Evaluation of the Introduction of the Makinson Incentive Scheme in HM Customs and Excise

Simon Burgess

Carol Propper

Marisa Ratto

Stephanie von Hinke Kessler Scholder

Emma Tominey

Centre for Market and Public Organisation

University of Bristol

October 2005

Executive Summary

This report uses economic theory and statistical analysis to provide an evaluation of the introduction into HM Customs and Excise of the team bonus aspects of the John Makinson report "incentives for change" scheme.

We investigate the effect of the introduction of an incentive scheme in two of the teams in which the scheme was implemented, named here as Team 1 and Team 2. Both are VAT division teams. The scheme provided financial incentives for greater output of some of the activities of these teams. Financial rewards were given for greater work on trader audit (events and yield) of three of the twelve VAT trader groups. We know that one of the two teams that we analyse here qualified for the bonus whilst the other did not.

We analyse whether performance on the incentivised tasks, which are total monetary yield and events (trader specific activity required to assure compliance), increases or decreases as a result of the scheme. The specific questions we address are:

- Does the scheme induce greater output?
- Do both yield and time spent on trader group activities increase?
- How do outcomes differ across the two teams in the scheme?
- How do the teams increase their output?

To address these questions, we use the fact that the scheme is designed with a control group, named here as the Control Team, and that we have data before and during the scheme. Our approach examines the effect of the scheme, netting out factors that may affect performance over and above the impact of the scheme. These are differences in staffing, local market conditions, and differences in dealing with the specific trader group. Once these are controlled for we find:

- Both incentivised teams, on aggregate, spend more time and recover more yield on all trader groups than the control team. However, Team 2 performs better than Team 1 on incentivised trader groups relative to non-incentivised trader groups.
- The analysis at officer level suggests that Team 2 officers increase their time spent on incentivised trader groups by 38 hours in the incentivised period compared to the non-incentivised period, holding all other variables fixed. On the contrary, officers in Team 1 spend 5 hours less on incentivised trader groups in the incentivised period compared to the pre-incentivised period, holding all other variables fixed.

2

- At officer level, for the total yield outcome, the scheme has a similar impact in both incentivised teams. In particular, officers in Team 1 increase total yield on average by £25,800 during the scheme while officers in Team 2 increase total yield by £29,800. Officers in team 1 increase yields per hour on an incentivised trader group on average by £71 in the nine months of the incentive scheme, whereas the increase in yields per hour on an incentivised trader group for officers in team 2 is only £27 on average, though not statistically significant. This contrasts with the results at team level.

- There is some evidence of strategic task assignment in Team 2. We find that in all three teams, efficient officers (we define these as the officers who were in the upper 25% of the productivity distribution in the period before the incentive scheme) increase the time on incentivised trader groups whilst the scheme was in operation. But results are statistically significant only for Team 1 and Team 2. We find a substantial difference in the behaviour of efficient officers in the two treated teams. Efficient officers in Team 2 substantially increase the time spent on incentivised trader groups (an average increase of 276 hours in the nine months of the scheme versus 99 hours for the other officers). They also spend significantly less time on non-incentivised trader groups than the other officers. Efficient officers in Team 1 increase their time spent on non-incentivised trader groups, though less than the other officers (an average increase of 29 hours for efficient officers versus 54 hours for the other officers). This suggests that efficient officers in Team 2 swap their time from non-incentivised trader groups to incentivised trader groups after the introduction of the scheme, whereas efficient officers in Team 1 do not. Hence the team strategy on how to allocate officers to trader groups differs across the two incentivised teams. We can therefore conciliate the difference in the response to the scheme at team level and the response at officer level: all officers in both incentivised teams work harder in the period of the incentive scheme, but Team 2 also works strategically smarter, as a team, moving efficient officers to tasks where time target counted.

Contents

Signposting the Report

This report first briefly documents the nature of the Team Based Incentive Scheme and the relevant organisational structure of HM Customs and Excise[1] at the time of its introduction. Second, we set out what economic analysis tells us of the likely impact of the scheme. Thirdly, we present the empirical analysis and discuss our main results. Finally, we provide a summary of the results. Appendices give further details on the theory, data and statistical techniques.

Clicking on the hyperlinks below will take you straight to the relevant sections:

The Incentive Scheme

What impact might the scheme have: the predictions for HM Customs and Excise from economic theory

Empirical Analysis

Summary of Findings

Appendix A: A brief review of the economic theory of team incentives

Appendix B: Statistical Approach

The Incentive Scheme

The Makinson approach

The team-based incentive scheme designed for HM Customs and Excise is a trial scheme, set as part of a programme to improve efficiency and productivity in the public sector. Similar incentive schemes have been designed for Child Support Agency and Jobcentre Plus. The idea of piloting a team-based incentive scheme in public agencies dates back to the Makinson report "Incentives for change" (2000). John Makinson (then Group Finance Director of Pearson plc) was recruited to the Government; Public Service Productivity Panel to analyse how performance-based incentives operated in four public organisations (Benefits Agency, Employment Service, HM Custom and Excise, and Inland Revenue) and how they might be improved. In his report, particular emphasis is placed on the use of team-based incentives, with the view that teamwork better reflects the way in which most public servants actually work. In particular, the reasons for adopting team-based rewards, according to the document, are as follows:

---

[1] On 18th April 2005, H.M C&E merged with IR to create HM Revenue & Customs.

- "The public service ethos stresses the importance of collective rather than individual achievement. Team-based rewards would, if properly executed, reinforce this positive affinity and motivation.

- Team based rewards are more capable of measurement. There is an abundance of benchmark data available in each agency on the relative performance of individual offices on a variety of measures.

- Team rewards would address concerns that individual performance measurement reflects biases against women, ethnic minorities and part-time workers.

- It will be easier to integrate pay incentive with non pay-based recognition in a team framework. The "office of the month", rewarded by a benefit in which everyone can share, is a less divisive idea than the "employee of the month". It also provides an opportunity to recognise team behaviour.

- Team rewards foster a spirit of internal competition between offices, which is more productive than internal competition within an office"[2].

The Makinson report makes some recommendations on how incentives should be designed and distributed. In particular:

- Every member of staff should have a bonus opportunity representing at least 5% of base salary.

- Incentives should relate to targets already embodied in the Public Service agreements (PSA) of the respective agencies - this is to ensure that incentives reinforce the strategic objectives of the organisation - and no employee should be judged on a wide range of targets.

- As a rule of thumb, five targets should be the maximum for junior grades and eight targets the limit for more senior staff.

- Team-based incentives should in general relate to the performance of an individual office. The relative performance of offices and districts should be made widely available within each agency.

- Incentives should be funded largely from improved productivity. The entitlement of individual agencies should be based on their overall performance. Extra funding for

---

[2] Makinson (2000), p. 17.

performance incentives should be released not just for making cost savings, but also for better than targeted service delivery and overall performance.

As we shall see, the designers of the incentive scheme at HM Customs and Excise have followed some of these recommendations.

Brief outline of the structure of HM Customs and Excise

Customs & Excise administers or enforces controls on the international movement of goods and is responsible for a wide range of Excise duties and other indirect taxes such as VAT. The agency went through a major reorganisation in April 2001, when all its activities were organised into one of two core businesses, Business Services and Taxes and Law Enforcement, which have separate regional management arrangements. Business Services and Taxes comprises all business taxes, the facilitation and information services of Customs, international movements and trade services. It focuses on legitimate business and has approximately 13,000 members of staff. Some of its services are delivered regionally and some managed centrally.

Law Enforcement comprises all investigation, intelligence and detection activities of the department. It focuses on fraud or other regulatory breaches where business activity, if it exists, is incidental to the main purposes of the activity. It has approximately 8,000 members of staff.

The incentive scheme affects only Business Services and Taxes, which is organised as follows:

Business Services and Taxes

Regional Business Services	*Large Business Group*	Policy	Business Design
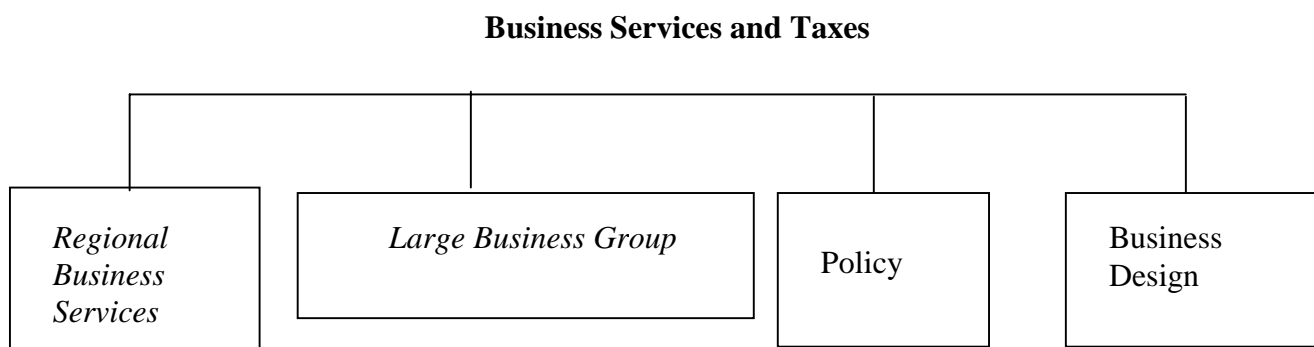
The trial sites were selected from the Regional Business Services (RBS) and Large Business Group (LBG).

The scheme design

This is a second trial implemented at Customs and Excise, which ran from April 2002 until December 2002[3].

The Teams

There are 12 teams, consisting of trial sites in RBS and LBG. These are:

- VAT Assurance Division, Solent, RBS South

- VAT Assurance Division, North East, RBS Central

- Excise Assurance, RBS Scotland

- Excise Assurance, RBS North

- International Trade Assurance, RBS North

- International Trade SME Assurance, RBS Scotland

- Risk Team (Operational Analysis Division), RBS North

- Risk Team, RBS Northern Ireland

- National Function – Registration Unit

- National Function – Debt Management Unit

- National Function - Nation Advice Service

- LBG Scotland and Northern Ireland

Note that the trial teams, apart from those defined at national level and LBG for Scotland and Northern Ireland, are paired. This is because the scheme designers adopted two different reward schemes for each team in a pair: in one team the bonus paid is equal across all team members, whereas in the other team the bonus varies according to the officer's job band. For each pair, a blind control team was chosen after the trial finished. The staff in the control team therefore never knew they were monitored as the control group for this incentive scheme.

For each team, there is a senior (band 11) manager responsible for the division. Staff in the team are located in different offices and targets are devolved by the divisional manager to office (band 9/10) managers. For bonus payment, performance is evaluated at the level of the whole division: if office managers do not reach their targets, it is only a management issue. The team (division) will

---

[3] The first team bonus trial scheme ran from July 2000 until 31st March 2001, and was implemented in six sites of the organisation.

get the bonus even if one or more locations do not reach their target, provided the others do or over-hit theirs, producing an overachievement as a whole.

We analyse the impact of the scheme in two trial teams: the VAT Assurance Division, Solent, RBS South (Team 1), and the VAT Assurance Division, North East, RBS Central- East Midlands (Team 2). We select this specific pair of teams, as their activities are more straightforward to quantify and analyse.

Team 1 comprises 154 officers, located in three offices: Poole, Southampton and Portsmouth. Team 2 has 158 officers, located in six offices: Grimsby, Lincoln, Leicester, Nottingham, Peterborough, King's Lynn.

The two trial teams for VAT assurance were chosen on the basis that they were the most stable teams, as all others went through a major restructuring during the trial. Moreover, these two teams were not subject to the previous team trial scheme[4]. The control team picked by HM Customs and Excise comprises 281 officers. The main reason for selecting this particular control group is that the managers in that region already knew about the scheme. The control team is the BS North, located in York and Middlesborough.

The targets

The incentive targets for each team are selected from their delivery agreement. Each incentive target consists of a baseline target (this is the delivery agreement target) plus a 'stretch', set at 5% of the baseline target. For the two teams we examine, the targets consist of 1) meeting a specific number of audits, or more precisely, assurance events to high risk traders, namely new registration and exceptional risk trader groups, and of 2) increasing the amount of yield or tax revenue for new registration, large traders and exceptional risk trader groups. Each team has five incentive targets. The following table summarises the targets set for the two teams.

