazing grace: tax amnesties and compliance" *National Tax Journal* 43,1990, 23-27.

Alm, James, Sanchez, Isabel, and Ana de Juan. "Economic and noneconomic factors in tax compliance" *Kyklos* 48, (1), 1995, 3-18.

Allingham, Michael G. and Agnar Sandmo. "Income tax evasion: a theoretical analysis" *Journal of Public Economics,* 1, 1972, 323-338.

Beck, Paul J., Davis, Jon S., and Woon-Oh Jung. "Experimental evidence on taxpayer reporting under uncertainty" *The Accounting Review* 66, (3), 1991, 535-558.

Becker, Gary S. "Crime and punishment: an economic approach" *Journal of Political Economy,* 76, 1968, 169-217.

Collins, Julie H. and R. David Plumlee. "The taxpayer's labor and reporting decision: the effect of audit schemes" *The Accounting Review* 66, (3), 1991, 559-576.

Friedland, N., S. Maital, and A. Rutenberg. "A simulation study of income tax evasion" *Journal of Public Economics* 10, 1978, 107-16.

Friesen, Lana. "Targeting enforcement to improve compliance with environmental regulations" Working Paper, Lincoln University, November 2000.

Greenberg, Joseph. "Avoiding tax avoidance: a (repeated) game-theoretic approach" *Journal of Economic Theory* 32, 1984, 1-13.

Harford, Jon D. "Measurement Error and State-Dependent Pollution Control Enforcement", *Journal of Environmental Economics and Management* 21, 1991, 67-81.

Harford, Jon. D and Winston Harrington. "A Reconsideration of Enforcement Leverage When Penalties are Restricted", *Journal of Public Economics* 45, 1991, 391-395.

Harrington, Winston. "Enforcement leverage when penalties are restricted" *Journal of Public*

Economics 37, 1988, 29-53.

Helland, Eric. "The Enforcement of Pollution Control Laws: Inspections, Violations, and Self-Reporting" *Review of Economics and Statistics*, 80, (1), 1998, 141-53

Hessing, Dick J., Elferrs, Henk, Robben, Henry S.J., and Paul Webley. "Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies" in *Why People Pay Taxes*, edited by Joel Slemrod, Ann Arbor: University of Michigan Press, 1992.

Landsberger, Michael and Isaac Meilijson. "Incentive generating state dependent penalty system" *Journal of Public Economics* 19, 1982, 333-352.

Reinganum, Jennifer, and Louis Wilde. "Income tax compliance in a principal-agent framework" *Journal of Public Economics* 26, 1985, 1-18.

Rickard, J.A., Russell, A.M., and Howroyd, T.D. "A tax evasion model with allowance for retroactive penalties" *Economic Record* 58, 1982, 379-385.

Table 1
Parameters Used in Experiment

	Random Audit Equivalent (RAE)	Past Compliance Targeting (PCT)	Optimal Targeting (OT)
Exogenous Parameters			
Endowment	100 points	100 points	100 points
c	40 points	40 points	40 points
F_{MAX}	100 points	100 points	100 points
Target Z	0.5	0.5	0.5
β	0.9	0.9	0.9
Enogenous Parameters			
p_G	0	0.062296	0
p_R	0.5	0.367851	0.290173
ρ	-	0.169350	-
θ	-	-	0.709827
Equilibrium Predictions			
Compliance | Green	0.0	0.0	0.0
Compliance | Red	1.0	1.0	1.0
Overall Compliance Z	.5	.5	.5
Fraction of Rounds in Green	.5	.5	.5
Inspection Rate | Green	-	p_G	-
Inspection Rate | Red	p_R	p_R	p_R
Overall Inspection Rate	0.25	0.2150734	0.1450864

