Zhixin Dai, Fabio Galeotti, Marie Claire Villeval

The Efficiency of Crackdowns: A Lab-in-the-Field Experiment in Public Transportations

Zhixin Dai, Fabio Galeotti, Marie Claire Villeval

Abstract:

The concentration of high frequency controls in a limited period of time ("crackdowns") constitutes an important feature of many law-enforcement policies around the world. In this paper, we offer a comprehensive investigation on the relative efficiency and effectiveness of various crackdown policies using a lab-in-the-field experiment with real passengers of a public transport service. We introduce a novel game, the daily public transportation game, where subjects have to decide, over many periods, whether to buy or not a ticket knowing that there might be a control. Our results show that (a) concentrated crackdowns are less effective and efficient than random controls; (b) prolonged crackdowns reduce fare-dodging during the period of intense monitoring but induces a burst of fraud as soon as they are withdrawn; (c) pre-announced controls induces more fraud in the periods without control. Overall, we also observe that real fare-dodgers fraud more in the experiment than non-fare-dodgers.

Working Paper

The Efficiency of Crackdowns: A Lab-in-the-Field Experiment in Public Transportations

Zhixin Dai[*], Fabio Galeotti[*], and Marie Claire Villeval[*†#]

[*] Université de Lyon, Lyon, 69007, France
CNRS, GATE Lyon Saint-Etienne, Ecully, 69130, France
[†] IZA, Bonn, Germany
[#] Department of Public Finance, University of Innsbruck

Abstract: The concentration of high frequency controls in a limited period of time ("crackdowns") constitutes an important feature of many law-enforcement policies around the world. In this paper, we offer a comprehensive investigation on the relative efficiency and effectiveness of various crackdown policies using a lab-in-the-field experiment with real passengers of a public transport service. We introduce a novel game, the daily public transportation game, where subjects have to decide, over many periods, whether to buy or not a ticket knowing that there might be a control. Our results show that (a) concentrated crackdowns are less effective and efficient than random controls; (b) prolonged crackdowns reduce fare-dodging during the period of intense monitoring but induces a burst of fraud as soon as they are withdrawn; (c) pre-announced controls induces more fraud in the periods without control. Overall, we also observe that real fare-dodgers fraud more in the experiment than non-fare-dodgers.

Keywords: Crackdowns, fraud, risk, monitoring, transportation, field experiment

JEL Classification Codes: C91, D83, K42

1. Introduction

A crackdown – i.e. the concentration of high-frequency or systematic controls in a limited amount of time and/or on a delimited geographical area or subset of the population – is a type of policy intervention widely used by law-enforcement institutions to fight against crime (Sherman, 1990; Di Tella and Schargrodsky, 2002, 2004; Kleiman, 2009; Eeckhout *et al.*, 2010; van der Weele, 2012). It was one of the most important policing policies in the U.S (Sherman, 1990) as well as in China in the 1980s (Schultz, 1989), and has still great relevance worldwide. Its applications include tax audits (Slemrod *et al.*, 2001; Kleven *et al.*, 2011), traffic enforcement (Eeckhout *et al.*, 2010), corruption and theft (Di Tella and Schargrodsky, 2003, 2004), drug markets (Farrell and Thorne, 2005), and urban violence (Moeller, 2009). It is also frequently used in public transportations to discourage fare-dodging.

Several economic theories have modeled the advantages of employing crackdown policies (Lando and Shavell, 2004; Lazear, 2006; Eeckhout *et al.*, 2010). These studies show that if the relationship between crime and expected punishment is concave, crackdowns will outperform a randomly distributed control strategy; if it is convex, then the opposite will hold. These models also demonstrate that crackdowns should be pre-announced in order to maximize their deterrent effect, especially to modify the behavior of those who have the highest cost of effort. From an empirical perspective, it is however hard to measure the shape of the relationship between crime and expected punishment using field data. In addition, the evaluation of pre-announced crackdowns is usually based on their immediate effects while they may also have effects in the long run. These long-term effects may be deleterious if more crimes are induced as soon as a crackdown ends; they may be positive if a high risk of control has an educative effect.

Empirical studies, such as Kastlunger *et al.* (2009), Di Tella and Schargrodsky (2003, 2004), Dai *et al.* (2015c), Banuri and Eckel (2015) offer mixed evidence on the effectiveness and efficiency of crackdowns. Moreover, they do not investigate the

relative efficiency and effectiveness of crackdowns to alternative mechanisms of crime prevention such as non-concentrated random controls keeping everything else constant.[1] Without a suitable comparison test, it is difficult to discern the real capacity of crackdown policies from these studies.

To fill this gap, we conducted an artefactual field experiment (Harrison and List 2004) using passengers of the public transport service in Lyon, the third largest city in France. We look at small-scale cheating of ordinary people which, quite contrary to what one would expect, has the largest social and economic impact for society (Bucciol *et al.*, 2013; Mazar and Ariely, 2006). More precisely, we investigate fare evasion.[2] According to some recent statistics, fare-dodging caused € 300 million loss to the French national railway company and € 100 million to the Parisian public transport operator in 2013 (Lefigaro.fr, 2013). It is not a problem circumscribed to France but it is widespread in all countries (Bonfanti and Wagenknecht, 2010). It is thus important to determine which and to what extent crackdown policies are effective and efficient to fight against fare evasion.

To investigate the relative efficiency and effectiveness of various crackdown policies we propose a novel game called the *daily public transportation game*. In this laboratory game, participants have to decide repeatedly (i.e. over many periods) whether to buy or not a (experimental) bus ticket knowing that there is a risk of control. We study how compliance behavior evolves over time depending on subjects' exposure to either repeated controls concentrated in a given period of time (crackdown) or irregular controls spread out over time. We analyze behavior during and after periods of control to investigate the immediate/short-term and long-term effects of these policies. We also test whether it is more effective to introduce

---

[1] These studies either did not explicitly compare the efficiency and effectiveness of a crackdown policy to an alternative mechanism of crime prevention such as random controls (Di Tella and Schargrodsky, 2003, 2004; Banuri and Eckel, 2015) or they vary at the same time other things than just the concentration of controls (Dai *et al.*, 2015c; Kastlunger *et al.*, 2009). For instance, Dai *et al.* (2015c) also vary the total number of audits, while Kastlunger *et al.* (2009) indirectly change subjects' expectation of an audit by informing them in advance of the overall probability of being audit.
[2] The cost of using a public vehicle is, in general, quite small (the price of a ticket is € 1.70 in the city where this study was conducted). It thus seems appropriate to classify fare dodging as a case of small-scale cheating.

crackdowns at an early or later stage of the game, and whether crackdowns should be pre-announced ex ante or not (i.e. whether or not subjects should be informed in advance of a control).

With respect to the latter comparison, our experiment also contributes to the literature on the efficiency of sanction mechanisms under ambiguity.[3] Indeed, when crackdowns are not pre-announced, subjects do not know whether there will be a control or not and what is the probability of its occurrence (i.e. condition of uncertainty). We therefore contribute to the study of the effects of ambiguity on behavior in the loss domain, whereas the impact of ambiguity is much more frequently studied in the gain domain. In contrast, when crackdowns are announced in advance, subjects are fully informed about whether a control will occur (i.e. condition of certainty). By comparing these alternative conditions, we can investigate whether uncertainty favors compliance or not.