---

[4] The idea of the organisation was to give as many different members of staff as possible the opportunity to be in a pilot scheme.
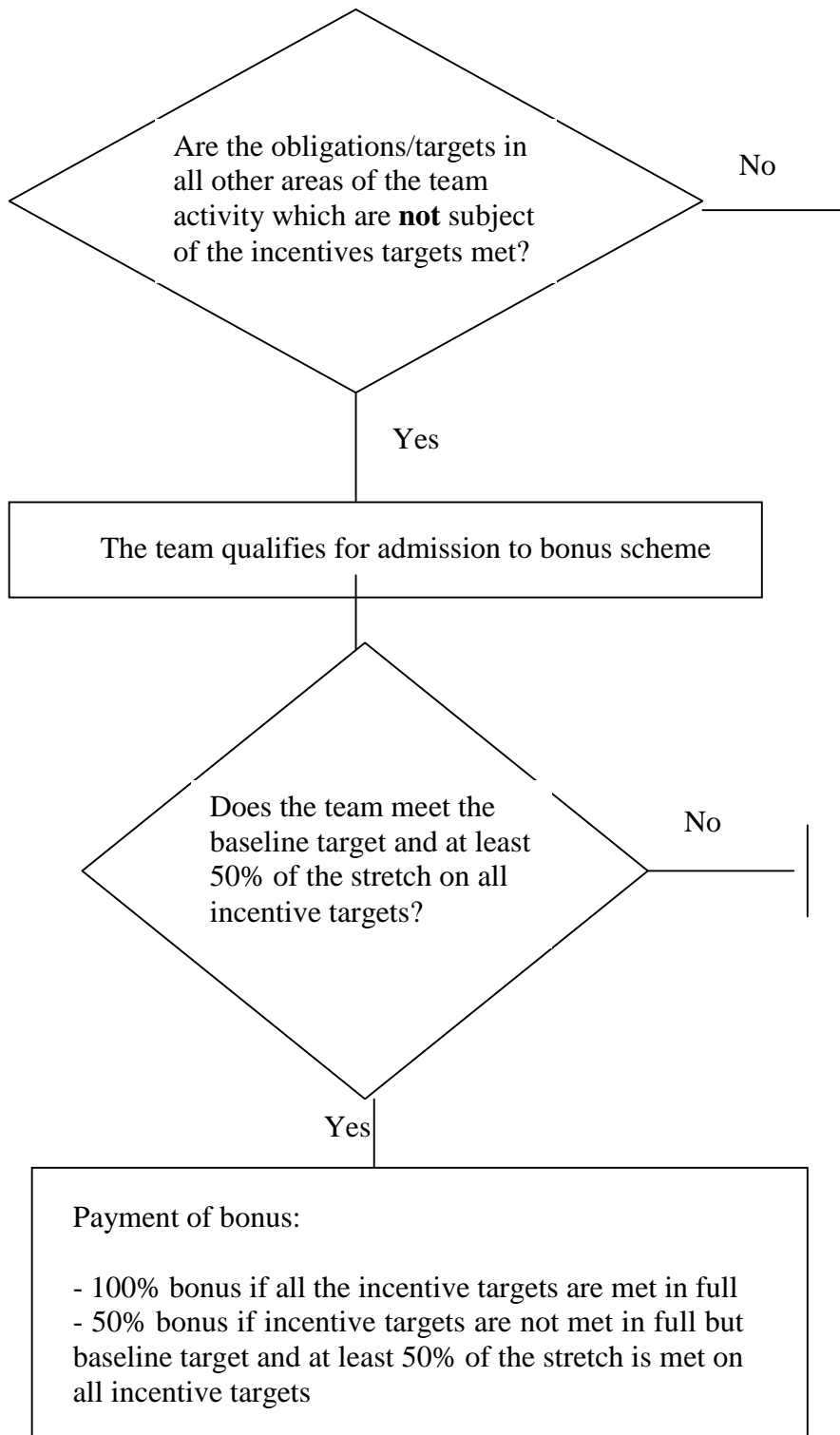
Incentive target per team (Delivery Agreement target + 5% stretch)	Team 1 (Solent)	Team 2 (North East)
Completing a minimum of **events** for:		
- New registrations trader group	1008 events	1192 events
- Exceptional risk trader group	887 events	1140 events
Achieving a minimum of **yield** for:		
- New registrations trader group	£ 7.82m	£ 10.62m
- Exceptional risk trader group	£16.18m	£ 19.34
- 999 (large) traders trader group	£ 1.90m	£ 3.60

The Bonus structure

A team potentially qualifies for the bonus after meeting all its other business objectives and targets as set out in the Delivery Agreement 2002-2003, which are not subject to the incentive payments. The bonus is then paid on achievement of the incentive targets. The maximum bonus will be paid if the team achieves all the incentive targets in full. In cases where a team fails to meet all incentive targets in full but meets the baseline target and at least 50% of the stretch on all of incentive targets, the bonus will be 50% of the total bonus. In cases where the achievement on any incentive target falls below 50% of the stretch no bonus is payable.

At the level of a single officer, an individual qualifies for a team bonus conditional on having worked a minimum qualifying period of 89 days through the duration of the trial. The amount of any bonus due is pro-rated according to the number of days served as part of the trial team. For the assessment of performance against the targets on yields, actual revenues from investigations on each trader group are capped to £2.5m. The rationale behind this is to avoid the possibility of hitting the targets thanks to one big entry. In our analysis we ignore this capping, as we do not aim at ascertaining whether or not teams hit their targets. Our aim is to investigate whether and how teams changed their behaviour after the introduction of the incentive scheme. In any case, as we shall consider, our results are not affected by the revenue capping.

The following diagram summarises the bonus award process.

Are the obligations/targets in all other areas of the team activity which are **not** subject of the incentives targets met?

No

Yes

The team qualifies for admission to bonus scheme

Does the team meet the baseline target and at least 50% of the stretch on all incentive targets?

No

Yes

Payment of bonus:

- 100% bonus if all the incentive targets are met in full
- 50% bonus if incentive targets are not met in full but baseline target and at least 50% of the stretch is met on all incentive targets

The bonus payable is calculated in two different ways for the two teams. The bonus in Team 1 varies in size across job bands. For example, for somebody working full time and the maximum days in trial, a band 4 officer would get a bonus of £550, a band 6 would get £ 720 and a band 11 manager would get £1520. For Team 2, the bonus is of a flat rate, equal across job bands, the amount of which is £740. No additional payments are made for performance above the level of the incentive target.

What impact might the scheme have: predictions for the HM Customs and Excise scheme from economic theory

The design of an optimal incentive scheme is a complex matter. The measurability of performance, the size of the team, and the multi-dimensionality of tasks are all elements to be considered. In Appendix A we briefly review what theory has to say about incentives in teams. This section uses this analysis to put forward hypotheses as to what might be the effect of the Team Based Incentive scheme in HM Custom and Excise.

The incentives in the HM Custom and Excise scheme

The nature of tasks in VAT assurance divisions

Officers within the VAT assurance divisions are typically engaged in multiple activities. In Figure 1, we illustrate all the potential activities that an officer can be engaged in. The activities (and trader groups), which are subject to the incentive scheme, are highlighted. Time spent on each activity is recorded by each officer on a weekly basis. For the trader audit activity, this is specified by trader group. The amount of yield recovered for each trader group is also recorded. There are two main tasks within trader audit activity: examination of trader's records and recovering any unpaid tax. The impact of the officer's effort on the outcome of these two tasks is quite different. For the first task, the officer's effort directly maps onto the outcome (a visit, and associated activity which make up an event). But for the second task, the amount of yield recovered is not only determined by the officer's effort and ability, but also by the intrinsic difficulty of each case and the uncertainty surrounding any debt recovery.

<u>The implications for the operation of the HM Customs and Excise incentive scheme</u>

Given this, what do we expect from the incentive scheme? We focus on which targets are chosen and how the reward system and the teams are defined.

Targets

There are two targets to reach: number of events and yield. However, officers cannot neglect the other activities; overall performance is checked, and to qualify for the bonus, teams must also meet their other delivery agreement targets.

The two outcomes, events and yield, are measurable with a very good degree of precision: they are both quantifiable. They are also measured at individual level, thus the performance measure is precise and disaggregated. However, the extent to which individual effort affects final output is very different for these two outcomes. For the number of events, just the action of allocating time to trader group activity is recorded as the outcome: the closure of audit activity to a client is the outcome, meaning there is perfect mapping of individual effort to output. However, for the yield outcome, the fact that the officer puts lots of effort into investigating a case does not necessarily mean achieving additional yield. External circumstances affect the final output more for the yield outcome than for the number of events. The mapping between individual effort and output is not so precise as in the case of events. The variance of the output distribution for the yield outcome is higher than the variance for the event outcome. Given that the two outcomes are rewarded in the same way, we expect an officer to be more willing to increase effort for the event outcome than for the yield outcome.

The design of the bonus

For Team 1 the bonus depends on the job band, it is roughly a fixed percentage of salary. This may induce some fairness considerations among the team members in Team 1: some officers (typically in job band 6 and 7) may be more directly involved in the visit and collection of yield than others. These workers would get a lower reward for reaching the target than other team members who are less directly involved in the production process, but are in a higher job band. This does not apply to Team 2, where the bonus is equal across all team members.

Moreover, the bonus is paid if the target is reached. Additional performance above the target is not rewarded. This may induce officers to increase their effort just to hit the target, or if performance is already close to the target, not to increase their effort, as the target is going to be easily met.

Teams

An important characteristic of the scheme is the structure of the teams. These are defined at the level of a division. Setting targets and assessing performance at the level of a division creates interdependencies among the offices in the same division. The expected reward of an office for hitting the target depends on how far actual performance at divisional level is from the target. And this is determined by all offices' output. However, production occurs at the level of offices, where complementarities in production may be at work (final output may depend on the effort of different colleagues). The office would be a team à la Holmström (1982), identified by complementarities in production. Hence the structure of the team as designed in the HM Customs and Excise incentive scheme is quite complex, and results in a two-level team: natural teams (offices) are included within reward teams (divisions). It is important to distinguish the team enforcement problem at office level and at division level.

At the level of an office, the fact that individual contributions to office output are not separately observable/rewarded creates a 'free riding' problem, where individuals have incentives to let others do the work for them – if everyone behaves this way then output falls (see Holmström (1982), discussed in appendix A). We expect that the greater the number of people in an office, the more serious the free riding problem. However, we also expect peer pressure to be present in an office where colleagues are able to observe each other, which can alleviate free rider problems. Kandel and Lazear (1992) show that peer pressure can offset free riding tendencies, but the offsetting strength of peer pressure varies with unit size. The smaller the unit, the more effective is peer monitoring in reducing free-riding; Knez and Simester (2001) find evidence which confirms these predictions.

At the level of a division, the probability of hitting the target depends on the contribution of each office to aggregate output. The scheme is such that the band 11 manager is responsible for the division and devolves the targets to band 9/10 managers in the different offices. If band 9/10 managers do not reach their targets it is only a management issue. The team (division) will get the bonus even if one or more locations do not reach their targets, provided that the division as a whole achieves/exceeds the target. So there may also be possible free riding across offices, as well as within offices. The literature on voluntary contribution to public goods[5] suggest that the greater the number of agents, the more serious the free rider problem. Hence, we expect the number of offices

---

[5] See Olson (1971)

in a division to be an important aspect to consider for the evaluation of the impact of the scheme on performance. The role of the band 11 manager is crucial for enforcing the offices' optimal level of effort. In particular, the band 11 manager is in a position to coordinate and monitor the contribution of each office to overall performance and may guarantee measurement of performance and feedback on how individual offices are doing against the target set for the team.

Without this, individual offices would have no idea what their contribution to final output is and hence would not be motivated enough to increase their effort to reach the target. The number of offices is going to affect the process of coordination and monitoring, and it seems reasonable to assume that the greater the number of offices within a division, the more difficult will be to coordinate and monitor.

In the case analysed here, both incentivised teams have nearly the same number of officers, but Team 1 has three offices, whereas Team 2 has six offices. The two effects go in different directions. Although coordination and monitoring might have been more difficult for the band 11 manager in Team 2, free riding at office level could have been less problematic in Team 2 than in Team 1, as the office size in team 2 is smaller. Hence, the prediction of the effect of team and office size is not clear; it depends on what issue is more important – free riding or coordination and monitoring.

Empirical Analysis

We begin with an overview of our methodology, and then present our main results.

Methodology

Policy Evaluation

The key question is whether the outcome with the new policy is any different than it would have been without it? The major difficulty in a policy evaluation, however, is to define the counter-factual; what would have happened without the policy? In this context, the outcomes are the measures of outputs. Some assumptions have to be made to *estimate* what the outcome would have been in the absence of the policy. The best approach (in this non-experimental setting) is to use another, similar, set of organisations as a control group. The control group has to be as similar as possible to the policy group and, importantly, subject to the same general set of influences. The idea is to compare the *change* in outcome in the targeted group with the *change* in outcome for the control group. For obvious reasons, this is called a "difference in difference" approach. This is the technique we use here. We compare the change between the incentivised and the pre-incentivised period in outcomes for the treatment group (teams 1 and 2) with the change in outcome for the control group.

Our Modelling Approach

Our approach is based on economic models of production where staff can apply more or less effort to raise output. The incentive scheme is meant to raise effort and so output. Output will depend on the number of people working, on the equipment they have to deal with, and their effort. The latter is unobservable to us, but is assumed to depend on the presence of the incentive scheme. This is what we estimate here: after controlling for as many other factors as we can observe, any remaining difference between the scheme and non-scheme VAT Assurance Divisions is due to the effects of the incentive scheme itself.

To measure the impact of the incentive scheme, we examine five outcomes:
- the time spent by each officer on each activity, and – for the trader audit work – the time spent on the different trader groups (this outcome is denoted time),
- yield per officer per trader group,

- positive yield per officer per trader group,

- negative yield per officer per trader group, and

- productivity, which we calculated for each officer as the ratio of yield per trader group in a nine months period to the time spent on the trader group in the same period, weighted by the share of time spent on the trader group[6].

We do not have a direct measure for the number of events (which is the incentivised outcome), so we need to use the time spent on each trader group as a proxy for that outcome. Yield is split into positive and negative as this corresponds to the Customs and Excise categories of under- and over-declared tax.

We look at the effect of the incentive scheme on the five different outcomes at *team level*, for *specific groups* of officers and at *officer* level. The analysis only considers <u>continuous frontline</u> staff, which we identify as those officers present for the first month of observation and the last month of the scheme (<u>continuous</u>) and spending time visiting traders and/or collecting yields (<u>frontline</u>). These sum up to 129 officers for Team 1 – out of 154 total officers –, 124 for Team 2 - out of 158 total officers –, and 197 in the control team – out of 281 total officers. The rest of continuous officers are clerical workers and managers, who are not directly involved in visiting traders or collecting yields.

At *team level* we consider the change in median outcomes between the pre-incentivised and the incentivised period for the two trial teams and the control team[7]. This allows us to look at how performance changed over the two periods for all three teams separately. We then consider the difference in the change in performance between each trial team and the control team (this is the difference-in-difference estimate of the impact of the scheme). This allows us to estimate whether outcomes for all trader groups have increased or decreased with the introduction of the scheme, implicitly controlling for officer fixed effects and team fixed effects. Finally, we estimate the impact of the scheme for the incentivised and non-incentivised trader groups separately. This allows us to consider whether the scheme had a different impact on incentivised and non-incentivised activities, controlling for officer fixed effects, team fixed effects and activity fixed effects.

---

[6] For more details, see Appendix B.

We then investigate whether the staff with different characteristics respond in different ways to the scheme. We do this by matching staff in treatment teams with staff in the control team by observable characteristics (grade, age, gender). The intention is to see if different types of staff respond differently to the scheme.

Finally, we use regression analysis at officer level to estimate whether the scheme has an impact on the five definitions of output at the officer level[8]. We control for individual officer characteristics, team membership (to reflect local market conditions), and type of trader group. Finally, we look to see how changes in output might have been achieved, by looking at the allocation of efficient officers across incentivised and non-incentivised tasks.