Table 2
Compliance Rates Observed Under Each Regime

		Order of Presentation			
		First	Second	Third	Pooled
A. When in Green Group					
Random Audit Equivalent	Mean	0.167	0.072	0.029	0.089**
	N	41	43	41	125
Past Compliance Targeting	Mean	0.175	0.108	0.062	0.117**
	N	40	32	36	108
Optimal Targeting	Mean	0.074	0.037	0.097	0.067**
	N	41	39	31	111
3 Combined	Mean	0.138	0.070	0.059	
	N	122	114	108	
B. When in Red Group					
Random Audit Equivalent	Mean	0.751	0.843	0.754	0.782**
	N	41	39	41	121
Past Compliance Targeting	Mean	0.691	0.749	0.845	0.754**
	N	41	27	30	98
Optimal Targeting	Mean	0.588	0.635	0.608	0.610**
	N	45	41	34	120
3 Combined	Mean	0.674	0.740	0.733	
	N	127	107	105	
C. Overall Compliance Rates					
Random Audit Equivalent	Mean	0.465	0.442	0.430	0.446
	N	47	47	47	141
Past Compliance Targeting	Mean	0.428	0.363	0.374	0.391**
	N	59	47	47	153
Optimal Targeting	Mean	0.336	0.356	0.303	0.332**
	N	47	47	47	141
3 Combined	Mean	0.411	0.387	0.369	
	N	153	141	141	

** For pooled results, indicates signicant difference from theoretical prediction at the 1% level. T-tests calculated on SPSS Version 10.0

Table 3
Inspection Rates Observed Under Each Regime

		Order of Presentation			
		First	Second	Third	Pooled
A. When in Green Group					
Random Audit Equivalent	Mean	-	-	-	-
	N				
Past Compliance Targeting	Mean	.0108	0.074	0.135	0.105*
	N	40	32	36	108
Optimal Targeting	Mean	-	-	-	-
	N				
3 Combined	Mean	-	-	-	-
	N				
B. When in Red Group					
Random Audit Equivalent	Mean	.508	.493	.478	0.493
	N	41	39	41	121
Past Compliance Targeting	Mean	0.368	0.445	0.400	0.399
	N	41	27	30	98
Optimal Targeting	Mean	0.351	0.304	0.412	0.352*
	N	45	41	34	120
3 Combined	Mean	0.407	0.408	0.434	
	N	127	107	105	
C. Overall Compliance Rates					
Random Audit Equivalent	Mean	0.261	0.243	0.227	0.244
	N	47	47	47	141
Past Compliance Targeting	Mean	0.207	0.194	0.188	0.197
	N	59	47	47	153
Optimal Targeting	Mean	0.190	0.161	0.184	0.178*
	N	47	47	47	141
3 Combined	Mean	0.218	0.199	0.200	
	N	153	141	141	

*.** For pooled results, indicates significant difference from theoretical prediction at the 5% and 1% level respectively. T-tests calculated on SPSS Version 10.0

Table 4
A Signed-Rank Test Comparison of Regimes

	PCT-RAE	OT - RAE	OT - PCT
	(1)	(2)	(3)
Compliance Rate | Green	1.056 (N=87)	-0.665 (N=97)	-0.309 (N=73)
Compliance Rate | Red	-1.053 (N=76)	-3.688*** (N=102)	-1.909* (N=74)
Overall Compliance	-1.107 (N=141)	-3.024*** (N=141)	-1.942** (N=141)
Fraction of Rounds in Green	0.730 (N=141)	-1.865* (N=141)	-2.028* (N=141)
Inspection Rate | Green	-	-	-
Inspection Rate| Red	-1.915* (N=76)	-3.720*** (N=102)	-1.217 (N=47)
Overall Inspection	-1.378 (N=141	-2.505** (N=141)	-0.589 {N=141)

[a] No Inspections were carried out in RAE or OT

*,**,*** denote significance at the 10%. 5%, and 1% levels, respectively.

Calculated on SPSS Version 10.0.