An additional originality of our study is that we ran our experiment with a non-standard subject pool consisting of passengers of the public transport. This enables us to test different crackdown policies with a representative sample of the people for which these policies can be directly relevant.[4] In addition, we are able to identify real fare-dodgers by verifying whether subjects can present a validated ticket at the end of the experiment and by asking subjects to self-report how often they travel without a valid ticket in their real life. This allows us to compare the behavior of actual fare-dodgers and non-fare-dodgers in our game, and to test whether our game captures important dimensions of actual behavior in the field. Hence, our study also contributes to the general discussion about whether experimental results can be generalized to the real world (for a recent overview on this topic, see Fréchette and Schotter, 2015).

---

[3] See, for example, Dai *et al*. (2015c) for a discussion on related studies.
[4] To strengthen the link of our experiment with the real world, we also frame the experimental setting in terms of the natural environment where these subjects take decisions about public transportation. This also facilitated the comprehension of the instructions for participants who possessed very different socio-economic and educational backgrounds.

To give a very brief preview of our findings, in the presence of ambiguity on audits (and thus on sanctions), random controls which are spread out over time are more effective at reducing fare-dodging in the laboratory than crackdowns, and this effect is even stronger for real-world fare-dodgers. We also find that a more prolonged crackdown reduces fare-dodging during the period of intense monitoring but induces a burst of fraud as soon as it ends. Furthermore, pre-announced controls induce more fraud in the periods without control. Compared to non-concentrated random controls, crackdowns have a positive impact on compliance but only in the short run. However, they entail potentially deleterious effects in the long run since, after a crackdown is over, fraud rate increases to levels that are even higher than before the introduction of the crackdown even when crackdowns are not pre-announced and individuals do not know the total number of audits. Finally, we find that real fare-dodgers cheat more in all treatments compared to non-fare-dodgers.

In the remainder of our paper, Section 2 surveys the related literature. Section 3 describes the experimental design and procedures. Section 4 reports the experimental findings while Section 5 discusses the results and concludes.

2. Related studies

Theoretical studies have shown the advantages of crackdowns (Lando and Shavell, 2004; Lazear, 2006; Kleiman and Kilmer, 2009; Eeckout *et al.*, 2010). These papers mainly focus on when crackdowns are optimal to use. The models by Lazear (2006) and Eeckhout *et al*. (2010) are more germane to our experimental setting as they focus on systematic controls in a limited amount of time or delimited geographical area.[5] Their results rely on the shape of the relationship between crime and expected punishment. In particular, concentrating the enforcement resources in a specific area or period of time is predicted to generate a greater reduction in crime than spreading the same resources over a broader area or longer period of time but only when the relationship between crime and expected punishment is concave. When the opposite is

---

[5] Lando and Shavell (2004) and Kleiman and Kilmer (2009) are more interested on the implications of concentrating sanctions on a subset of violators rather than a period of time or geographical area.

true (i.e. the relationship is not concave), a less concentrated control policy is more effective. It is however difficult to test these models empirically since the shape of the relationship between crime and expected punishment cannot be easily identified in real settings. Our results can indirectly shed light on this relationship: if crackdowns are more efficient than less concentrated controls, we may infer that the relationship between crime and expected punishment is concave.

Some empirical studies have investigated the efficiency of crackdown policies. Di Tella and Schargrodsky (2003) showed that the prices paid by hospitals for basic and homogenous inputs decreased by 15% during the first 9 months of a crackdown. In a natural experiment, Di Tella and Schargrodsky (2004) found a large deterrent effect on car thefts once police protection had increased after a terroristic attack. In these two studies, the measure of efficiency is however imprecise for several reasons. First, the total enforcement input is unknown. Second, there may be other changes during the crackdown period that may affect behavior. For example, in Di Tella and Schargrodsky (2003) the evolution of wages may also play a role in limiting corruption, and, therefore, the pure effect of the crackdown cannot be totally isolated (Banuri and Eckel, 2015).

We are only aware of three experimental studies explicitly related to crackdowns. They show that crackdowns impact, if anything, only in the short run, and have no effect or even a deleterious one in the long run. The first study by Kastlunger *et al.* (2009) examines the impact of various sequences of audits on tax compliance when subjects know the probability of audits but not their distribution over time. They show that the decrease of compliance after an audit is caused more by misperception of chance than by the willingness to repair a loss. In contrast, Dai *et al.* (2015c) investigate compliance under ambiguity when subjects ignore not only the distribution of audits over time but also the overall probability of audits. Their vehicle of research is a public good game with a centralized mechanism of punishment that sanctions, with a certain probability, subjects who contributed less than the average group contribution. The results indicate that even under ambiguity, the contribution

level is the same under a continuous audit regime (in which the –unknown-probability of being punished is equal to 1) compared to a random audit regime (in which the probability of being punished is equal to 1/3 but with a fine coefficient three times higher). Interestingly, the results also show that contributions were significantly lower in the former than in the latter regime after audits were completely removed. This suggests that when individuals are not informed on the risk of being audited, crackdowns may be less efficient than more irregular audits because subjects are less able to update their beliefs when audits are more irregular.

The third experimental study is Banuri and Eckel (2015). The authors use a bribery laboratory experiment to investigate whether a crackdown has long-term effects on compliance. Crackdowns were introduced in the second of three ten-period blocks. By comparing the corruption levels in the pre-crackdown block with the post-crackdown block, the authors found no significant difference. They therefore conclude that short-term crackdowns are largely ineffective in changing behavior in the long run.

Compared to the previous studies, the current study allows us not only to compare the relative efficiency and effectiveness of various crackdown policies but also to study the impact of reintroducing a crackdown, the length and the timing of a crackdown, and the consequence of combing crackdowns with random controls. In addition, it enables us to investigate the role of pre-announcement in both crackdown and random-control policies. On this respect, theoretical papers suggest that crackdowns should be pre-announced to deter illegal behaviors (Lazear, 2006; Kleiman and Kilmer, 2009; Eeckout et al., 2010). However, to the best of our knowledge, there is no systematic study that investigates the short-run and long-run implications of pre-announcing or not a crackdown.[6] Finally, our study enables us to

---

[6] We acknowledge that there are a few studies in the literature on tax evasion that investigate the effect of sending threat-of-audit letters (e.g. Slemrod et al., 2001; Kleven et al., 2011; Fiorio and Santoro, 2013). What is usually found is that informing people about the audit activity has a positive and statistically significant effect on tax fraud compliance. More relevant for us is the study of Alm et al. (2009) who investigated, by means of a lab experiment, different ways in which information regarding the audit activity can be transmitted ex-ante and ex-post to the taxpayers. They ran sessions where the

measure the impact of different law-enforcement policies using real-world subjects for whom these policies may be directly relevant in their everyday life.