Descriptive Statistics

<u>Time Allocation</u>

We look at time reports for each officer. Time recorded by an officer is allocated to a trader group and an activity. Time reports take the form of weekly allocation of officer hours for each trader group for each activity. All descriptive statistics are – unless stated otherwise – done using continuous frontline officers working on trader audit activity. The variable 'time' is capped at 50 hours per week, to remove outliers or errors in coding (this affects only 67/71325 observations). Table 1 shows that the mean time spent by continuous frontline officers on the trader audit activity for a trader group is 8.24 hours per week[9]. The median is lower, at 6.5 hours. There is large variability in time, as reflected in the standard deviation and the quartiles of the distribution.

Trader Groups and Work Activities

Table 2 shows, for the trader audit activity, the share of time the teams spend on each trader group. Period 1 refers to the 9 months prior to the incentive scheme, period 2 to the 9 months of the scheme and period 3 to the 3 months after the scheme was in operation. Each entry shows the share of total team time within the period spent on the particular trader group. For example, in period 1, Team 1 spent just over 41% of its time on trader audit for new registrations. 12 of the 14 trader

---

[7] We consider median outcomes because the yield series have very large outliers and this makes the mean an unreliable statistic. The patterns in the data are discussed in more detail below.

[8] This analysis assumes that there are no interdependencies in production between officers.

[9] NB: on average, continuous frontline officers in team 1, 2 and the control team visit 3.02, 3.13 and 3.04 trader groups per week respectively. The maximum number of trader groups visited is 9, 8, and 8 per week for teams 1,2 and the control respectively.

group categories are VAT trader groups. The Non-trader audit category is trader audit work which is not assigned to a particular trader group. We combined all non-VAT trader groups to form the category "non-VAT". It includes a range of groups, for example tobacco duties and international trade exports. Table 2 shows that this category takes up very little of the trial teams' time and a little more of the control team's time. All three teams spend most of their time on New Registration, High risk and Exceptional risk trader groups. If we consider the incentivised trader groups, which are highlighted in the table, the time spent on these groups differs slightly across the three teams. There is a decrease in the share of time spent on new registration in all three teams. However, this decrease is smaller for Team 2. The time spent on exceptional risk increases in all three teams, though to a greater extent in Team 2 (18 % as opposed to 9 and 10% for Team 1 and the Control Team). Among the non-incentivised trader groups it is worth noting the increase in the share of time spent on the low and medium risk trader groups in Team 1 and the Control team, which is not experienced in Team 2. Figure 2 shows time allocation over trader groups for the incentivised period. Time allocation is very similar in the three teams. New registrations and exceptional risk traders account, in the incentivised period, for 30% and 24% respectively in Team 1, 36% and 29% in Team 2 and 34% and 26% in the control team.

Table 3 examines the number of hours teams spend on each trader group for the trader audit activity. Period 2 represents the 9 months of treatment and period 1 represents the corresponding 9 months pre-treatment (i.e. April – December). This is to enable a comparison of hours before and during treatment without working in percentages of total time.

For Team 1 there is a significant increase in trader audit hours on 8 out of the 12 VAT groups and a fall in the remaining 4 groups. For Team 2, trader audit hours fall for 3 out of 12 groups and increase for 9 out of 12 groups. We find the same trend for the control group. Table 3 shows that, for the Control team, hours on trader audit activity increase between periods for 8 out of 12 VAT groups and fell for the remaining 4 groups. The biggest increase in all three teams is on time spent on exceptional risk: in Team 1 and the Control team the hours spent on this trader group double, while in Team 2 they increase by more than three times. Note also that the time spent on large traders increases only for Team 1 and Team 2[10]. The Control team on the other hand, decreases its time.

---

[10] Note that this trader groups is incentivised for yield, not for events.

Figure 3 shows time allocation over all five activities. It shows that all three teams spend more than half their time on trader audit, a significant amount of their time on non-trader work and smaller amounts of time on the other work activities. Of the 5 work activities, time on trader audit contributes to the event target, whilst time on the other activities does not.

Time Series pattern

Figure 4 shows the time series variation in total time allocation per team per month. The first point to note is the seasonality in the series. We see an upward trend in this seasonality for the second period. There are peaks in April, July, October and January – every three months. Investigating the seasonality in total time more closely, we see that the number of officers does not change in correspondence with this peak. Therefore we analyse which job band of officers are driving the peaks and whether overtime hours increase sharply during these months. We find that the peaks in the number of hours worked is driven primarily by job band 6, and to a lesser extent by job bands 7 and 8. This is not surprising, as job bands 6 and 7 account for the bulk of the workforce. Given almost fixed staff numbers within teams across months but peaks in the number of hours recorded, it seems likely that the number of overtime hours within a month increases during the peak time recording months. However, the pattern of overtime records does not reflect in a clear way the peaks in total hours recorded.

Yield

Yield is an incentivised target for three trader groups - new registrations, exceptional risk and 999 large traders. When a decision is made by HM Customs and Excise regarding visits to a particular trader, officers assign a result code to a trader group to describe the outcome - for example, over-declaration or under-declaration - and attach a monetary value to the outcome. Yield can be positive or negative. We analyse positive outcomes and negative outcomes separately, as both provide information on behaviour of officers.

Trader Groups and Work Activities

Table 1 describes positive and negative yields. The values are weekly yield per trader group, per officer. Hence, the mean positive yield collected is £301,811 for one officer in one week, which is allocated to one trader group. However, this large figure is not representative of a typical yield: the median positive yield is only £4,591. It is clear there are some very large outlier entries for yield, where officers collect a very large sum from a trader.

Figure 5 shows the positive yield across VAT trader groups as a percentage of total yield. The composition of positive yield is slightly different across the three teams. The largest share of positive yield comes from the exceptional risk trader group in all three teams; it counts for 38% in Team 1, 65% in Team 2 and 40% in the Control team. New registrations and high risk follow in order of magnitude, though their shares differ across the three teams.

Figure 6 presents the allocation of negative yield across VAT trader groups. The distribution of negative yield across VAT trader groups is pretty similar to the distribution of positive yield. Exceptional risk traders account for the largest share of negative yield; for the three teams respectively, this is 34%, 56%, and 51%. New registrations and high risk traders account for a relatively big share of the total negative yield (above 15% each). The share of negative yield from 999 large traders accounts for 22% only for Team 1 and is much smaller in the other two teams.

Time Series pattern

Unlike the time series pattern for time allocation, there is no clear evidence of seasonality in yield collection for Team 2. There is a weak seasonal pattern in the series for Team 1 and the control team, for example, positive yield in the control team peaks in July and August for both years. However, these patterns are not as strong as in the time series for time allocation.

Staffing

Table 4 details the characteristics of staff by team, distinguishing between total staff and continuous frontline staff[11]. The gender variable equals 0 for males and 1 for females. Looking at the first four columns, the total sample contains approximately 50% males and females. In Team 1 and 2 there are slightly more males compared to in the control team. The mean age of workers in the total sample varies between 43-46. There is very little variance across teams in the average age. The same is true for the mean job band, which is (approximately) job band 6. Officers have spent on average about 4 years within their job band in the total sample. In teams 1 and 2, the average is slightly higher, whereas the average in the Control team is somewhat lower. The average salary is approximately £20,000 per annum, £1,600 gross monthly (both including or excluding overtime)

---

[11] As specified earlier, these are staff employed throughout the period before and during the scheme and who are directly involved in the trader audit activity. This is the group used in the descriptive statistics and in the analysis which follows.

for all teams. The variable overtime per month confirms that very little overtime occurs within a month, in particular in the two treatment teams – only 16/60 hour (27 minutes) in Team 1 and practically no overtime in Team 2. The part-time marker shows that less than 1/5 of the total sample works part-time. In teams 1 and 2 there are fewer than in the control team; 16% and 13% respectively. Part-time workers work on average 25 hours per week. This is consistent for Team 1 and the Control team, but there is no record for Team 2. Very few sick days are taken within a month in all three teams: the number of sick days is 1.1 for Team 1, 1.5 for Team 2 and 0.9 for the Control team. The variable 'days in scheme' presents the number of days that the trial teams participated in the incentive scheme - 259 for Team 1 and 261 for Team 2. In order to qualify for a bonus, officers had to work at least 89 days. All officers in Team 1[12] and all except one in Team 2 satisfied this criterion. The average potential bonus for Team 1 was £688 and for Team 2 was £681 - approximately 3% of mean annual salary. The final 4 columns show the figures for continuous frontline staff. This is very similar to the total sample of workers for all variables.

---

[12] However, one officer from team 1 has no staff information. Therefore we do not know his/her days in scheme or potential bonus.

Results

The effect of the Makinson scheme

Team level analysis

We first look at the impact of the scheme at the level of a team, as the scheme is designed at this level.

In tables 5a-c, we calculate the difference in the median of each of the five outcomes (time, total yield, positive yield, negative yield and productivity (defined as weighted yield/time) between the incentivised period and non-incentivised period. As above, we only consider these outcomes for continuous frontline workers and for the activity *trader audit work* (activity 1), as all other categories of activity were not incentivised (see Fig 1). We calculate the median difference for all trader groups together (table 5a), only incentivised trader groups (table 5b) and only non-incentivised trader groups (table 5c)[13]. As the results in table 5a show, the median time spent on audit work across all trader groups increases for all three teams, but increases more for Team 2 than the other two teams. The median total yield and positive yield increases for the two treatment teams, but not for the control team, and so does the median productivity. The median negative yield does not change over the period.

Table 5b shows the results for all the incentivised trader groups. The results are somewhat different. Only in Team 2 does the median time spent on auditing the incentivised trader groups increase. The median total and positive yield increase in both treatment teams but substantially more in Team 2 (3 and 5 times more than in Team 1 respectively). Productivity also increases more in Team 2 than in Team 1.

Table 5c shows the results for all non-incentivised trader groups. In this case, Team 2 decreases the median time spent auditing the non-incentivised trader groups, whereas Team 1 and the control team spend more time on them. The median total and positive yield and also productivity increase only in Team 1.

---

[13] To reiterate, we examine medians to minimise the effect of the very large outliers

These results show that there is a difference in performance of the two treatment teams if compared with the Control team over the incentivised period, but Team 2 is performing much better on incentivised trader groups than Team 1 in terms of time spent, yield recovered and productivity.

Tables 5d-f show the difference in the change in performance between each trial team and the control team (the difference-in-difference estimates). For example, the difference-in-difference in median total time in the second column of table 5d is the difference between the change in median total time in Team 1 (second column table 5a) and the change in median total time in the control team (fourth column in table 5a). These difference-in-difference results allow us to estimate the impact of the scheme. We do not calculate standard errors as we only have three observations. In table 5d we consider all trader groups. The results suggest that both treatment teams increase their performance on all outcomes apart from negative yield, relative to the control team. If we only consider all incentivised trader groups (table 5e), Team 2 increases the time spent auditing the incentivised trader groups more than the control group whereas Team 1 does not. Total yield and positive yield from investigating the incentivised trader groups increase in both treatment teams relative to the Control team, and this is true also for productivity, but the increase is more substantial for Team 2 (in Team 2 total yield increases by more than three times than in Team 1 relative to the Control team, positive yield by more than five times and productivity by 1.6 times). Note that productivity on all trader groups (table 5d) increases more in Team 1 than in Team 2. However, productivity on incentivised trader groups (table 5e) increases more in Team 2 than in Team 1. In table 5f we consider only all non-incentivised trader groups. The results suggest that the scheme has no effect for Team 2 for the yield outcomes and productivity. Time spent on these groups decreases in Team 2 relative to the control team. The opposite is true for Team 1, which seems to have increased the time spent on non-incentivised trader groups relative to the control team and also the revenues from investigations and productivity on these trader groups.

Table 5g presents the difference-in-difference results for the incentivised trader groups (which are labelled *M*, which stands for Makinson) relative to non incentivised trader groups (*nonM*) in the two treatment teams. For example, the value of the median total time as in the second column of table 5g is the difference between the value in the second column of table 5e (the change in the median total time allocated to all incentivised trader groups in Team 1 relative to the control team) and the second column in table 5f (the change in the median total time allocated in all non-incentivised trader groups in Team 1 relative to the control team). These difference-in-difference-in-difference results allow us to consider whether the scheme had a different impact on incentivised

and non-incentivised activities in the two treated teams, controlling for officer fixed effects, team fixed effects and activity fixed effects.

The results indicate that in Team 1, the incentive scheme has a negative impact on the time spent on incentivised trader groups: Team 1 allocates more time to non-incentivised trader groups. The opposite happens in Team 2 where the median time allocated to incentivised trader groups increases relative to the time allocated to non-incentivised trader groups. Given that time contributes to events, this suggests that the scheme has the expected result on the *events* outcome only for Team 2. For the median total yield and positive yield outcomes we have similar results for both treated teams: they increase for the incentivised trader groups relative to the non-incentivised trader groups in both teams, though the increase is more substantial for Team 2 than for Team 1. Productivity on incentivised trader groups relative to non-incentivised trader groups decreases in Team 1 and increases in Team 2. Hence the incentive scheme has a positive impact on productivity on incentivised trader groups only in Team 2.

Officers with similar observable characteristics: matching analysis

In order to see whether officers with different observable characteristics respond to the incentive scheme in different ways, we analyse the effect of the incentive scheme for specific groups of officers. We perform a matching analysis. All staff is split up into one of 16 cells, which are created according to part-time/full-time status, gender, age, and pay. Details on how cells are created and their specification can be found in Appendix B. The cell matching allows us to separate officers into groups, which are homogeneous in terms of part-time/full-time status, age, gender and pay. We can then investigate whether officers with certain characteristics get a higher outcome compared to those officers without those specific characteristics.

We compare the median difference in outcome between incentivised and non-incentivised period at officer level in the trial team with the median difference in outcome between incentivised and non-incentivised period at officer level in the control team. This enables us to isolate any performance improvement that is driven by the incentive scheme for groups of officers with the same observable characteristics. We perform this analysis to investigate whether the response of officers to the incentive scheme differs according to their age and pay. One possibility is that younger workers may be better motivated to increase their effort, as they know performance will be evaluated and

hence they can influence their manager's perception about their ability for future promotion. Another possibility is that, as the bonus payment differs in the two trial teams (it is a flat rate varying across job bands in Team 1, whereas it is equal across job bands in Team 2), we might expect the high-paid officers in Team 1 to put in some more effort, as they get a higher bonus.