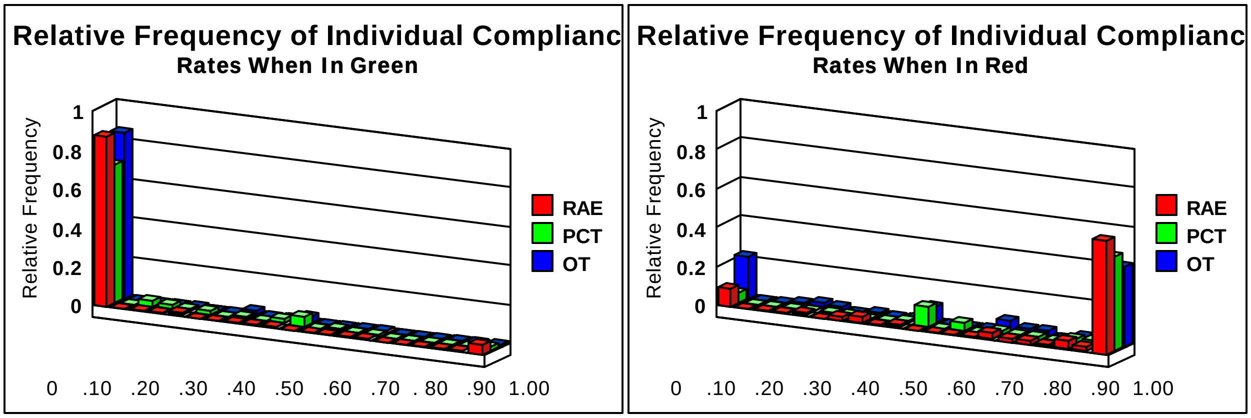

Figure 3. The Distribution of Average Compliance Rates

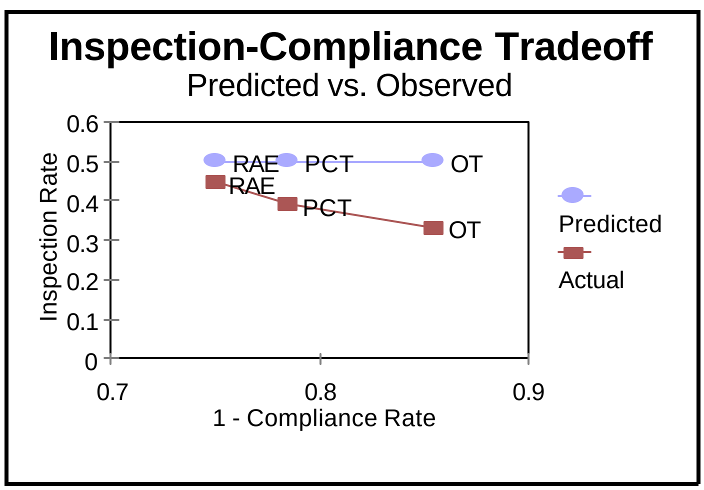

Figure 4. The Inspection -Compliance Tradeoff

NOTES

1. Where his transition rules required a probability of transfer, however, these were set optimally. For example, Harrington assumed that those audited in group 1 should be transferred with certainty if found in violation, and kept in group 1 with certainty if found in compliance. But those who were audited in group 2 and found in compliance should have only a probability of escape back to group 1. Given the rules he assumed for transition, this probability was set optimally.

2. The cost advantage of either rule decreases as the desired compliance rate rises. When desired compliance increases beyond a critical upper bound, both schemes would need to induce compliance in both audit groups rather than just the second. When this happens inspection costs become no cheaper than under random auditing.

3. Eric Helland (1998) has used data from the American pulp and paper industry to test whether environmental regulators are behaving according to the predictions of Harrington's (1988) model. He finds that indeed, following a discovered violation, firms experience a one or two quarter period of more frequent inspections. However the basis of return to low enforcement is found to be self-reported violations, rather than demonstrated compliance.

4. Alm et al. set the probability of audit at .04 in every scheme, but it may fall below this level in the conditional audit reduction scheme.

5. Half of subjects were placed in the Red and Green groups for the first real round of each mechanism, and they were informed of this. Placement was varied across mechanisms so that no subject was consistently placed in one group, and every possible sequence was equally represented.

6. Tests for order effects and learning *within* a regime are difficult to make, as subjects in different sessions experienced a given regime or audit pool for very different numbers of real rounds. (All experienced 10 practice rounds with each regime.) This is why our unit of observation is the average behaviour of a subject under a given regime and audit pool.

7. While the distribution generating the underlying compliance or inspection observations for these variables may not be normal, the sample mean for each individual's inspection or compliance rate should be distributed normally in large samples. Hence the t test is appropriate.

8. Subjects were predicted to spend 50% of rounds in the Red Group for all three regimes. In fact they spent 49.7% of rounds in Red under RAE, 46.4% in Red under PCT, but 56.8% in Red under OT. Only the OT difference is significant at the 5% level, and was caused by under-compliance.

9. *C/I* ratios may also be constructed at the individual level, but are undefined for individuals who are never inspected.

10. We thank an anonymous reviewer for suggesting this comparison.

11. A comparison of the *net* revenues generated by the PCT and RAE is complicated by the fact that the RAE generates higher fine revenues, but also requires more inspections.

12. The unusually high number of inspections generated randomly in Red under OT would not in itself detain subjects in Red. Given that subjects were not complying, a low or average inspection rate would also have detained them there.