3. Experimental design and procedures

3.1 Experimental design

The experiment is divided in five different parts. In the first part, we elicited participant's attitudes towards risk and ambiguity both in the domain of gains and in the domain of losses. In the second part, we measured subjects' tendency to behave dishonestly by asking them to privately roll a dice and report the outcome (a task similar to the one used in Fischbacher and Föllmi-Heusi, 2013). In the third part, we measured the tendency to behave dishonesty in a contextualized laboratory task which was described as a decision on whether to buy or not a ticket for using public transportation and covering different trip lengths.[7] The fourth part was what we refer to as the *daily public transportation game* which is the main focus of this paper and is described in detail below. The final part was a socio-demographic questionnaire which was administered to the subjects at the end of the experiment. A detailed description of the second and third parts, together with their main results, are reported in a companion paper (Dai *et al*., 2015b). No feedback or information about future tasks was given to subjects in parts 1-3. In addition, all subjects went through the same parts (in the same order), and, therefore, earlier tasks could not explain across-treatment differences in the *daily public transportation game*.

Elicitation of attitudes towards risk and ambiguity. We elicited the participants' risk attitudes using a task developed by Eckel and Grossman (2008) and adopting a similar visual representation as Eckel *et al.* (2012) and Cardenas and Carpenter (2013). Subjects made four consecutive lottery choices, first in the gain domain, then in the loss domain, next in the gain domain under ambiguity, and finally in the loss domain

---

audit probability was pre-announced and sessions where it was not, and found that compliance was on average lower in the former. However, differently from our experiment, they did not consider periods of intense control (i.e. crackdowns).

[7] We asked subjects to make several decisions under different audit probabilities and fines.

under ambiguity. Every time the choice was displayed in a new screen.[8] Each choice consisted in selecting the most preferred of six possible lotteries. Table 1 displays the six lotteries for both the choices in the gain and loss domain. When the risk was known by the subjects, each lottery involved a 50/50 chance of a low or a high payoff. Moving from Lottery 1 to Lottery 6, both expected return and risk (standard deviation) increase. Hence, subjects who are extremely risk-averse should choose the sure Lottery 1. Moderately risk-averse individuals should choose an intermediate lottery (Lotteries 2-4). Risk-neutral subjects should choose Lottery 5 or 6. Risk-seeking persons should choose Lottery 6. In the loss domain, lotteries could give negative payoffs. To prevent participants to make real losses in the experiment, we assigned subjects an initial endowment of € 40. This makes the expected payoffs equivalent in the loss and in the gain domains. Hence, the losses were only a framing manipulation. Under ambiguity, the payoffs were the same as in the choices under certainty but the probability of each outcome was uncertain (between 3/10 and 7/10).

[Table 1 about here]

The daily public transportation game. Each participant was initially endowed with € 35 to avoid a net loss in this game. To facilitate subjects' understanding of the experimental environment and enhance the external validity of our experimental setting in relation to the real-world environment where our subjects usually operated, we described the laboratory game as a choice on whether to pay or not a ticket for using a bus.[9] The game consists of 60 periods ("days") in total divided in 2 blocks of 30 days. For each day, participants were asked to decide whether or not buy a ticket to use a bus. The cost of a ticket was €1.70.[10] Subjects were told that a control of the ticket could occur in certain days. Subjects received no information on the number of audits and their distribution over time. If a participant was caught travelling without a

---

[8] The order in which subjects made their choices was kept fixed to facilitate decision-making and avoid confusion.

[9] We however refrain from using strong terminology which could lead to experimenter demand effects.

[10] This was also equivalent to the real cost of using a public transport in the French city where the experiment took place.

ticket, he or she had to pay a fine of € 10 plus the price of the unpaid ticket (€ 11.70 in total); this was made common knowledge. At the end of each experimental day, subjects received an immediate feedback on whether they had been audited (or not) and fined. However, this information does not remain visible on the computer screen once the subject moves to the next day. This means that subjects have to memorize the previous sequence of audits. At the end of the task, three periods (days) were selected at random for actual payment.

The experiment consisted of 8 treatments. These are summarized in Table 2. In all treatments, we employed the same number of audits (12 in total). In the Random-Random treatment (*RR*), 6 audits were randomly distributed over the first block (days 1 to 30), and other 6 over the second block (days 31 to 60). In the Crackdown-Crackdown treatment (*CC*), we concentrated 6 consecutive audits in each block. In particular, subjects were audited in days 3 to 8 and 33 to 38.[11] To break the regularity of a crackdown's occurrence, we ran the Crackdown-Random treatment (*CR*), which is similar to the CC treatment except that, in the second block, audits were spread out as in the RR treatment. In the Random-Crackdown treatment (*RC*), we reversed this order by having random audits in the first block and a crackdown in the second block.

[Table 2 about here]

In order to measure the impact of a longer crackdown policy, in the Early-Long-Crackdown treatment (*C_Long1*) we concentrated 12 consecutive audits in the first block, from day 3 to 14, and we applied no audit in the second block. To measure whether the timing of a crackdown matters, we also ran the Late-Long-Crackdown (*C_Long2*), where we applied the 12 consecutive audits in the second block from day 33 to 45.

As explained earlier, previous studies suggest that crackdowns should be

---

[11] The reason why we started the crackdown in day 3 and day 33 is that we did not want subjects to easily learn the occurrence of an audit.

pre-announced (Eeckhout *et al.*, 2010). This motivates our last two treatments: Random-Random with pre-announcement (*RR_A*) and Crackdown-Crackdown with pre-announcement (*CC_A*). The treatments RR_A and CC_A are similar to the RR and CC treatments. The only difference is that subjects were informed in advance each day (i.e. before making a decision) whether an audit would occur this day.

In our setting, if subjects knew the exact audit probability – which is 20% given 12 audits across 60 periods – buying a ticket would be a dominant strategy.[12] However, the audit probability is unknown. Rational and risk-neutral subjects should fraud if their perceived probability of audit is lower than 14.53%.

3.2 Experimental procedures

279 participants were recruited for a 45-minute experiment at the exit of the bus and tram in front of the main railway station of Lyon, the third largest city in France, between 11:00am and 6:30pm. Recruiters wore no uniform (to signal that they were not ticket inspectors). The recruiters invited passengers to participate in a study conducted by researchers of the university. People holding a monthly-pass were excluded (by definition, these persons are not fare-dodgers).

Recruiters informed passengers individually that the study would last between 45 and 60 minutes, they would earn between € 25 and € 75, on average € 50, and they would be paid immediately at the end of the experiment in vouchers which could be exchanged for goods and services in several places, including a near commercial center. They also indicated that all decisions were strictly anonymous and that no special knowledge was required. Once passengers agreed to participate, recruiters informed them to keep their ticket (if they had one) because it will be exchanged for a new free ticket at the end of the experiment to compensate them for the time spent in the lab.[13]

---

[12] If the audit probability was known, rational and risk-neutral agents should always choose to buy a ticket since the payoff would be higher than not to buy a ticket.
[13] After having been punched, a ticket was valid for one hour. Hence, it was entirely legitimate and credible to compensate a ticket-holding subject with a new ticket for the time spent in the lab.