Table 6 presents the results of the cell matching difference-in-difference analysis. For each of the five outcomes, we calculate the median officer level difference in output (time, yield, positive yield, negative yield and productivity) for Team 1 (or 2) minus the median officer level difference in output for the Control team. A positive (negative) difference, for example for the time outcome, means that the officers in the treatment group spent more (less) time than those in the Control group for a specific cell. The opposite is true for the negative yield.

For the median outcomes for time, the results show that the male full time, younger, lower paid workers (cell 1) in Team 2 spend significantly more time than the same type of officers in the Control group. Female full-time, older, higher paid workers in Team 2 (cell 8) also spend more time than their Control group counterparts. However, female full-time, younger, higher paid workers (cell 6) in Team 2 spend significantly less time compared to those in the Control team. Hence, although there are some statistically significant differences across groups, there is no clear pattern as to *which* characteristics influence the median officer level difference in time between Team 2 and the control group. There are no significant differences within groups between Team 1 and the Control.

The same lack of a clear pattern across types of officers appears for the other four outcomes. When looking at the cell matching results of median positive yield, we see that cell 1 shows a significant positive difference both for Team 1 and Team 2. Cell 1 also has a positive effect for the median total yield for Team 1. We also look separately at incentivised and non-incentivised trader groups and derive similar conclusions. The specific officer characteristics do not have any consistent and clear effect on the median officer level difference in outcomes between the treatment and the control groups.

Officer level analysis

To estimate the incentivisation effect at officer level for the five different outcomes, we undertake various regression analyses. As we describe in Appendix B, the dependent variable in the regressions is the officer level difference in outcome (respectively time[14], yield, positive yield, negative yield, and productivity). For each of the five outcomes, we regress the officer level difference in outcome between the incentivised period and the non-incentivised period against a set of individual specific characteristics (years in grade, age, gender, part-time/full-time status, job band), team membership to reflect local market conditions, the type of trader groups dealt with, and a dummy variable, which takes the value 1 for the incentivised trader groups in the period of the scheme and 0 otherwise. The key parameter of interest is the coefficient estimate for this dummy variable, as it measures the effect of the Makinson incentive scheme on output.

Table 7 presents the results. We focus on the results for the coefficient of the dummy variable, so that each entry in the table measures the change in each of the five outcomes due to the incentive scheme, controlling for all the other variables. Robust standard errors are given in parentheses below the coefficient estimates[15]. The footnotes in the table specify which other variables are controlled for.

We present the results for four different specifications. We have performed other analysis using different specifications of regressions, however, we have chosen not to include these for reasons of space. In particular, we analysed regressions with variables indicating the interaction between the incentivisation dummy and the officer-specific variables (years in grade, age, mean age squared, gender, part-time/full-time worker and job band). These did not show any interaction effects, indicating no differences in responses across different grades and gender of officer. Additionally, we carried out median regressions (the regressions reported here are at the mean). However, since it is not possible to cluster the error term in these regressions, the estimated standard errors for these

---

14 As is already described in appendix B, the variable time is capped at 50 hours per week, to remove outliers or errors in coding. This affects 67 / 71325 observations.

[15] Significant coefficients are in bold. Three asterisks imply that a coefficient is significant at 1%, i.e. there is a 1% probability of rejecting the hypothesis that the variable does not have an effect. Two asterisks imply that a coefficient is significant at 5% and one asterisk implies that a coefficient is significant at 10%.

analyses were incorrect and are therefore not presented here[16]. The regressions without the outliers, which are shown here, produce somewhat similar coefficients as the median regressions.

For each specification we run the regression for both teams together, controlling for Team 2 and the Control team (see model in Appendix B) and also consider the two teams separately. Hence there are three sets of results for each specification.

The first specification is for all trader groups together and includes dummy variables for each trader group among the controls. The first row in the table shows that, when we evaluate the effect of the incentive scheme for both teams, the scheme has a positive effect on the change in outcome in all five regressions. However, none of the coefficients is statistically significant. If we consider the two teams separately, the Makinson coefficient estimate on time is negative for Team 1, although not statistically significant, and positive for Team 2. The result for Team 2 is statistically significant at the 1% level. Therefore, the incentive scheme has a negative effect on the time allocated to incentivised activities by an officer in Team 1. An officer in Team 1 spends 5 hours less on incentivised trader groups in the incentivised period compared to the pre-incentivised period, holding all other variables fixed. In contrast, officers in Team 2 increase, on average, their time spent on incentivised trader groups by 38 hours in the incentivised period compared to the previous period, holding all other variables fixed. For total yield and positive yield the coefficient estimates are positive for both teams. However, results are not statistically significant from 0 for any of the three regressions. Negative yield increases[17] in both teams though results are not statistically significant.

The second specification uses the same controls, but omits the non-VAT and non-trader audit categories from the outcome variables. The non-VAT category includes a range of trader groups, for example, tobacco duties and international trade exports, that take up very little of the trial teams' and control team's time but count for a very substantial amount of positive yields collected. Hence, it is likely that this category represents money paid to HM Customs and Excise as standard procedures, without officers making (many) visits. Therefore, these are not necessarily representative trader groups and are excluded in these regressions. The same is true for the non-

---

16 We wish to allow for clustering at team level to allow for error correlation across officers within a team.

17 As noted above, negative yield enters the regression with a negative sign, so that a negative coefficient implies an increase in negative yield.

trader audit group, which represents all activities recorded under activities other than trader audit, which are not assigned to a particular trader group.

The results from the regressions for both teams again show a positive incentive effect for four out of five regressions, but these differences are not statistically significant. If we consider the two teams separately, the time-coefficients are very comparable to the previous ones. The Makinson coefficient is negative for Team 1 but not statistically significant and significantly positive for Team 2.

We consider a third specification for total yield, positive yield, negative yield and productivity. This specification excludes outliers of the dependent variables. We take out the 5% at the top and bottom of the distribution. The motivation behind this choice is that a very high/low yield may not necessarily reflect the officer's effort in the period considered (nine months). Some officers may spend almost a year on one case, while not getting any yield in this period (for example because the case is extremely complicated to assess), or they could make a big hit in the period considered but this may not be necessarily representative of the trader group. It is also easier to make reporting errors when reporting the amount of yield. There is no upper boundary (in contrast to time, where restricting this variable to a maximum of 50 hours a week only affected 67 out of 71325 observations). When we look at the distribution of yields, we see that the highest recorded yield in the two years of data is £284,000,000. The mean is only £54,161 and the median equals zero. We therefore decide to explore the effect of incentivisation leaving out the outliers[18]. This also allows us to control for the revenue capping introduced by the designers of the incentive scheme to assess team performance[19].

The regressions for both teams excluding outliers show a significant positive effect for yield, positive yield and negative yield (significant at 1% for yield and positive yield and at 5% for negative yield). The first two have large coefficients: both total yield and positive yield per officer in both teams increases on average by more that £24,000 for the incentivised trader groups in the incentivised period relative to the previous period. Negative yield increases on average by more than £634. When we consider the two teams separately, the scheme has a positive

---

[18] For yield, positive yield and negative yield, this affected 292 out of 2935 observations. For productivity, this affected 294 out of 2935 observations.

[19] By taking out the outliers we trim our data by 88 observations out of 869 in Team 1, 88 out of 867 in Team 2 and 120 out of 1199 in the Control team. If we had capped revenues in the same way as the scheme designers did, we would have taken out 72 observations for Team 1, 88 for Team 2 and 121 for Team 3.

effect on total yield and positive yield for both teams and results are statistically significant at 1%. In Team 1 total yield for the incentivised trader groups increases on average by £25,815 over the period and in Team 2 by £29,781. Positive yield increases on average by £25,696 in Team 1 and by £28,955 in Team 2. Negative yield increases in both teams but less in Team 1 than in Team 2. Productivity increases more substantially in Team 1 than in Team 2, and results are statistically significant (at 5%) only for Team 1. Officers in Team 1 increase yields per hour on an incentivised trader group on average by £71 in the nine months of the incentive scheme. Officers in Team 2 increase yields per hour on an incentivised trader group on average by £27, though the result is not statistically significant. We should note that the big increase in productivity in Team 1 is linked to the results we obtain for total yields and time spent on incentivised trader groups: officers in Team 1 spend less time and collect more total yields on incentivised trader groups, so that productivity necessarily increases.

In the forth specification we exclude outliers *and* omit the non-VAT and non-trader audit trader groups. The results are very similar to the third specification, both in terms of coefficients and standard errors. Again, the incentive scheme has a statistically significant positive effect on total yield and productivity for both teams.

Summary
To summarise, our results indicate the effects of the scheme were as follows:

- Team 1 and 2, on aggregate, spend more time on all trader groups and recover more yield than the control team.

- If we distinguish between incentivised and non incentivised trader groups, Team 2, on aggregate, increases the time spent on incentivised trader groups relative to non incentivised trader groups. This is not true for Team 1, which decreases the time spent auditing the incentivised trader groups.

- The median total yield and positive yield increase for the incentivised trader groups relative to the non-incentivised trader groups in both teams. The increase is larger for Team 2 than for Team 1.

- Productivity on incentivised trader groups relative to non-incentivised trader groups increases in Team 2 but decreases in Team 1.

- There is no clear pattern of how different observable characteristics of officers drive any performance improvement.

- The analysis at officer level gives similar results to the aggregate analysis at team level for the time outcome. Officers in Team 2, on average, increase their time spent on incentivised trader groups by 38 hours in the incentivised period compared to the previous period, holding all other variables fixed. This does not happen in Team 1, where officers, on average, spend 5 hours less on incentivised trader groups, though the result is not statistically significant.

- For the total yield and productivity outcomes results at officer level differ from results at team level: officers in both teams increase total yield and productivity for the incentivised trader groups. Officer in Team 1 increase total yield on average by £25,000 during the scheme while officers in Team 2 increase total yield by £29,000. The increase in productivity on incentivised trader groups is more substantial in Team 1 than in Team 2: yields on an incentivised trader group per hour increase on average by £71 in the nine months of the incentive scheme in Team 1 and £27 in Team 2, though results are statistically significant only for Team 1.

- Hence, for the yield outcome, the scheme has a similar impact on officers in the two treatment teams, and productivity increases more in Team 1 than in Team 2. However, the response at team level differs and Team 2, on aggregate, performs better than Team 1 in all outcomes on incentivised trader groups. This difference in response at individual and team level suggests that the strategy at team level differs between the two treated teams. An important team strategy is the allocation of officers across trader groups. It makes sense to assume that this decision is taken by managers. In particular, managers can strategically allocate more efficient officers on incentivised trader groups in order to meet the targets. In what follows, we investigate how the allocation of efficient officers differs across the two treated teams and the control team.

Additional issues

Strategic task management

We explore whether the officers with a high productivity in the first period were allocated to the incentivised trader groups in the incentivised period in order to increase the chance of hitting the targets. An efficient officer is defined as one whose productivity on VAT-trader groups in the non-incentivised period was in the top 25% of the productivity distribution of all officers in that period. We analyse whether there is a statistically significant difference in time spent between the two periods on both the incentivised and the non-incentivised trader groups by efficient officers and

other officers in each team. The increase in time spent is defined as the time spent in the incentivised period minus the time spent in the previous period.

We first test whether the mean change in time spent by efficient officers on incentivised trader groups is equal to the mean change in time spent by the other officers on incentivised trader groups (this would imply that there wasn't any strategic task allocation). We then do the same for the non-incentivised trader groups. For incentivised (non-incentivised) trader groups, we expect the efficient officers to have a relatively higher (lower) mean change (i.e. a more (less) positive number) than the other officers.

Table 8 presents these results. For the incentivised trader groups, looking at all teams simultaneously, the results show that both types of officers increase the time spent on incentivised trader groups over the two periods. However, the mean increase for efficient officers is more than that of the other officers, and the result is statistically significant at 1%. When we consider the three teams separately, all teams have a positive mean change but results are statistically significant only for Team 1 (at 10%) and for Team 2 (at 1%).

In Team 2 the efficient officers substantially increase the time spent on incentivised trader groups compared to other officers. The mean time spent on incentivised trader groups increases by 276 hours in the nine months of the scheme for efficient officers, whereas the other officers increase the time spent on the same trader groups by 99 hours. So efficient officers in Team 2 increase the time spent on incentivised trader groups by almost three times as much relative to the other officers. In Team 1 efficient officers increase the time spend on incentivised trader groups by 87 hours, whereas the other officers increase their time spent on the same trader groups by 35 hours. Hence efficient officers in Team 1 spend more time on incentivised trader groups than the other officers, but the mean increase is substantially less than for efficient officers in Team 2. The mean change is positive for efficient officers and for the other officers in the control team, but is not significant.

For the non-incentivised trader groups for all three teams, we see that efficient officers spend less time on these in the incentivised period. The other officers, however, increase the time spent on non-incentivised trader groups.

When we analyse the three teams separately, we see that the efficient officers in Team 2 and the Control team spend significantly less time on non-incentivised trader groups than the other officers

in their team. In Team 2 also the other officers decrease their time on non-incentivised trader groups, but less than efficient officers (an average decrease of 122 hours for efficient officers versus 8 hours for the other officers). In the Control team the other officers increase their time on non-incentivised trader groups. In Team 1 efficient officers increase their time spent on non-incentivised trader groups, though less than the other officers (an average increase of 29 hours for efficient officers versus 54 hours for the other officers), though results are not statistically significant.

It is clear that efficient officers in Team 2 reallocate their time to incentivised trader groups much more than in the other two teams. If we consider the net increase in time spent on incentivised trader groups, defined as the increase in time spent on incentivised trader groups minus the increase in time spent on non-incentivised trader groups, efficient officers in Team 2 reallocate 398 more hours on incentivised trader groups: they increase the time spent on incentivised trader groups by 276 hours and decrease the time spent on non-incentivised trader groups by 122. In Team 1 the net reallocation is 58 hours and in the control team 137 hours.

If we look at the net reallocation by other officers, in Team 2 they increase the time spent on incentivised trader groups by 107, in Team 1 they *decrease* the time spent on incentivised trader groups by 19 hours and in the control team they increase by 19 hours.

Hence officers in Team 2 work strategically harder: they swap their time from non-incentivised trader groups to incentivised trader groups and the swap is much greater for efficient officers. As already mentioned, overtime increased very little during the scheme in both teams, so that officers' relocation to tasks may have had a substantial impact on team level performance. Strategic task allocation is less clear in Team 1 where efficient officers spend more time than the rest of officers on all trader groups, both incentivised and non incentivised.

To see how officers changed their time allocation across the different trader groups after the introduction of the scheme, we regress the officer level change in time between the incentivised period and the previous period on the incentivisation dummy and 11 interaction dummies, controlling for the same variables as in specification 2 of table 7. The 11 interactions are the VAT trader group dummies interacted with the 'efficient officer'-dummy. The coefficient on these variables gives the officer level change in time for efficient officers on the VAT trader groups.

The results are presented in table 9. Note that the incentivised trader groups for the *time* outcome are new registrations and exceptional risk, which are highlighted in the table. Efficient officers in Team 1 increase, on average, the time spent on exceptional risk traders by 63 hours in the incentivised period relative to the previous period, holding all other variables constant. They also increase time spent on other non-incentivised trader groups like insolvent, median risk, missing traders, branches, deregistered, though results are statistically significant for the insolvent trader group. They, however, decrease the time spent on the other incentivised trader group - new registrations - by 15 hours over the period, though the result is not statistically significant.

Efficient officers in Team 2 increase their time spent on exceptional risk traders by 123 hours in the incentivised period and on new registrations by 44 hours, which are the incentivised trader groups for the *time* outcome. They decrease time spent on all other trader groups, apart from deregistered and insolvent, though coefficients are statistically significant only for the deregistered trader group. Overall, although results are not significant in either team for most of the trader groups, differences are present. And the results confirm that Team 2 has a more strategic approach in allocating efficient officers to the incentivised trader groups in order to hit the targets. Ex-post we know that Team 2 hit all its targets, but Team 1 missed the target on events for the new registrations trader group and also the yield on this trader group, while it hit all other targets. This is the trader group on which efficient officers spent less time in Team 1 during the incentivised period.

Whether this strategic task reallocation is due to the incentive scheme is not clear. It could be the managers' response to the incentive scheme, but it could also be an already existing difference in the managerial strategy across the two incentivised teams. One team may have a more flexible task assignment strategy because officers are used to dealing with different trader groups, so that they have enough experience to be effectively reallocated across trader groups. And this could be due to the initial choice on the organisation of work in offices or it could also be due to the fact that some trader groups are concentrated in certain areas, so that officers working in those areas need to focus on a limited number of trader groups and hence there is less scope for reallocation. In both cases a more/less flexible task allocation strategy is not related to the incentive scheme.

But the lack of strategic allocation of officers across tasks by the managers is only one possible explanation for why Team 1 missed its targets. Other issues arising from the nature of the scheme may be responsible for this outcome. As we discussed in section 3, the team structure designed for the Makinson incentive scheme is quite complex. In particular, it is important to distinguish the team enforcement problem at office level from the team enforcement problem at division level. A

greater number of workers in an office makes the free-riding problem more difficult to tackle and a greater number of offices within a division makes the coordination and monitoring of offices behaviour more complicated for the band 11 manager. Hence there would be a greater incentive to free-ride in teams with bigger and more numerous offices. In our case, the trial teams have nearly the same number of officers, but Team 1 has three offices, whereas Team 2 has six offices. We could not make a clear prediction on what to expect as, although coordination and monitoring might have been more difficult for the band 11 manager in Team 2, free-riding at office level could have been less problematic in Team 2 than in Team 1. On the basis of the ex-post results of the two teams, one possible explanation of why Team 1 performed worse that Team 2 could be that for Team 1 the free-riding problem within an office was a more important problem than coordination. This remains only a speculation on our side, as we didn't have the details of which officers were located in the same office and so could not test how office size affects the results.

There is an issue also regarding the way the structure of the bonus differs across the two teams. For Team 1 the bonus depends on the job band, it is roughly a fixed percentage of salary, whereas in Team 2 it is equal across job bands. This means that the team members in Team 1 (typically in job band 6 and 7) who are more directly involved in the visit and collection of yield than others would get a lower reward for reaching the target than other team members who are less directly involved in the production process, but are in a higher job band. This may reduce their effort. We have some evidence on the fairness considerations regarding the design of the bonus. One result of the qualitative study commissioned by HMCE was that 69% of interviewed staff thought that if successful, everyone should receive the same bonus. If this affected behaviour, it might have reduced effort in Team 1.

Summary of findings

Our statistical approach has been to examine the effect of the scheme, netting out factors that may affect performance over and above the impact of the scheme. The factors that are netted out are differences in staffing, local market conditions, and differences in dealing with specific trader groups. Once these are controlled for we find:

- Both treated teams, on aggregate, spend more time and recover more yield on all trader groups than the Control team. However, if we distinguish between incentivised and non-incentivised trader groups, officers in Team 2 perform better than officers in Team 1: they increase the time spent and productivity on incentivised trader groups relative to non-incentivised trader groups. Officers in Team 1 decrease both the time spent and productivity on incentivised trader groups relative to non-incentivised trader groups. Officers in both teams increase total yield and positive yield on incentivised trader groups relative to non-incentivised trader groups, but the increase is larger for Team 2 than Team 1.

- The analysis at officer level provides similar results to the aggregate analysis at team level for the time outcome. The incentive scheme has a positive effect on the time allocated by officers in Team 2 to incentivised trader groups. Team 2 officers increase their time spent on incentivised trader groups by 38 hours in the incentivised period compared to the non-incentivised period, holding all other variables fixed. On the contrary, officers in Team 1 spend 5 hours less on incentivised trader groups in the incentivised period compared to the pre-incentivised period, holding all other variables fixed.

- For the total yield outcome, the scheme has a similar impact on officers in both treated teams. Productivity increases more for officers in Team 1 than for officers in Team 2. This contrasts with the results at team level, according to which Team 2 performs better than Team 1 on incentivised trader groups. In particular, officers in Team 1 increase total yield on average by £25,800 during the scheme while officers in Team 2 increase total yield by £29,800. Officers in team 1 increase yields per hour on an incentivised trader group on average by £71 in the nine months of the incentive scheme. The increase in productivity in Team 2 is on average £27, but not statistically significant

- There is some evidence of strategic task assignment in Team 2. In all three teams, efficient officers increase the time on incentivised trader groups whilst the scheme was in operation, but results are statistically significant only for Team 1 and Team 2. Efficient officers in Team 2 substantially increase the time spent on incentivised trader groups (an average

increase of 276 hours in the nine months of the scheme versus 99 hours for the other officers). Moreover, efficient officers in Team 2 spend significantly less time on non-incentivised trader groups than the other officers. On the other hand, efficient officers in Team 1 seem to have increased their time on non-incentivised trader groups, though less than the other officers (an average increase of 29 hours for efficient officers versus 54 hours for the other officers). This suggests that efficient officers in Team 2 swapped their time from non-incentivised trader groups to incentivised trader groups after the introduction of the scheme, while efficient officers in Team 1 did not. This implies that the team strategy on how to allocate officers to trader groups differs across the two treated teams. We can therefore reconcile the difference in the response to the scheme at team level and the response at officer level: all officers in both teams work harder in the period of the incentive scheme, but officers in Team 2 also work smarter strategically, so hitting the targets.

References

A. Dixit "Incentives and organisations in the public sector: an interpretative review" mimeo, 2000.

B. Holmström "Moral hazard in teams", Bell Journal of Economics, 13, 1982, 324-340.

H. Itoh "Incentives to help in multi-agent situations", Econometrica, 59, (3),1991, 611-636.

Kandel, E. and Lazear, E. (1992) "Peer Pressure and Partnerships", *Journal of Political Economy*, 100 (4), pp. 801-817.

Knez, M., and Simester, D. (2001) "Firm-wide incentives and mutual monitoring at Continental Airlines", *Journal of Labor Economics*, 19(4), pp.743-772.

G. MacDonald, L. M. Max "Adverse Specialization", Journal of Political Economy, 109 (4), 2001, 864-899.

J. Makinson "Incentives for change. Rewarding performance in national government networks", Public Service Productivity Panel, 2000.

Olson, M (1971) *The logic of collective action.* Harvard University Press, Cambridge

J. M. Wooldridge (2002) *Econometric analysis of cross section and panel data.* The MIT Press, Cambridge, Massachusetts

Tables and Figures

Table 1: Output statistics per officer, per trader group

Variable	Mean	Std Dev	P25	P50	P75	P99	No. obs
Time	8.24	7.67	2	6.5	11.5	32.7	70130
Pos. Yield (£)	301,811	4,430,267	1,325	4,591	16,751	4,786,177	17741
Neg. Yield (£)	-3,485.73	39,457.77	0.00	0.00	0.00	0.00	17741

The unit of observation is weekly recorded time/yield per officer per trader group in the two periods.

Table 2: Share of total time spent on each trader group within trader audit (in %).

| Period | Category | Trader audit | | |
		Team 1	Team 2	Team C
1	Non-trader audit	0.07	0.04	0.19
	Non-VAT	3.24	2.88	6.19
	New registration	41.36	40.47	38.82
	Low risk	5.44	6.29	3.38
	Medium risk	6.05	7.47	5.23
	High risk	24.98	24.17	24.61
	Exceptional risk	14.61	10.02	16.82
	999 large traders	1.85	3.70	0.44
	Corporate groups	0.04	.	0.03
	Branches	0.01	0.22	0.08
	Insolvent	0.02	0.18	0.09
	Deregistered	0.38	0.77	0.39
	Missing traders	0.06	0.10	0.07
	Sifted complaint	1.91	3.71	3.67
2	Non-trader audit	.	.	.
	Non-VAT	4.238	3.103	8.078
	New registration	30.124	35.573	33.692
	Low risk	8.112	5.812	7.064
	Medium risk	11.988	7.691	9.546
	High risk	17.466	13.336	14.791
	Exceptional risk	24.388	28.838	25.532
	999 large traders	1.874	4.001	0.217
	Corporate groups	.	0.001	0.043
	Branches	0.023	0.043	0.016
	Insolvent	0.143	0.266	0.149
	Deregistered	0.857	0.929	0.641
	Missing traders	0.779	0.400	0.216
	Sifted complaint	0.008	0.005	0.014
3	Non-trader audit	.	.	.
	Non-VAT	4.635	4.009	7.910
	New registration	30.741	35.048	34.273
	Low risk	17.762	6.337	16.727
	Medium risk	12.108	7.821	10.744
	High risk	12.522	12.859	13.003
	Exceptional risk	16.996	24.292	15.546
	999 large traders	3.324	6.863	0.378
	Corporate groups	0.023	0.037	0.233
	Branches	.	.	0.117
	Insolvent	0.278	0.369	0.120
	Deregistered	0.995	1.913	0.792
	Missing traders	0.542	0.441	0.154
	Sifted complaint	0.075	0.011	0.007

The shares sum to 100% in each period.

Table 3: Hours of total time spent on each trader group within trader audit.

Period	Category	Trader audit Team 1	Team 2	Team C
1	Non-trader audit	34	23	147
	Non-VAT	1,571	1,854	4,866
	New registration	20,075	26,093	30,506
	Low risk	2,639	4,059	2,659
	Medium risk	2,939	4,814	4,108
	High risk	12,123	15,585	19,338
	Exceptional risk	7,090	6,459	13,220
	999 large traders	898	2,384	345
	Corporate groups	17	.	27
	Branches	7	139	62
	Insolvent	10	117	69
	Deregistered	183	496	306
	Missing traders	27	66	55
	Sifted complaint	927	2,390	2,884
2	Non-trader audit	.	.	.
	Non-VAT	2,577	2,406	7,210
	New registration	18,315	27,583	30,073
	Low risk	4,932	4,507	6,305
	Medium risk	7,299	5,963	8,520
	High risk	10,619	10,341	13,201
	Exceptional risk	14,827	22,361	22,789
	999 large traders	1,139	3,103	194
	Corporate groups	.	1	38
	Branches	14	34	14
	Insolvent	87	206	133
	Deregistered	521	720	572
	Missing traders	474	310	193
	Sifted complaint	5	4	12

Table 4: Team staff mean characteristics.