We also recruited 35 subjects from the Fine Collection Office (FCO). These people did not pay immediately their fine after having being caught travelling without a ticket and they were issued a fine ticket to be paid later at the FCO. After they made their payment, recruiters used a similar procedure to invite them to participate in our experiment.

The experiment was conducted using our mobile laboratory in a building located near the recruitment point. Mobile partitions were used to isolate participants. Subjects made their decisions via tactile tablets. A maximum of 9 subjects participated at the same time in the laboratory. Since there was no interaction between participants during the experiment, subjects could start the experiment as soon as they arrived in the room.

Upon arrival and after signing a consent form, an experimenter randomly guided each participant to a workstation. Subjects were reminded that decisions were anonymous. All the instructions were displayed on the tablet. They were informed that they could raise their hand at any time and an experimenter would come to them and answer their questions privately.

Once a subject completed all the tasks, he or she was escorted to a separate payment room where he or she was paid in vouchers. Then we exchanged his or her validated ticket – if he or she had one – with a free day pass (providing a real incentive for exchanging a validated ticket).

Subjects on average completed the experiment in 37 minutes (12.81 of standard deviation) excluding the payment procedures. On average, participants earned about € 60 (20.59 of standard deviation). Earnings were calculated based on subjects' decisions in parts 1, 2 and either 3 or 4.

3.3 Identification of fare-dodgers

One major novelty of our study is the fact that we are able to measure the impact of different law-enforcement policies on actual fare-dodgers and non-fare-dodgers.

Therefore, it is important to describe how we identified fare-dodgers and non-fare-dodgers.

We collected three different measures of fare-dodgers. First, we categorized as fare-dodgers subjects who were recruited at the FCO. Second, fare-dodgers were those subjects who could not display a validated ticket at the end of the experiment. Finally, we classified as fare-dodgers those participants who, in the final questionnaire, self-reported to travel without a valid ticket at least once every 10 trips.[14] Note that we define a subject as fare-dodger if he or she fits in one or more of the three categories above. We believe this is a relatively reliable and comprehensive way to identify fare-dodgers for the following reasons.[15] First, we know for sure that subjects recruited at the FCO were fare-dodgers. Second, we can confidently say that subjects who could not display a valid ticket were fare dodgers since we recruited them when they were getting off the bus/tram and we explicitly told them to keep their ticket[16]. Hence, they were supposed to hold a validated ticket unless they fare dodged. Third, there is no reason why subjects who self-reported as fare-dodgers lied about this.[17] Of course, we recognize that our classification has some limitations and may not capture all the fare-dodgers who participated in our experiment. Some of them may have in fact bought the ticket on the day of the experiment and lied about their illegal behavior.[18] We should however point out that, if we accidentally classified some subjects as non-fare-dodgers, this would only strengthen our results if we find differences in behavior between the two groups (fare-dodgers and non-fare-dodgers).

To provide some summary statistics on our measures of fare-dodgers, excluding subjects from the FCO, 41.80% of the remaining participants could not show a

---

[14] In the questionnaire, they were asked to report how often (from 0 to 10) they travel without a valid ticket out of 10 trips.

[15] An alternative procedure would be that of relying only on one of the three measures. The risk is of course that of underestimate the number of fare dodgers.

[16] The lab was also near the tram and bus stop (about 100 meters from the stop).

[17] The opposite is likely to be true, that is subjects who self-reported as non-fare-dodgers were actually fare-dodgers.

[18] We believe these are more likely to be occasional fare dodgers.

validated ticket, 54.92% self-reported as fare-dodgers, 68.44% self-reported to be fare-dodgers and/or could not show a validated ticket, and 28.28% both self-reported as fare-dodgers and could not show a validated ticket.[19]

4. Experimental results

In this section, we first examine, both in aggregate and at individual level, the relative efficiency and effectiveness of our different treatments in terms of induced behavior, and whether fare-dodgers behave differently than non-fare-dodgers. We then compare the relative efficiency of the various crackdown policies in terms of revenue collected.

4.1 Relative efficiency and effectiveness of different treatments

Table 3 displays some summary statistics on the average fraud rate by treatment for actual fare-dodgers and non-fare-dodgers.

[Table 3 about here]

One first important observation is that actual-fare-dodgers fraud more than non-fare-dodgers in all treatments and this difference is significant except in the RR and C_Long1 treatments (see the Test column in Table 3). This finding is important as it suggests that our experiment has some external validity.

Table 3 also indicates that the average fraud rate differs across treatments. Straight comparisons of the fraud rates between treatments using non-parametric tests may, however, not be right. Since fare-dodgers behave differently than non-fare-dodgers, differences between treatments may in fact be explained by a different proportion of fare-dodgers rather than a change in behavior induced by the treatment manipulation.

Although the recruitment process was random, we indeed detect some variations in the participation of actual-fare-dodgers in each treatment, which may bias our

---

[19] These figures are consistent with data from the local public transport company (see, e.g., *Keolis*, 2014), and other studies conducted in comparable countries (Bucciol *et al.*, 2013). Our high percentages are also due to the fact that we excluded people with a monthly pass from our experiment.

measure of relative efficiency of crackdowns.[20] In order to control for this potential bias, we estimate a random-effects probit model with robust standard errors clustered at the individual level to account for the fact that the 279 players made 60 successive decisions. The dependent variable is a dummy which is equal to 1 if a subject chose to fraud on a given experimental day. Independent variables include a dummy variable for each treatment with RR taken as the reference category. The "Intensity Crackdown" variable indicates the consecutive number of audits occurred before the current day and captures the impact of crackdowns. We also control for whether a subject had been fined in the previous experimental day. In addition, we include measures of risk aversion in the gain and the loss domain with or without uncertainty.[21] Finally, we control for the behavior of fare-dodgers in two complementary ways. First, we include a dummy for fare-dodgers who did not display a valid ticket or self-reported as frauders. Second, we control, with a separate dummy, for the behavior of the fare-dodgers recruited at the FCO. We treat these latter fare-dodgers separately as their behavior is susceptible to be different from other fare-dodgers since they had just paid a fine.

[Table 4 about here]

In Table 4, Columns (1) and (2) report the results of the regression conducted on the data from the first block (days 1-30) and the second block (days 31-60) respectively, while Column (3) pools the data from the two blocks together. Columns (4-6) are identical to Columns (1-3) except that we only use data from actual fare-dodgers. In Columns (7-9), we only use data from non-fare-dodgers.

[Figure 1 about here]

To help discussing the main results of Table 4, we also plot the fraud rate over time for each of our different treatments (Figure 1). For the sake of readability, we pool some days together. Precisely, we divide each block into four parts:

---

[20] The number of actual fare-dodgers varies between 61% (C_Long1) to 79% (RR).
[21] The higher the value, the less risk averse the subject is.

pre-crackdown days (1-3, 31-33), crackdown days (4-9, 34-39), post-crackdown days (10-12, 40-42), and the remaining days (13-30, 43-60).[22] Crackdown days are identified with a vertical line. In each panel of Figure 1, we always display the baseline condition (i.e. the RR treatment).