	All staff				Continuous frontline staff			
	Overall	Team 1	Team 2	Control	Overall	Team 1	Team 2	Control
Gender	0.47	0.39	0.44	0.52	0.44	0.40	0.40	0.50
Age	44.2	45.7	45.4	42.8	44.6	45.0	45.5	43.7
Job band	5.9	6.2	5.9	5.8	6.2	6.3	6.3	6.0
Years in job band	3.9	4.0	4.3	3.6	4.2	4.0	4.7	4.1
Years at Customs &Excise	18.1	19.1	19.1	17.1	19.5	19.3	20.0	19.2
Annual pay (£)	20247	20797	20066	20047	20671	20742	20789	20551
Gross monthly pay (£)	1594	1630	1599	1571	1632	1619	1657	1625
Gross monthly pay + overtime (£)	1599	1632	1599	1582	1639	1621	1657	1639
Overtime per month (hours)	0.38	0.16	0.02	0.70	0.44	0.15	0.02	0.88
Part-time marker	0.16	0.16	0.13	0.17	0.16	0.19	0.12	0.16
Part-time hours per week	24.8	24.1	-	25.1	24.7	23.8	-	25.2
Days of sickness	1.1	1.1	1.5	0.9	1.0	1.2	1.3	0.8
Days in scheme	260.2	258.8	261.5	-	264.0	262.4	265.6	-
Potential bonus (£ per person)	-	688	681	-	-	695	696	-

Table 5a

For all trader groups

Differences between incentivised & pre-incentivised period

Variable	Team 1	Team 2	Control
Total Time	50.5	103.35	41.5
Total Yield	67193	73441.5	0
Positive Yield	84983	80877	0
Negative Yield	0	0	0
Productivity	137.54	120.13	0

Table 5b

For all incentivised

Differences between incentivised & pre-incentivised period

Variable	Team 1	Team 2	Control
Total Time	-3.8	73.89999	-1.5
Total Yield	10901	31725	0
Positive Yield	8370	46550.5	0
Negative Yield	0	0	0
Productivity	58.70	89.78	9.17

Table 5c

For all non-incentivised

Differences between incentivised & pre-incentivised period

Variable	Team 1	Team 2	Control
Total Time	13.5	-13.4	1.099998
Total Yield	6406	0	0
Positive Yield	6406	0	0
Negative Yield	0	0	0
Productivity	58.24	0	0

Table 5d

Difference-in-Difference

Variable	T1-C	T2-C
Total Time	9	61.85
Total Yield	67193	73441.5
Positive Yield	84983	80877
Negative Yield	0	0
Productivity	137.54	120.13

Table 5e

Difference-in-Difference

Variable	T1-C	T2-C
Total Time	-2.3	75.40
Total Yield	10901	31725
Positive Yield	8370	46550.5
Negative Yield	0	0
Productivity	49.53	80.61

Table 5f

Difference-in-Difference

Variable	T1-C	T2-C
Total Time	12.4	-14.5
Total Yield	6406	0
Positive Yield	6406	0
Negative Yield	0	0
Productivity	58.24	0

Table 5g

Difference-in-Difference-in-Difference

Variable	M*(T1-TC) – $nonM$*(T1-TC)	M*(T2-TC) – $nonM$*(T2-TC)
Total Time	-14.7	89.89999
Total Yield	4495	31725
Positive Yield	1964	46550.5
Negative Yield	0	0
Productivity	-8.71	80.61

Table 6 – All Trader Groups
Cell Matching difference-in-difference results (median)

Cell [20]	Time		Yield		Positive yield		Negative yield		Productivity	
	T1-TC	T2-TC	T1-TC	T2-TC	T1-TC	T2-TC	T1-TC	T2-TC	T1-TC	T2-TC
1	72	**144.7***	**95638**	120906	**89314**	**133816***	0	0	137.5	206.4
	(52)	**(52)**	**(44743)**	(76731)	**(45338)**	**(80345)**	(1694)	(4223)	(126.7)	(192.7)
2	81	193	392558	207689	372516	408537	-541	0	372.8	594.8
	(68)	(150)	(310574)	(352568)	(280107)	(345756)	(9373)	(21599)	(1783)	(965.1)
3	-16	-18	94498	7618	97876	**50616***	0	-5737	120.6	-28.3
	(57)	(93)	(62740)	(62566)	(77816)	**(79397)**	(5255)	(8983)	(369.2)	(1004.9)
4	-23	16	76257	228586	68504	340771	-1239	-1239	420.9	220.2
	(55)	(78)	(194244)	(155255)	(221802)	(203883)	(11403)	(10878)	(621.1)	(214.9)
5	-96	-39	0	56993	0	62033	0	-1889	0	39.7
	(197)	(278)	(20438)	(46120)	(25357)	(49289)	(4487)	(2876)	(626.5)	(85.2)
6	-62	**-224.2***	67334	88684	111876	56127	-15062	0	-2.2	110.5
	(114)	**(130)**	(4708698)	(431890)	(4408510)	(376026)	(25192)	(11378)	(7896.5)	(459.7)
7	149	25	0	21578	0	18435	0	12332	205.9	145.7
	(117.6)	(126.4)	(416399.8)	(66363.4)	(373088.8)	(71816.3)	(2927.3)	(7650.7)	(3573.8)	(143.7)
8	4	**229.9***	181932	252247	195229	235397	0	2330	331.6	100.3
	(134.4)	**(106.3)**	(396084.2)	(1834150)	(1854586)	(197525.9)	(10471.6)	(11108.6)	(9004.4)	(503.1)
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0
13	-32	-71	-1455	-7853	-1081	-7479	-43	-43	-98.0	0
	(96)	(64)	(66068)	(61472)	(75126)	(72020)	(13599)	(13565)	(172.8)	(0)
14	49	-1	7839	25769	8288	30848	0	0	387.0	131.8
	(131)	(116)	(4265628)	(5239848)	(4730489)	(4820144)	(35476)	(15909)	(22000.6)	(23843.6)
15	41	0	0	42852	0	41951	0	-1292	118.7	**112.0***
	(83)	(157)	(66242)	(137125)	(57747)	(180374)	(1028)	(49432)	(400.4)	**(421.6)**
16	-38	-74	105342	89100000	172382	87500000	10240	1535934	135.9	133409
	(151)	(235)	(46700000)	(7.85e+07)	(4.61e+07)	(7.65e+07)	(41638.9)	(1340698)	(127320)	(116670.5)
Weighted Sum	18	52	117206	1680216	119074	1696540	-606	26564	212	2489

---

[20] The cells are defined in Appendix B.

Table 7
Difference in difference regressions estimating the Makinson effect at officer level.

	Time	Yield	Positive yield	Negative yield	Productivity
Specification 1[a]					
Both teams	16.662	862,226.531	879,553.131	-17,326.599	382.772
	(12.538)	(652,598.418)	(653,435.097)	(12,221.249)	(3,608.746)
Team 1	-4.922	136,162.869	145,919.115	-9,756.246	239.004
	(12.373)	(98,686.498)	(98,767.889)	(7,560.350)	(470.589)
Team 2	**37.894 *****	978,548.109	1,001,806.908	-23,258.800	271.321
	(14.520)	(860,264.244)	(860,344.116)	(21,407.965)	(408.845)
Specification 2[b]					
Both teams	14.010	435,750.800	453,579.263	-17,828.463	40.633
	(12.902)	(412,809.891)	(414,757.543)	(12,356.234)	(144.382)
Team 1	-4.581	47,088.951	57,081.507	-9,992.555	166.134
	(12.420)	(42,239.992)	(43,773.840)	(7,575.689)	(249.878)
Team 2	**37.937 *****	896,445.946	919,626.879	-23,180.933	58.324
	(14.541)	(860,973.361)	(860,979.605)	(21,053.030)	(151.494)
Specification 3[c]					
Both teams		**24,517.457*****	**24,629.887*****	**-633.594*****	28.494
		(5,292.030)	(5,838.805)	(226.192)	(23.125)
Team 1		**25,815.132*****	**25,696.292*****	**-562.030****	**70.981****
		(7,004.909)	(7,864.822)	(218.070)	(32.555)
Team 2		**29,781.494*****	**28,954.846*****	**-899.411*****	26.951
		(6,927.112)	(7,314.759)	(297.831)	(22.341)
Specification 4[d]					
Both teams		**22,360.027*****	**22,154.517*****	**-585.759****	32.439
		(5,254.457)	(5,842.350)	(236.258)	(23.020)
Team 1		**25,180.406*****	**25,583.105*****	**-549.367****	**70.908****
		(6,873.621)	(7,827.547)	(219.232)	(32.484)
Team 2		**28,775.089*****	**28,803.199*****	**-879.165*****	27.019
		(6,879.612)	(7,288.997)	(296.545)	(22.287)
* significant at 10%; ** significant at 5%; *** significant at 1%					

[a] This specification controls for the following variables: team membership, years in grade, age, mean age squared, gender, part-time/full-time worker, job band, all trader groups.

[b] This specification controls for the same variables as above, but excludes the non-VAT and non-trader audit trader groups.

[c] This specification controls for the same explanatory variables as specification 1, but excludes the outliers of the dependent variables (defined as the top and bottom 5%).

[d] This specification excludes the outliers of the dependent variables and omits the non-VAT and non-trader audit trader groups among the explanatory variables.

Table 8 – Strategic task allocation

Officer level difference in time spent per month on VAT trader groups during the 9 months of the scheme relative to the same period prior to the scheme.

Productivity for officer *i* on trader group *j* is defined as the sum of yield (9mths) over time (9mths) on trader group *j*, weighted by share of time spent on that trader group. Productivity on all trader groups considered by officer *i* is:

$$productivity_i = \sum_{TG=1}^{j} \left[\left(\frac{\sum_{m=1}^{9} yield_{i,TG_j,m}}{\sum_{m=1}^{9} time_{i,TG_j,m}} \right) \times \left(\frac{\sum_{m=1}^{9} time_{i,TG_j,m}}{\sum_{TG=1}^{j}\sum_{m=1}^{9} time_{i,TG_j,m}} \right) \right]$$

The number of efficient officers for each team is: 28 (out of 129) for Team 1, 29 (out of 124) for Team 2 and 35 (out of 197) for the control team. (Total=92 out of 450).

Incentivised TG

	All teams	Team 1	Team 2	Control team
Efficient officers	131.94	87.12	275.91	48.51
Others	56.89	35.02	98.84	45.92
	t = 2.9816	t = 1.4958	t = 3.4616	t = 0.0641
	Significant at 1%	Significant at 10%	Significant at 1%	Not significant

Non-incentivised TG

	All teams	Team 1	Team 2	Control team
Efficient officers	-63.56	29.31	-122.17	-89.29
Others	26.04	54.05	-8.31	28.73
	t = -4.3070	t = -0.7410	t = -2.6602	t = -3.7477
	Significant at 1%	Not significant	Significant at 1%	Significant at 1%

Table 9 – Time change per trader group

	Team 1	Team 2
Efficient officer * New registrations	-15.001 (21.992)	**43.716 *** (25.351)
Efficient officer * Low risk	-3.071 (8.022)	**-26.502 *** (9.769)
Efficient officer * Median risk	14.049 (9.806)	-18.213 ** (8.882)
Efficient officer * High risk	-15.169 (20.790)	-45.998 (29.948)
Efficient officer * Exceptional risk	**62.801 *** (25.398)	**122.785 *** (42.559)
Efficient officer * (999) Large traders	-14.579 (23.095)	-55.919 (41.144)
Efficient officer * Corporate groups	0 (0.000)	-9.942 (6.982)
Efficient officer * Branches	0.495 (4.063)	0 (0.000)
Efficient officer * Insolvent	**9.547 *** (3.865)	4.715 (3.299)
Efficient officer * Deregistered	3.002 (3.116)	**9.241 *** (3.895)

Efficient officer * Missing traders	10.339 (6.056)	-0.232 (2.515)

Figure 1

Possible activities that an officer can be involved in within a VAT Assurance Division.
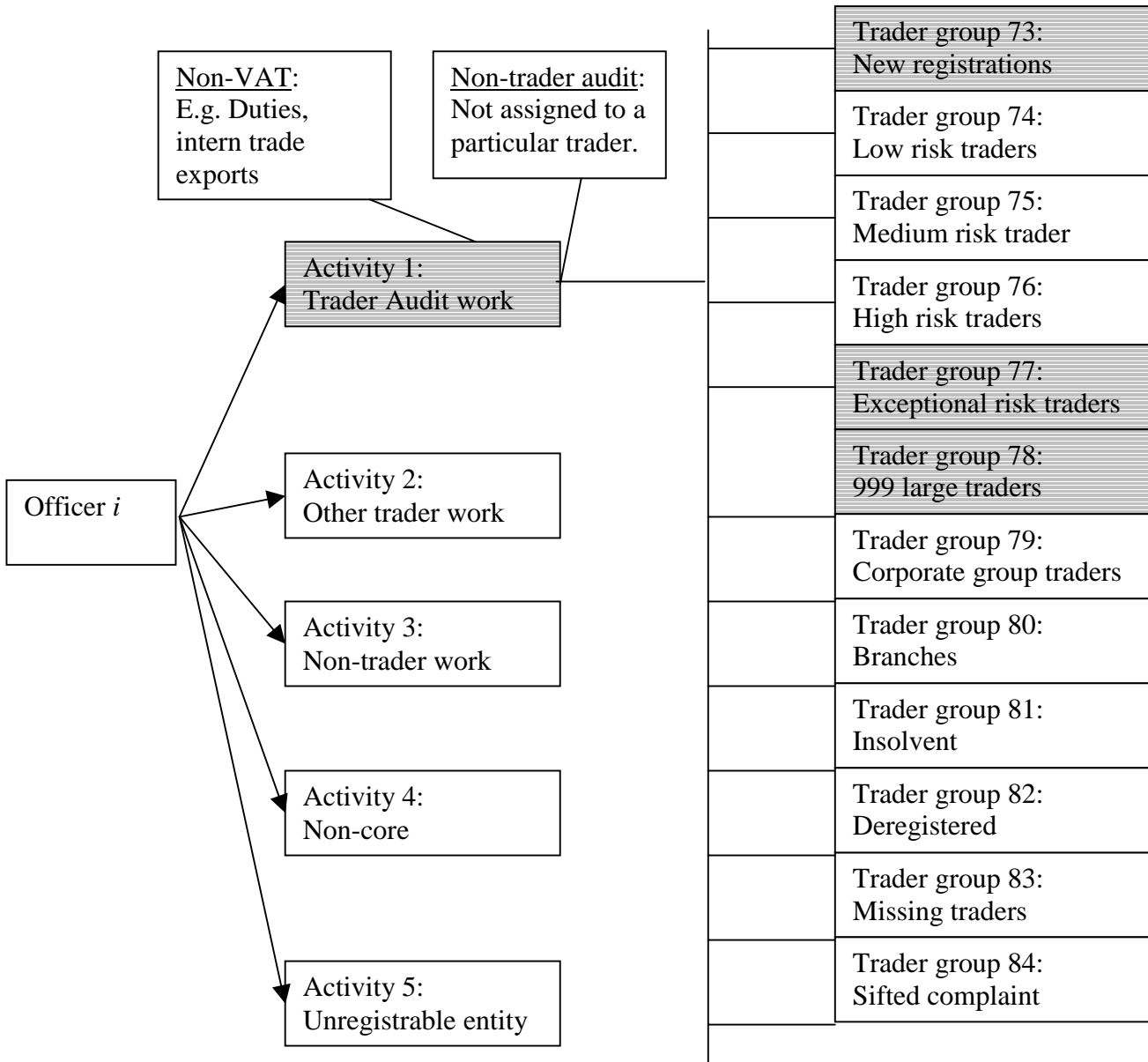
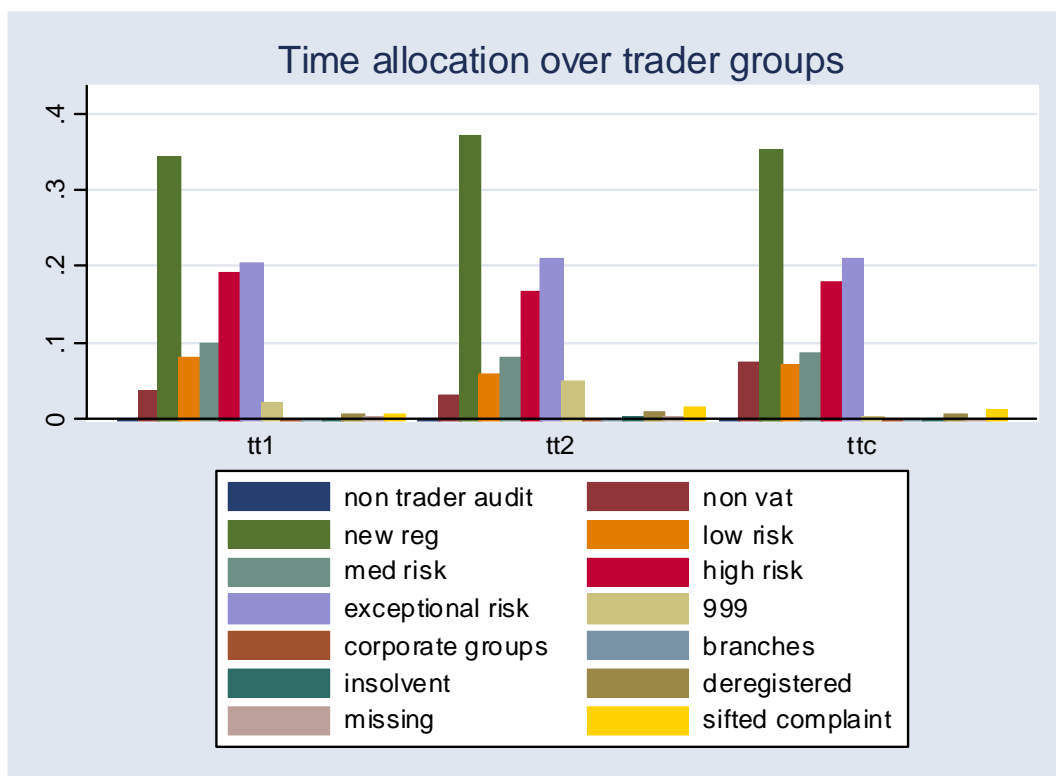Figure 2: Time allocation across trader groups.

Time allocation over trader groups

Figure 3: Time allocation across activities.
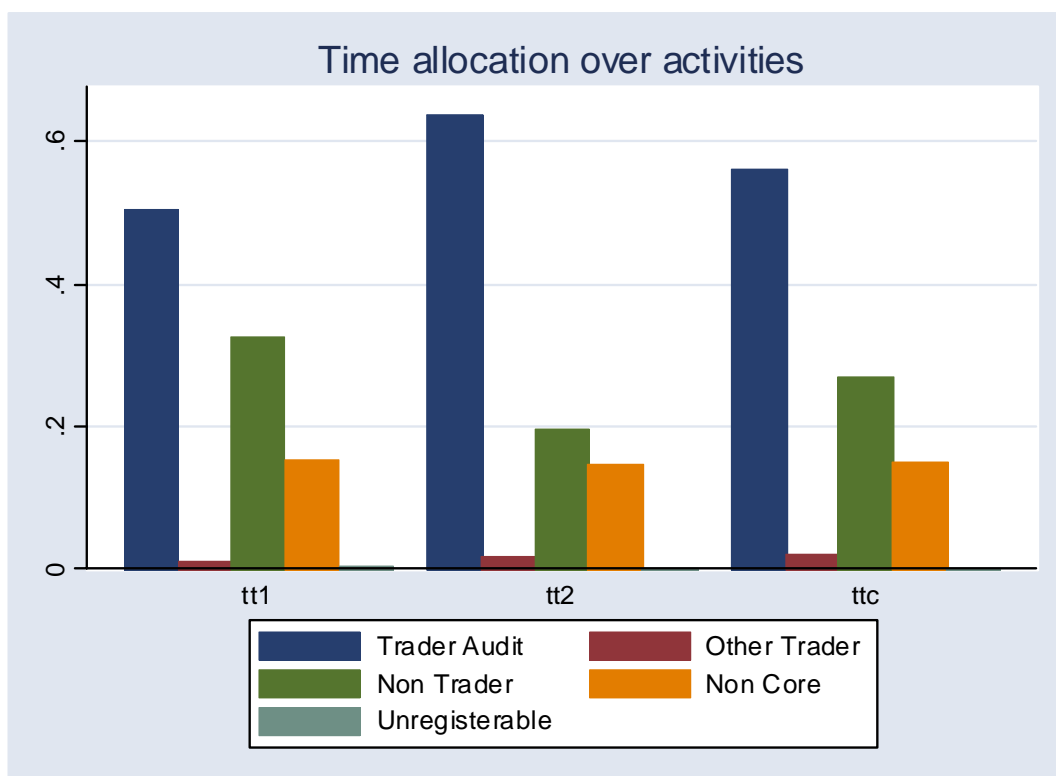

Time allocation over activities
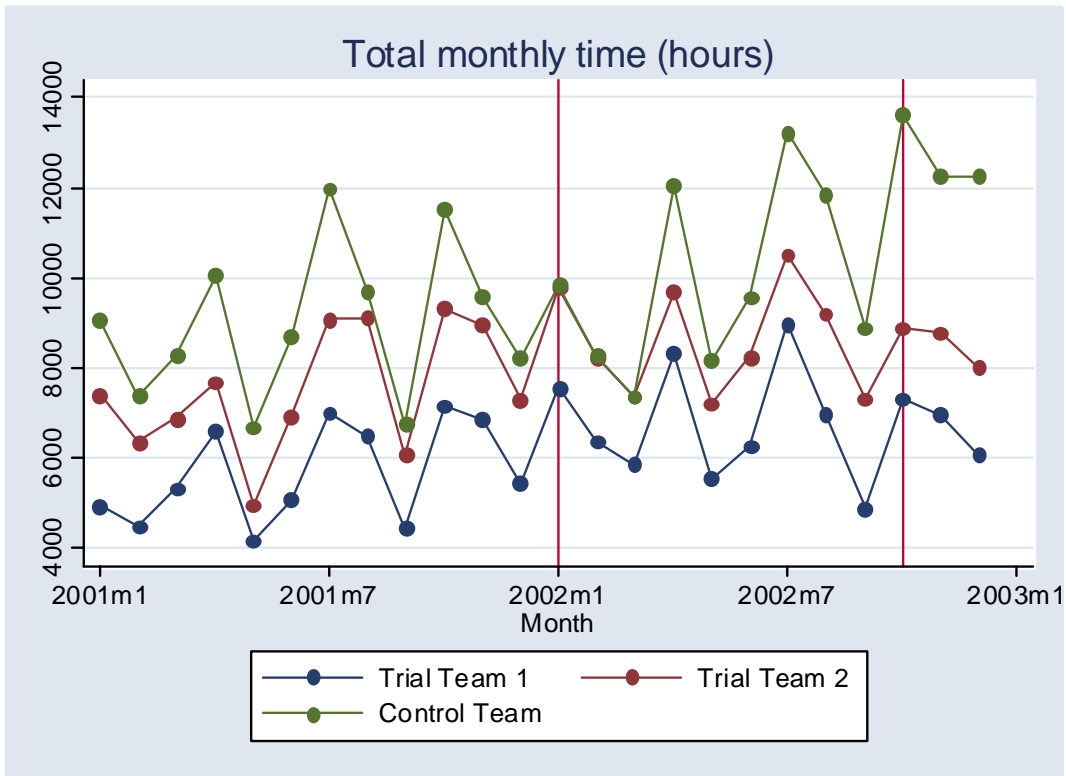
Figure 4: Time series variation in time.

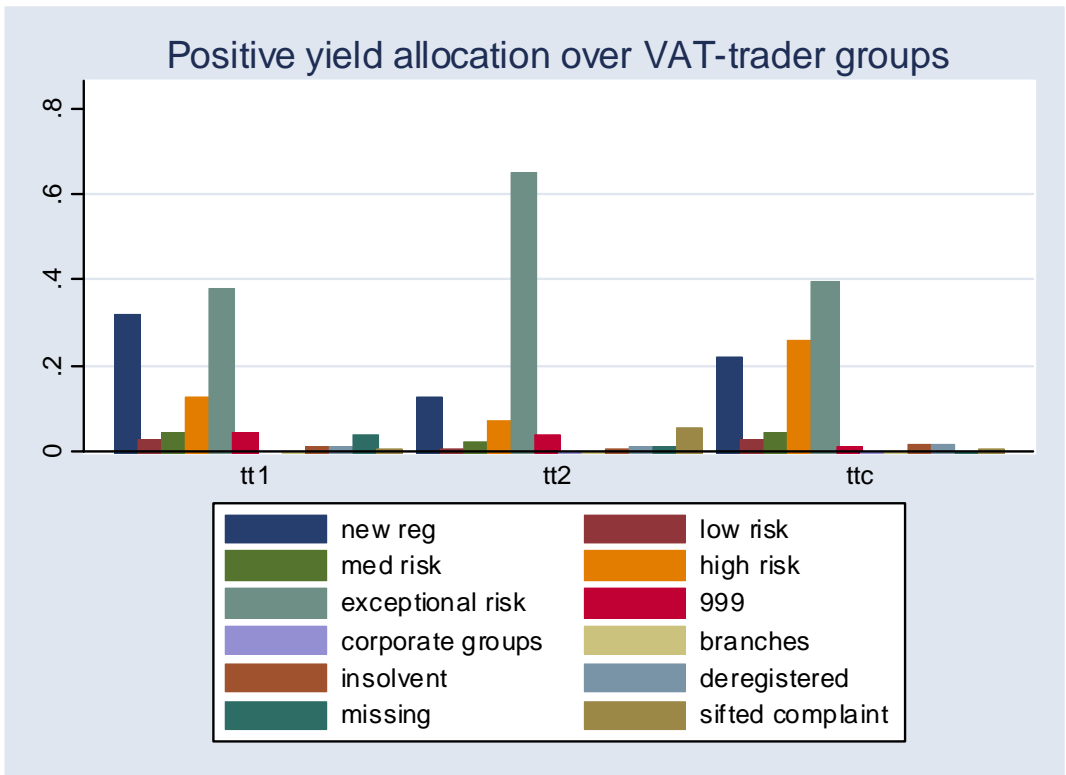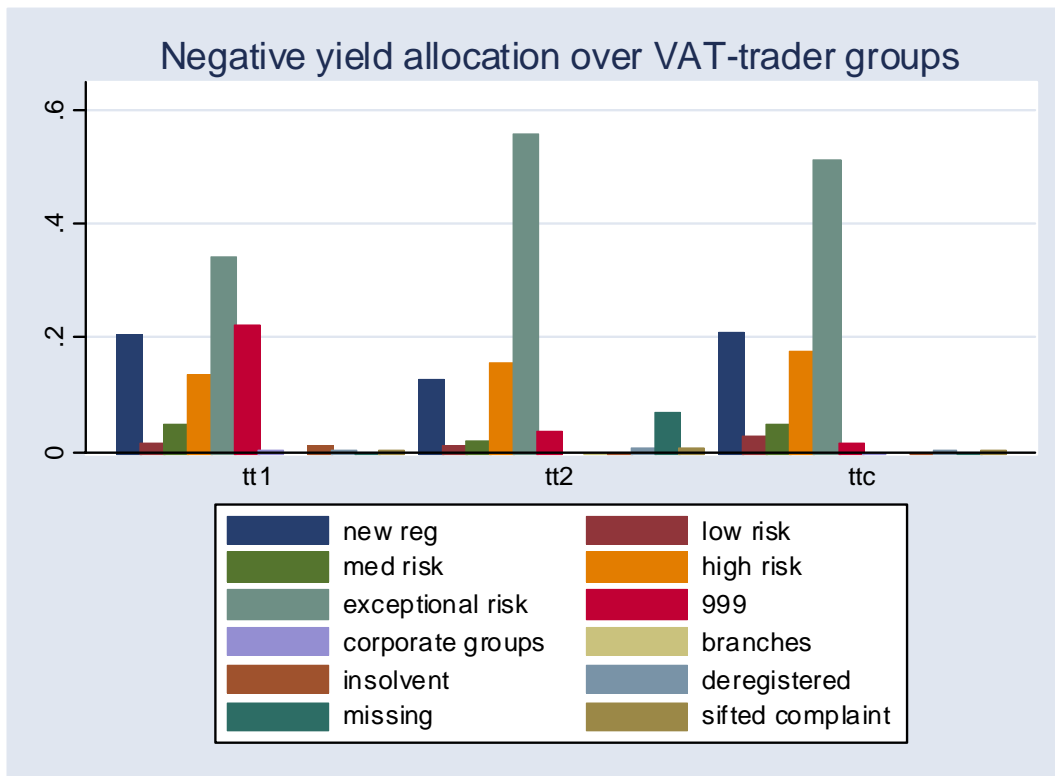Figure 5 Allocation of positive yield across VAT trader groups.

Figure 6 Allocation of negative yield across VAT trader groups.

Appendix A: A brief review of the economic theory of team incentives

A general insight from the economic literature on the theory of incentives in teams is that the effectiveness of group rewards on individual and group performance depends on the type of the organisation where the scheme is implemented and on the characteristics of its production process. More precisely, there are several aspects related to the type of the organisation and to how output is produced and measured that need to be considered in designing the optimal incentive scheme. The optimal incentive scheme is that, which delivers the desired effects on performance with the minimum cost for the designer.

In this section, we briefly review the theory of incentives in teams. We focus on the predictions of how the optimal reward scheme should be designed and on the characteristics of public sector agencies that have an impact on the delivery of incentives. Given this, we identify the points of strength and weaknesses of the incentive scheme designed at HM Custom and Excise.

The use of incentive schemes in teams

Incentive schemes are used whenever there are two parties writing a contract and one has imperfect information on some elements/aspects that have an important impact on the outcome of the transaction but cannot be contracted for. One example, which is relevant for our context, is the case of moral hazard. Problems of moral hazard arise in situations in which one individual (the principal) hires another to take an action for him/her as the "agent" and the latter has private information on the action (e.g. effort) (s)he chooses.