Table 4 shows that the probability of fraud is 14.2% higher in the CC treatment compared to the RR treatment. This also appears in Figure 1 where the average fraud rate of the CC treatment is always greater or similar to the average fraud rate of the RR treatment except during the second crackdown. Figure 1 (Panel A) also shows that, in the CC treatment, subjects almost did not react to the first crackdown but only to the second one as if there was a learning effect at work. Alternating crackdowns with random controls does not make any difference compared to only random controls. As Table 4 (and also Figure 1) indicates, both CR and RC have no impact on the probability of fare dodging compared to RR. This leads to our first result.

Result 1: Crackdowns fail reducing overall fare evasion.

We next consider the impact of pre-announcing audits. Model (3) of Table 4 reveals that participants are more likely to fraud in treatment CC_A and RR_A compared to the baseline (RR). The probability is 22.8% and 14.6% higher, respectively. This is also confirmed in Figure 1 (Panel B) where the average fraud rate is generally higher in RR_A compared to RR, and CC_A to CC. The reason, which is well captured by Figure 1, is that pre-announcement makes people fare dodging more when there is no audit. This leads to our second result:

Result 2: Pre-announcement of audits decreases their overall effectiveness.

We can investigate the effect of having longer crackdowns by looking at C_Long1 and C_Long2. The probability of committing a fraud is significantly higher

---

[22] A crackdown ran from day 3 (33) until day 8 (38) but its effect operated from day 4 (34) until day 9 (39). In other words, the effect of a crackdown was one-day delayed. This is because subjects were not informed in advance of a crackdown. There is however an exception: in the CC_A treatment, the control was pre-announced and, therefore, the effect and execution of a crackdown coincided. For this treatment, crackdown days are 3-8 and 33-38. This is taken into account in Figure 3.

in C_Long1 (17.1%) compared to RR while there is no significant difference between RR and C_Long2 (see Table 4). Figure 3 (Panel C) shows that fraud increases considerably after the introduction of an early longer crackdown (C_long1 treatment) and, to a lesser extent, before the introduction of a late longer crackdown (C_long2 treatment). This leads to our third result:

Result 3: Longer crackdowns are overall less effective if they occur in the early stage of the experiment.

To measure, other things being equal, the short-run effect of a crackdown (i.e. the immediate effect of a crackdown once it is introduced), we can look at the variable "Intensity Crackdown", that is the number of consecutive audits over time. We find that the probability to fraud, on average, drops by 5.8% in the immediate run for each consecutive day with an audit. This means that frequent consecutive audits discipline people in the short run. In Figure 2, we plot the marginal effect of each consecutive day of audit on the probability of not buying a ticket. We can see that the effect is not linear. The drop is sharp in the first days and it softens in the following. Fraud decreases by about 50% after 3/4 days, and it is almost eradicated after 9 days (only 1% or less of fraud). We can also look at the effect of being fined in the previous day. Surprisingly, it has no additional impact on fraud. A possible explanation is that some people cheat more to recover from the loss while others cheat less because of the educative effect of the fine. The net effect is neutral.

The results of Table 4 also confirm our previous finding that actual fare-dodgers cheated more than actual non-fare-dodgers. In particular, self-reported fare-dodgers or non-ticket holders are much more likely to cheat than non-fare-dodgers (+39.7%). The same applies to fare-dodgers recruited at the FCO (+29.9%). We also find that all the previous results (i.e. treatment effects and short-term impact of crackdowns) are even stronger among actual fare-dodgers (Columns 4-6). In contrast, there are no treatment effects among non-fare-dodgers. This provides important support to the external validity of our experiment and leads to our fourth result:

17

Result 4: Actual fare-dodgers fraud more than non-fare-dodgers and are more sensitive to our treatment manipulations.

On a final note, Table 4 shows that the risk attitude of subjects – in different domains (gains or losses) and information sets (uncertain or certain probabilities) – does not explain behavior in our *daily public transportation game*. This suggests that the decision to act dishonestly is, in our setting, better explained by other factors, such as emotional, moral or social concerns, than simple cost-benefit considerations.

4.2 Relative efficiency in terms of payoffs

We now investigate the relative efficiency of the various audit policies from the point of view of the public transportation company by calculating the total revenue (paid tickets plus fines) that the company is able to collect. Figure 2 displays the total revenue for all our different treatments. Based on a Wilcoxon rank-sum test, we find no significant difference in revenue between treatments CR, RC, and RR ($p > 0.1$ for all possible comparisons). The revenue is instead significantly lower in treatments CC, C_Long1, C_Long2, CC_A, RR_A compared to RR ($p = 0.020$ for CC, and $p < 0.001$ for the other comparisons). This is in line with some of our previous findings and leads to our last result:

Result 5: Crackdowns do not increase the revenue of the law-enforcement institution.

[Figure 2 about here]

5. Conclusion

Our artefactual field experiment provides a comprehensive analysis of the relative efficiency and effectiveness of various crackdowns policies. The main finding is that crackdowns, alone, decrease overall effectiveness compared to random controls mostly because they induce a burst of fraud in the subsequent periods with no control. Only when they are alternated with random controls or protracted for a longer time at

the late stage of the game,[23] they achieve a level of overall compliance similar to a randomly distributed control policy. We also find that pre-announcing a control decreases overall compliance although it increases the effectiveness of a crackdown during its launch.

We believe all these results are very valuable for policy-makers who are contemplating which policy interventions to adopt against law-breaking behavior. In particular, our results suggest that under ambiguity in the domain of losses, systematic controls in a limited amount of time are, at best, as effective as random controls and potentially very inefficient in the long-run both in terms of achieved compliance and revenue for the law-enforcement institution.

One however should be careful in generalizing our results to any environment where crackdown policies may be relevant. Our study focuses on a specific field of application (i.e. public transportations), and look at a specific type of law violation (e.g. fare-dodging). It may be that crackdowns are very effective and efficient in other domains of application. This should be investigated in future studies.

Finally, although we try to be as comprehensive as possible by considering different and alternative policies of crackdown, there are still many grey areas which deserve future empirical investigation. For example, crackdowns may arise endogenously (Dai, 2015a). Quoting Van de Weele (2012, p. 417), "enforcement strength is not exogenous, but determined by feedback effects between society and enforcers and social processes within the enforcement agencies themselves". Crackdowns may also cause displacement effects, e.g. participants who are subjected to crackdowns will move to other areas during crackdown periods and thus decrease their efficiency. All these issues can be addressed in future studies.

---

[23] When the crackdown lasts longer and is localized towards the end of the game, subjects comply more during the period of no control as if they were expecting a control at any time.

References

Alm, J., Jackson, B. R., & McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, 93(3-4), 392–402.

Banuri, S., & Eckel, C. (2015). Cracking down on bribery. *Social Choice and Welfare*, 1-22.

Bonfanti, G., & Wagenknecht, T. (2010). Human factors reduce aggression and fare evasion. *Public Transport International*, 59, 28-32.

Bucciol, A., Landini, F., & Piovesan, M. (2013). Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization*, 93, 248-257.

Cardenas, J. C., & Carpenter, J. (2013). Risk attitudes and economic well-being in Latin America. *Journal of Development Economics*, 103, 52-61.

Dai, Z. (2015a). On the optimal design of crackdowns, theory and experimental evidence. *Mimeo*.

Dai, Z., Galeotti, F., & Villeval, M. C. (2015b). Cheating in the lab predicts cheating in the field! An experiment in public transportations. *Mimeo*.

Dai, Z., Hogarth, R. M., & Villeval, M. C. (2015c). Ambiguity on audits and cooperation in a public goods game. *European Economic Review*, 74, 146-162.

Di Tella, R., & Schargrodsky, E. (2003). The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires. *The Journal of Law and Economics*, *46*(1), 269–292.

Di Tella, R., & Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, *94*(1), 115–133.

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results*, 1, 1061-1073.

Eckel, C. C., Grossman, P. J., Johnson, C. A., de Oliveira, A. C., Rojas, C., & Wilson, R. K. (2012). School environment and risk preferences: Experimental evidence. *Journal of Risk and Uncertainty*, 45(3), 265-292.

Eeckhout, J., Persico, N., & Todd, P. E. (2010). A theory of optimal random crackdowns. *American Economic Review*, *100*(3), 1104–1135.

Farrell, G., & Thorne, J. (2005). Where have all the flowers gone? Evaluation of the Taliban crackdown against opium poppy cultivation in Afghanistan. *International Journal of Drug Policy*, 16 (2), 81–91.

Fiorio, C., Iacus S., & Santoro, A. (2013). Taxpaying response of small firms to an increased probability of audit: some evidence from Italy, *University of Milan Bicocca Working Paper 251*.

Fréchette, G. R., & Schotter A. (Eds.) (2015). Handbook of Experimental Economic Methodology. Oxford University Press.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, 1009-1055.

Kastlunger, B., Kirchler, E., Mittone, L., & Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology* 30, 405-418.

Keolis (2014). "Fraude: comment lutter?" Keo', February. Available online at http://www.keolis.com/fileadmin/documents/medias/publications/keo/KEO1402.pdf (accessed 13 October 2015).

Kleiman, M. (2009). When Brute Force Fails: How to Have Less Crime and Less Punishment. Princeton University Press, Princeton.

Kleiman, M., & Kilmer, B. (2009). The dynamics of deterrence. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(34), 14230–14235.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat?: evidence from a tax audit experiment in Denmark. *Econometrica*, *79*(3), 651–692.

Lando, H., & Shavell, S. (2004). The advantage of focusing law enforcement effort. *International Review of Law and Economics*, *24*(2), 209–218.

Lazear, E. (2006). Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, (8), 1029–1061.

Le Figaro (2013). " La fraude coûte 400 millions d'euros à la SNCF et la RATP". lefigaro.fr, 29 August. Available online at http://www.lefigaro.fr/conjoncture/ 2013/08/29/20002-20130829ARTFIG00252-la-fraude-coute-400-millions-d-euros-a-l a-sncf-et-la-ratp.php (accessed 13 October 2015).

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy and Marketing*, 25, 117–126.

Moeller, K.K. (2009). Police crackdown on Christiania in Copenhagen. *Crime, Law and Social Change*, 52 (4), 337–345.

Schultz, C. B. (1989). Economic Crimes in the People's Republic of China: A Swinging Door Policy. *American University International Law Review*, 5, 161-206.

Sherman, L. W. (1990). Police Crackdowns: Initial and Residual Deterrence. *Crime and Justice*, *12*(1990), 1.

Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, *79*(3), 455–483.

Van Der Weele, J. (2012). Beyond the state of nature: Introducing social interactions in the economic model of crime. *Review of Law and Economics*, *8*(2), 401–432.

Tables and figures

Table 1: Lotteries payoffs

	Event	Probability (%)		Payoff		Expected Payoff		Risk
		Certain	Uncertain	Gain domain	Loss domain	Gain domain	Loss domain	
Lottery 1	A	50	30-70	€ 28	€ −12	€ 28	€ −12	0
	B	50	30-70	€ 28	€ −12			
Lottery 2	A	50	30-70	€ 24	€ −16	€ 30	€ −10	6
	B	50	30-70	€ 36	€ −4			
Lottery 3	A	50	30-70	€ 20	€ −20	€ 32	€ −8	12
	B	50	30-70	€ 44	€ 4			
Lottery 4	A	50	30-70	€ 16	€ −24	€ 34	€ −6	18
	B	50	30-70	€ 52	€ 12			
Lottery 5	A	50	30-70	€ 12	€ −28	€ 36	€ −4	24
	B	50	30-70	€ 60	€ 20			
Lottery 6	A	50	30-70	€ 2	€ −38	€ 36	€ −4	34
	B	50	30-70	€ 70	€ 30			

Note: The level of risk is represented by the standard deviation of the expected payoff.

Table 2: Summary of the treatments

Treatment	Days 1-30	Days 31-60
Random-Random (RR)	Random audits	Random audits
Crackdown-Crackdown (CC)	Crackdown (days 3-8)	Crackdown days (33-38)
Crackdown-Random (CR)	Crackdown (days 3-8)	6 Random audits
Random-Crackdown (RC)	Random audits	Crackdown (days 33-38)
Early-Long-Crackdown (C_long1)	Crackdown (days 3-14)	No audit
Late-Long-Crackdown (C_long2)	No audit	Crackdown (days 33-44)
Crackdown-Crackdown with Pre-announcement (CC_A)	Crackdown (days 3-8)	Crackdown (days 33-38)
Random-Random with Pre-announcement (RR_A)	Random audits	Random audits

Note: The number of audits (12 in total) was constant across treatments.

Table 3: Summary statistics on the fraud rate

	All	**Fare-dodgers**	**Non-fare-dodgers**	**Test**
RR (n = 38)	0.28	0.29	0.24	NS
	(0.45)	(0.45)	(0.43)	
CC (n = 38)	0.36	0.44	0.14	***
	(0.48)	(0.50)	(0.55)	
CR (n = 36)	0.25	0.34	0.11	**
	(0.43)	(0.47)	(0.31)	
RC (n = 38)	0.29	0.35	0.06	***
	(0.45)	(0.48)	(0.23)	
C_Long1 (n = 31)	0.38	0.44	0.30	NS
	(0.49)	(0.50)	(0.46)	
C_Long2 (n = 36)	0.32	0.41	0.08	***
	(0.47)	(0.49)	(0.26)	
CC_A (n = 31)	0.49	0.54	0.23	**
	(0.50)	(0.50)	(0.42)	
RR_A (n = 31)	0.39	0.48	0.20	**
	(0.49)	(0.50)	(0.40)	
Tot. (n = 279)	0.34	0.34	0.41	***
	(0.47)	(0.47)	(0.49)	

Notes: The "Test" column informs whether the fraud rate is different between fare-dodgers and non-fare-dodgers, according to a Wilcoxon rank-sum test. Standard deviations are shown in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$, NS $p>0.1$.