If the action were observable, the contracting problem between the two individuals would be relatively straightforward and the contract would simply specify the exact action to be taken by the agent. However, when effort is not observable by the principal, the compensation scheme has to be designed in a way that indirectly gives the agent the incentive to take the correct action. In the case of individual production, paying an individual the full value of his output will induce the efficient level of effort, in that final output, if perfectly measured as in the case of a monetary outcome, provides a good indicator of the agent's effort. The incentive to exert the desired level of effort is provided by linking the agent's compensation to his performance. An incentive scheme can then be defined as a reward system where part or all of the reward depends on performance. The power of the incentive scheme is measured by how much the compensation

depends on final output. High-powered incentives schemes are those schemes where the proportion of the reward – which depends on performance – is predominant.

How precisely performance is measured is of paramount importance for the effectiveness of an incentive scheme. When a precise measure of output is available, the principal can better infer the agent's effort from the output and hence can set higher incentives.

Delivering incentives in a team is more complex than in the case of individual output. In a team, final output is made up by all team members' contributions, and hence even if the principal observes final output perfectly, it is not a good indicator of each agent's inputs. Holmström (1982) provides one of the seminal contributions to the theory of incentives in teams. He shows that, in a setting where team members depend on each other to produce final output, i.e. where there are complementarities in production, if all the output of the team is shared among team members, team members are induced to free ride.

The intuition behind this idea is that paying an amount equal to total output to team members creates a negative externality for the team. If output is fully shared among team members, when an agent decreases her contribution, the value of total output will decrease and the sum of all agents' shares will decrease. Hence the agent who cheats will not pay in full for the consequences of his/her act. The cost of one person's shirking (in terms of the share of lower joint output) will be passed onto the others. The private marginal cost of shirking will be less than the social marginal cost (borne by all members of the team) and the level of effort chosen by the individual will be lower than the Pareto efficient level. Intuitively, this free-rider problem becomes greater in large organisations.

Holmström shows that it is possible to solve this free-rider problem by choosing an appropriate reward system. The optimal system of rewards in teams depends on how precisely output can be measured. First, if final output is perfectly measurable, a first best solution is attainable by imposing a system of group penalties whenever output falls below the desired level. This acts as an incentive to exert efficient levels of effort. Second, if output is measured with error (this will occur whenever output is not a monetary outcome) the size of the team becomes relevant for designing the optimal compensation scheme. Defining the optimal reward scheme is easier in small teams, where each agent has a substantial impact on the probability distribution of final output. In this case it is possible to elicit the appropriate level of effort through a reward system similar to the certainty case. In a large team, the contribution of each worker to joint production is less clear to identify and financial rewards are not enough to promote the desired action. The greater the uncertainty in output measurement and the greater the size of the team, the more

complex is the design of an optimal incentive scheme and explicit (financial) rewards will not be enough to get the desired actions. Some form of monitoring will become necessary.

<u>The use of team-based rewards in the absence of complementarities in production</u>

The Holmström analysis focuses on teams where agents' contributions to final output are interdependent. However, team-based rewards can also be used when the contribution of team members can be separately observed, so that there are no complementarities in production, but there are some positive aspects of teamwork that one might wish to promote.

Team-based rewards can be used to foster *co-operation* among team members. Milgrom-Roberts (1991) suggest that if co-operation within a group of individuals is important for the overall organisational objectives, then rewarding individual performance can detract from team performance by raising the marginal cost of effort in co-operating. Itoh (1991) analyses the relationship between financial incentives and 'helping' effort. He addresses the issue of whether it is always the case that, in moving from an individual based contract (i.e. one where individuals are paid only for their own output) towards one where rewards are based on teamwork, agents are induced to increase the level of helping effort. He finds that whether co-operation can be induced through financial rewards depends on the strategic interactions among agent's attitudes towards performing multiple tasks. In particular, agents can be induced to provide help, even for a small change in the wage schedule, if they get positive benefit from both types of effort. If, instead, tasks are similar and agents only care about the total amount of effort, they are reluctant to provide even a small amount of help. In this case a large perturbation of the individual-based contract is required to induce any helping effort from the individual.

Team based rewards may be used to benefit from the positive externalities of teamwork. Teamwork may facilitate a process of *communication and sharing of job experience*. It may also induce *peer monitoring*. If members of a team work in the same location and the organisation of work is such that they are able to observe each other, if their reward is linked to the team performance, they are more inclined to monitor how their peers are performing. This can help in enforcing proper levels of effort and tackle the free-rider problem. Additionally, teamwork can help insulate individuals against poor outcomes beyond their control as rewards are shared at a level bigger than the individual - *risk pooling*.

Which of these is more important depends on the organisation and the nature of output. In the context of HM Customs and Excise the positive externality of risk pooling created by the use of team-based rewards may be relevant for the yield outcome, for which the uncertainty of getting

back the unpaid tax may be quite substantial, so that the final outcome may be hardly under the control of the officer.

Delivering incentives in the public sector

There are some characteristics of public sector agencies that play an important role in the design of an incentive scheme. Dixit (2000) provides an overview of the differences between the private and public sector and how these impact on the optimal incentive scheme. Here we briefly mention those that are relevant for HM Customs and Excise.

Measuring performance in the public sector

An important aspect in designing the optimal reward system is how easy output is to measure. In the public sector output is often vaguely measured. This means that the information on the actions of agents is quite difficult to infer from the available measure of performance. The theory suggests in this case that weaker incentive schemes should be used.

Multi-tasking

Public sector agencies are complex organisations and are generally required to deliver a range of outcomes. This has an important impact on the incentive scheme. In particular, the interaction among the different tasks affects the power of the incentive scheme: if actions are substitutes the use of high-powered incentive schemes may have undesirable effects on overall performance. This is because exerting more effort on one task increases the marginal cost of any task that is a substitute and the agent can end up neglecting some tasks. In this case each outcome cannot be rewarded in isolation and lower powered incentives should be used.

An interesting case arises when activities are substitutes from the perspective of the agents (more time spent on one activity means less time on others), but they are complements from the perspective of the principal (the principal wants high performance in all of them). Hence the agent is willing to devote more time to the less difficult activities, whereas the principal prefers him to devote time to all activities. Marx and MacDonald (2001) show that if the principal is unsure about the agent's preferences over tasks, setting rewards on success on individual tasks may be sub optimal in that it may induce workers to focus and specialise in the less costly tasks. Results suggest that in this situation the system of reward should be non-monotonic, in that it defines different rewards according to the observed failure, partial success or full success on all tasks. The authors show that it is useful to reward failure on all tasks to some degree since this reduces the risk the agent has to bear for spreading his time across multiple activities. Of course

the reward for overall failure cannot be too large in that this outcome is easily achieved by the agent. Likewise the payment of success for achieving a subset of the tasks shouldn't be too large to avoid giving the agent an incentive to specialise only on some tasks. To avoid adverse specialisation it is typically optimal to reward no success more than partial success. In order to make general success attractive to the agent, since success on multiple activities is difficult to achieve, the compensation for succeeding must be very high. In conclusion, in an optimal contract, full success and only full success is rewarded highly and little specific compensation is awarded for each task.

Low powered incentive schemes should also be used when the different outcomes are measured with different errors. If each outcome could be rewarded in isolation, then the optimal incentive scheme would set higher incentives on the more easily measurable outcomes, as they provide a more accurate indicator of the effort exerted by the agent. However, in a context where there are multiple dimensions of output, this would make the agent concentrate on the tasks which are more accurately measured. Therefore the principal has to weaken the incentives on the more accurately measured tasks.

In conclusion, in a multitasking environment higher incentive schemes may be vulnerable to gaming by agents. This needs to be given particular weight in the public sector where principals may face more risk than in the private sector due to the fact that (a) it is more difficult to diversify the risks of bad outcomes of public policies and (b) there are critical threshold levels of public tolerance for failure by politicians. So politicians and senior civil servants can be very risk averse (in contrast with the assumption of the standard model where principals are risk neutral and do not have any concern for risk). The consequence of this is that politicians and senior civil servants may be *more* inclined to use high powered incentive schemes rather than setting low powered incentives schemes.

Appendix B: Statistical Approach

Definition of Productivity

Productivity for officer *i* on trader group *j* is defined as the sum of yield (in a 9 months – *m-* period) from trader group *j* over time spent (in the 9 months) on trader group *j*, weighted by share of time spent on trader group *j*.

By summing up across all trader groups dealt with by officer *i* we get the productivity for officer *i*:

$$productivity_i = \sum_{TG=1}^{j} \left[\left(\frac{\sum_{m=1}^{9} yield_{i,TG_j,m}}{\sum_{m=1}^{9} time_{i,TG_j,m}} \right) \times \left(\frac{\sum_{m=1}^{9} time_{i,TG_j,m}}{\sum_{TG=1}^{j}\sum_{m=1}^{9} time_{i,TG_j,m}} \right) \right]$$

Difference in difference approach

The aim of our evaluation is to examine whether the use of team-based incentive scheme raised productivity, and if so, **how** did any improvement come about. We adopt a "difference in difference approach". This compares the change in outcome **for each individual** (*i*) in the treatment group with the change in outcome for each individual (*j*) in the control group, at time *t* (when the scheme was implemented) and the period before (*t-1*):

$$\underbrace{\left(output_{i,t} - output_{i,t-1} \right)}_{\text{Treatment group}} - \underbrace{\left(output_{j,t} - output_{j,t-1} \right)}_{\text{Control group}}$$

Cell Matching

16 individual cells were created according to officer characteristics like part-time/full time status, gender, age and pay. The cells were crated based upon the treatment team status. We first divided officers into groups according to part-time/full time status and gender. Then we calculated the median of treatment Team 1 for each of the 4 groups. Individuals were split according to age below the median or age above or equal to the median. In a similar manner we calculated the median pay for these 8 groups and the individuals were split into 16 groups.

All trader groups

Cell	Team 1		Team 2		Control		Total
	#	%	#	%	#	%	#
1	19	14.73	21	18.26	33	17.74	73
2	19	14.73	10	8.7	17	9.14	46
3	19	14.73	17	14.78	19	10.22	55
4	19	14.73	23	20	28	15.05	70
5	7	5.43	4	3.48	15	8.06	26
6	7	5.43	9	7.83	22	11.83	38
7	7	5.43	7	6.09	14	7.53	28
8	8	6.2	14	12.17	12	6.45	34
9	0	0	0	0	1	0.54	1
10	1	0.78	0	0	0	0	1
11	0	0	0	0	0	0	0
12	1	0.78	0	0	0	0	1
13	5	3.88	1	0.87	4	2.15	10
14	6	4.65	4	3.48	8	4.3	18
15	5	3.88	3	2.61	9	4.84	17
16	6	4.65	2	1.74	4	2.15	12
Total	129		115		186		430

Cell = 1 for male full time younger low pay workers
Cell = 2 for male full time younger high pay workers
Cell = 3 for male full time older low pay workers
Cell = 4 for male full time older high pay workers
Cell = 5 for female full time younger low pay workers
Cell = 6 for female full time younger high pay workers
Cell = 7 for female full time older low pay workers
Cell = 8 for female full time older high pay workers
Cell = 9 for male part time younger low pay workers
Cell = 10 for male part time younger high pay workers
Cell = 11 for male part time older low pay workers
Cell = 12 for male part time older high pay workers
Cell = 13 for female part time younger low pay workers
Cell = 14 for female part time younger high pay workers
Cell = 15 for female part time older low pay workers
Cell = 16 for female part time older high pay workers

Regression analysis

We run the following regressions:

$$\Delta time = \beta_0 + \beta_1 Dummy + \beta_2 Team2 + \beta_3 TeamC + \beta_4 Years\_in\_grade +$$
$$\beta_5 Age + \beta_6 Age^2 + \beta_7 Gender + \beta_8 Parttime + \beta_9 Job\_band + \sum \beta_j TG_j$$

$$\Delta Yield = \beta_0 + \beta_1 Dummy + \beta_2 Team2 + \beta_3 TeamC + \beta_4 Years\_in\_grade +$$
$$\beta_5 Age + \beta_6 Age^2 + \beta_7 Gender + \beta_8 Parttime + \beta_9 Job\_band + \sum \beta_j TG_j$$

$$\Delta PositiveYield = \beta_0 + \beta_1 Dummy + \beta_2 Team2 + \beta_3 TeamC + \beta_4 Years\_in\_grade +$$
$$\beta_5 Age + \beta_6 Age^2 + \beta_7 Gender + \beta_8 Parttime + \beta_9 Job\_band + \sum \beta_j TG_j$$

$$\Delta NegativeYield = \beta_0 + \beta_1 Dummy + \beta_2 Team2 + \beta_3 TeamC + \beta_4 Years\_in\_grade +$$
$$\beta_5 Age + \beta_6 Age^2 + \beta_7 Gender + \beta_8 Parttime + \beta_9 Job\_band + \sum \beta_j TG_j$$

$$\Delta productivity = \beta_0 + \beta_1 Dummy + \beta_2 Team2 + \beta_3 TeamC + \beta_4 Years\_in\_grade +$$
$$\beta_5 Age + \beta_6 Age^2 + \beta_7 Gender + \beta_8 Parttime + \beta_9 Job\_band + \sum \beta_j TG_j$$

where $\Delta time$ is the difference in time allocated to trader groups by an officer in the period of the incentive scheme relative to the pre-scheme period, the dummy variable equals one for the incentivised trader groups in the incentivised period and zero otherwise. We control for team membership (Team2 and TeamC variables, where C stands for Control), officer individual characteristics - in particular age, gender part-time status and job band - and type of trader group (TG).

$\Delta Yield$ is the difference in yield collected by an officer in the period of the incentive scheme relative to the previous period. The same applies to the other outcomes. We take the difference between the values of each outcome in the two periods in order to eliminate the officer fixed effect, i.e. to control for any unobserved characteristics of the officer, like ability, which would impact on the outcome, but for which we do not have any data.

The assumption we are making is that output is affected by personal observable characteristics of the officer, the type of trader groups visited, the type of team the officer is in (which reflects the local market) and the incentive scheme.

They key parameter of interest is β_1 - the effect of the Makinson incentive scheme on output.