Table 4: Random-effects probit model

Dependent variable: Fraud	All Observations			Fare-dodgers only			Non-Fare-dodgers only		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1 if Fraud, 0 if not)	1st Month	2nd Month	All	1st Month	2nd Month	All	1st Month	2nd Month	All
RR	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
CC	0.113	0.175**	0.142*	0.183*	0.282***	0.225**	0.006	0.007	0.014
	(0.080)	(0.080)	(0.079)	(0.103)	(0.103)	(0.103)	(0.046)	(0.052)	(0.038)
CR	-0.025	0.022	-0.013	-0.005	0.078	0.021	-0.016	-0.020	-0.015
	(0.083)	(0.084)	(0.082)	(0.112)	(0.115)	(0.113)	(0.044)	(0.049)	(0.034)
RC	0.034	0.042	0.040	0.096	0.123	0.110	-0.026	-0.028	-0.018
	(0.081)	(0.082)	(0.080)	(0.102)	(0.104)	(0.102)	(0.054)	(0.060)	(0.041)
C_Long1	0.118	0.218**	0.171**	0.117	0.283**	0.204*	0.042	0.048	0.040
	(0.086)	(0.085)	(0.084)	(0.117)	(0.115)	(0.115)	(0.051)	(0.062)	(0.047)
C_Long2	0.050	0.131	0.071	0.184*	0.261**	0.193*	-0.053	-0.022	-0.026
	(0.084)	(0.084)	(0.082)	(0.106)	(0.106)	(0.106)	(0.056)	(0.054)	(0.040)
CC_A	0.186**	0.283***	0.228***	0.270***	0.410***	0.333***	0.046	0.050	0.039
	(0.083)	(0.082)	(0.082)	(0.104)	(0.101)	(0.102)	(0.065)	(0.073)	(0.054)
RR_A	0.103	0.182**	0.146*	0.178	0.303***	0.239**	0.013	0.019	0.022
	(0.085)	(0.086)	(0.084)	(0.112)	(0.111)	(0.111)	(0.046)	(0.052)	(0.040)
Intensity crackdown	-0.039***	-0.076***	-0.058***	-0.049***	-0.095***	-0.077***	-0.007**	-0.025*	-0.009*
	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.004)	(0.004)	(0.013)	(0.005)
Fined in the previous period	0.025	0.009	0.023	0.044	0.008	0.037*	-0.008	0.011	-0.003
	(0.021)	(0.023)	(0.015)	(0.029)	(0.032)	(0.021)	(0.012)	(0.016)	(0.007)

| Continued	All Observations			Fare-dodgers only			Non-Fare-dodgers only		
Dependent variable: Fraud	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(7)	(9)
(1 if Fraud, 0 if not)	1st Month	2nd Month	All	1st Month	2nd Month	All	1st Month	2nd Month	All
Risk index in the gain domain	-0.012	-0.014	-0.013	-0.014	-0.015	-0.017	-0.001	-0.001	0.001
	(0.013)	(0.013)	(0.013)	(0.017)	(0.017)	(0.017)	(0.007)	(0.008)	(0.006)
Risk index in the loss domain	0.019	0.018	0.019	0.024	0.016	0.021	-0.000	0.005	0.002
	(0.015)	(0.015)	(0.015)	(0.020)	(0.021)	(0.020)	(0.009)	(0.010)	(0.007)
Risk index in the gain domain	-0.014	-0.011	-0.013	-0.009	-0.009	-0.008	-0.007	-0.006	-0.007
under uncertainty	(0.013)	(0.013)	(0.013)	(0.017)	(0.018)	(0.017)	(0.008)	(0.009)	(0.007)
Risk index in the loss domain	-0.007	-0.009	-0.008	-0.018	-0.017	-0.018	0.003	-0.000	0.002
under uncertainty	(0.015)	(0.015)	(0.015)	(0.019)	(0.020)	(0.019)	(0.008)	(0.009)	(0.006)
Fare-dodgers (Self-report	0.388***	0.383***	0.397***	-	-	-	-	-	
Or no ticket)	(0.052)	(0.051)	(0.051)						
Fare-dodgers (fine recovery	0.298***	0.282***	0.299***	-	-	-	-	-	
office)	(0.076)	(0.076)	(0.075)						
Day	0.003***	0.001	0.001***	0.004***	0.001	0.001***	0.001*	-0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	8,091	8,370	16,461	5,858	6,060	11,918	2,233	2,310	4,543
Wald-chi2	260.25	461.18	718.94	153.74	379.16	570.63	65.44	37.17	99.19
Prob >chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
Log likelihood	-3579.906	-3396.3567	-6825.861	-3000.962	-2818.507	-5707.403	-586.3902	-570.4097	-1109.6325

Notes: The table reports marginal effects. These are calculated assuming that the random effects are zero. Standard errors are clustered at individual level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

**Figure 1**: Evolution of the average fraud rate by blocks of days over the sixty days
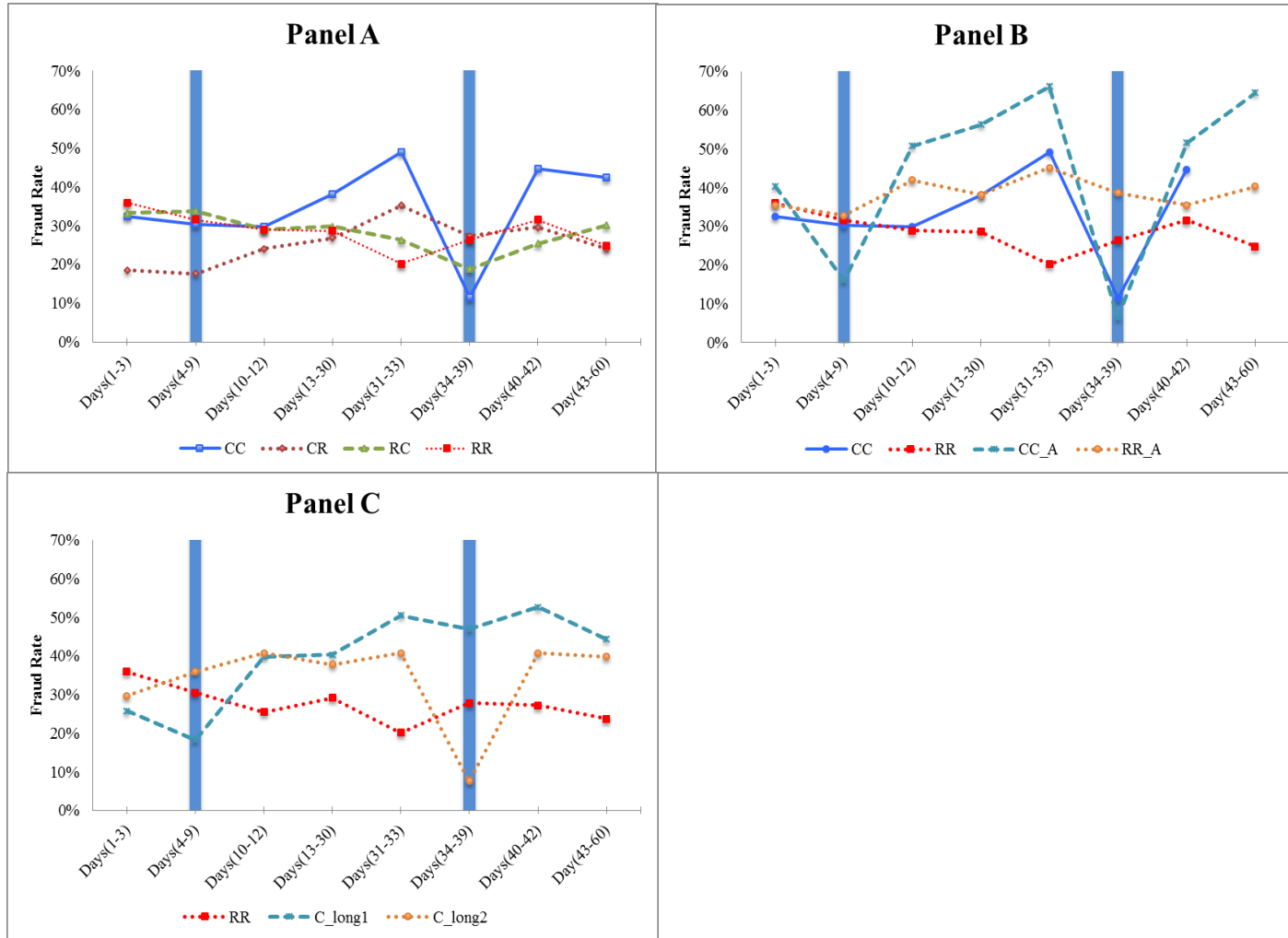
*Note*s: The vertical bars identify crackdown days. In treatment CC_A, the effect and execution of a crackdown coincided. In Panel B, crackdown days are 3-8 and 33-38 for this treatment.

Figure 2: Predictive margins of the fraud probability over consecutive days with audit
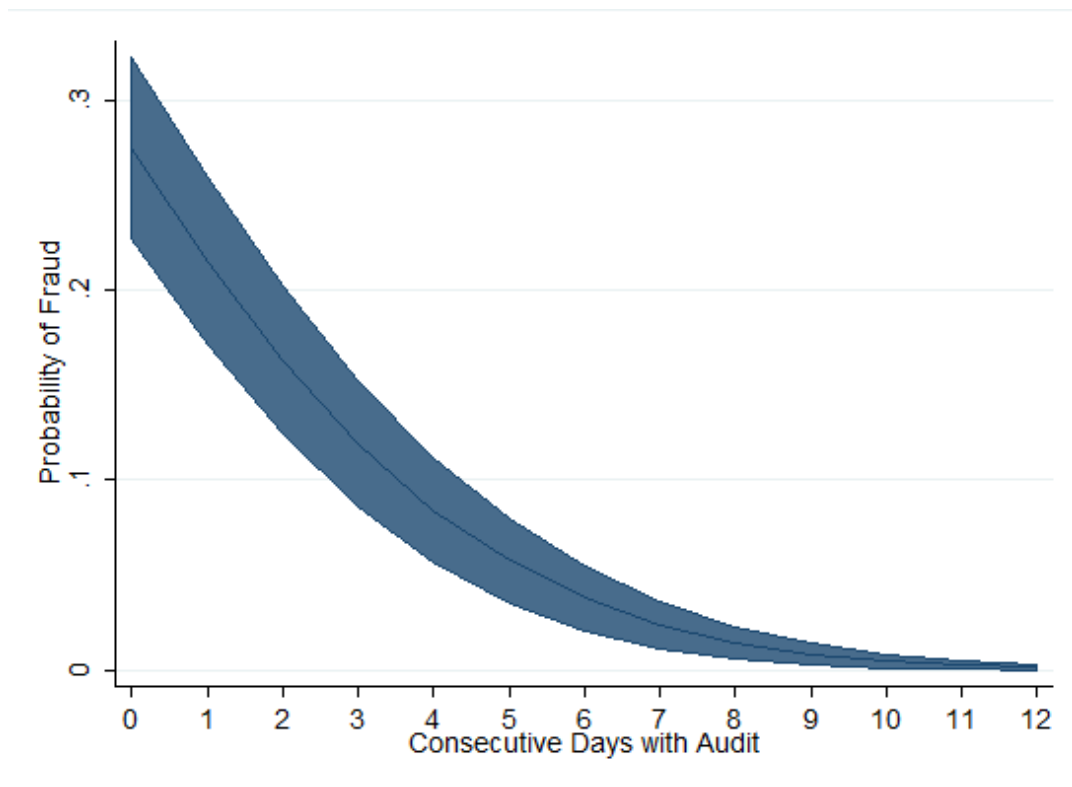
Note: the band identifies the 95% confidence interval.

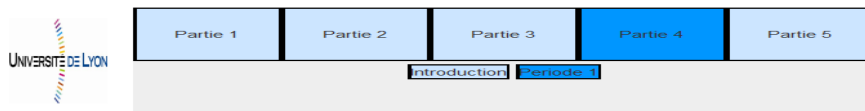**Figure 3**: Total revenue per ticket by treatment
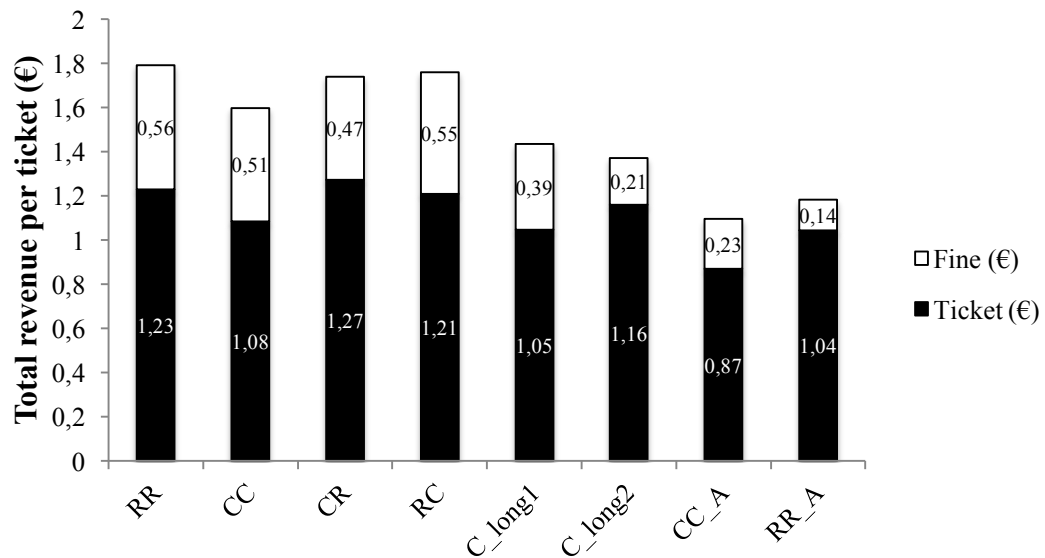
Appendix

Figure A1: Lab in the field

Figure A2: Decision-making screen shot

Chaque jour, cliquez sur « O » si vous achetez le ticket ou sur
« N » si vous ne l'achetez pas. Validez.

Vous saurez immédiatement si vous avez été contrôlé.

Semaine	LUN	MAR	MER	JEU		VEN
1				Amende	Suivant	
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						