A Theory of Optimal Random Crackdowns[*]

Jan Eeckhout

University of Pennsylvania

Nicola Persico

New York University

Petra Todd

University of Pennsylvania

March 2009

Abstract

This paper develops an incentives-based theory of policing that can explain the phenomenon of random "crackdowns," which are intermittent periods of high interdiction/surveillance. We show that, when police minimize the crime rate, random crackdowns can emerge as part of an optimal policing strategy. We develop several variations of the basic policing model that would apply in different monitoring situations, such as speeding, drug interdiction, or screening to deter terrorism. For a variety of police objective functions, random crackdowns can be part of the optimal monitoring strategy. We demonstrate support for implications of the crackdown theory using traffic data gathered by the Belgian Police Department and use the model to estimate the deterrence effect of additional resources spent on speeding interdiction.

A Theory of Optimal Random Crackdowns

March 2009

Abstract: This paper develops an incentives-based theory of policing that can explain the phenomenon of random "crackdowns," which are intermittent periods of high interdiction/surveillance. We show that, when police minimize the crime rate, random crackdowns can emerge as part of an optimal policing strategy. We develop several variations of the basic policing model that would apply in different monitoring situations, such as speeding, drug interdiction, or screening to deter terrorism. For a variety of police objective functions, random crackdowns can be part of the optimal monitoring strategy. We demonstrate support for implications of the crackdown theory using traffic data gathered by the Belgian Police Department and use the model to estimate the deterrence effect of additional resources spent on speeding interdiction.

1 Introduction

Police often engage in "crackdowns" on crime, which are intermittent periods of high intensity policing. This paper develops a theoretical framework for modeling police monitoring behavior and individuals' decisions to engage in crime. Within this framework, we show that there are situations where it will be optimal for a crime-minimizing police agency to engage in random crackdowns. When they occur, crackdowns also provide a way of estimating the deterrence effect of policing. We illustrate the application of the model in analyzing a speed deterrence program used by police in Belgium. In particular, we estimate the deterrence effect of additional resources spent on ticketing speeders and assess whether the current level of deterrence is socially optimal.

Two features characterize our notion of random crackdowns. First, they are *arbitrary*, in the sense that they subject certain groups (identified by presence in a particular time or place, or by other observable characteristic) that are not notably different from other groups in criminal propensities, to higher intensity police monitoring. Second, they are *publicized*, i.e., those who are subjected to crackdowns are informed about them *before* they engage in criminal activity.[1] Crackdowns are employed in a number of policing situations. Some examples include drunk driving interdiction accomplished using sobriety checkpoints, crackdowns on speeding achieved through announced greater police presence on certain highways, or crackdowns on drug trafficking aimed at particular neighborhoods.[2]

Criminologists sometimes rationalize the use of crackdowns by appealing to psychological theories according to which the impression created by the temporary show of force (the crackdown) is a psychological "bluff" that leads potential criminals to overestimate the risk of detection during non-crackdown periods.[3] This view relies on the

---

[1]Our definition of crackdown is different from the conventional use of the term in the literature on policing (see e.g. Di Tella and Schargrodsky 2002, 2004) because we require that crackdowns be arbitrary. We will return to this point when we discuss the related literature.

[2]For example, operation "safe streets" in Philadelphia, which puts heavy law enforcement on particular city blocks, received extensive media coverage. Other examples of crackdowns include the NHTSA campaign "You Drink & Drive. You Lose" which instituted highly visible enforcement against drunk driving. Another example is "Checkpoint Tennessee," Tennessee's statewide sobriety checkpoint program.

[3]Sherman (1990), Ross (1984). Sherman, p. 11, recommends that crackdowns be highly publicized

potential criminals' expectations being systematically wrong and therefore is inconsistent with rationality. In this paper, we take a different approach and develop a model in which potential criminals are not fooled about the odds of detection and yet the crime-minimizing police finds it optimal to employ crackdowns. In our model, crackdowns arise as a rational (indeed, optimal) response of both police and citizens' incentives. They also arise under a variety of ways of specifying police objectives, for example, if the police minimize total crime, or *undetected* crime, or if they solve the social planner problem of trading off the costs and benefits of crime.

We next illustrate the main idea behind this result through a simple example where the police minimize crime subject to a resource constraint.

Example *Consider a population of 100 citizens, half of whom would never commit a crime, and half of whom would commit a crime unless they are certain that they will be caught. A citizen's propensity to commit a crime is unobservable to the police. The police resources are such that they can only check 50 citizens. Suppose that the police check citizens at random (note that all citizens look the same to police), so that each citizen has a probability 1/2 of being checked. Then, only the high-propensity citizens will commit a crime, giving rise to a crime rate of 1/2. Suppose now that half of the citizens have blue eyes, half have brown eyes and that eye color is known to be independent of the propensity to commit a crime. Nevertheless, suppose that police crack down on brown-eyed citizens and check them all, and completely ignore the blue-eyed citizens. Then no brown-eyed ever commits a crime because they are sure that they would be caught, and only those blue-eyed citizens commit a crime who have high criminal propensity. Thus, the crime rate with a crackdown on brown-eyed persons is 1/4, which is lower than the crime rate of 1/2 obtained without crackdowns.*

This thought experiment shows that crackdowns can reduce crime by introducing disparate treatment within a population of observably identical individuals. We have not proved that the specific way in which citizens are divided and policed (blue-eyed v. brown-eyed) is the optimal one for reducing crime, though this is indeed the case. We show later in the paper that given *any* distribution (continuous or discrete, unimodal or

---

and be followed by secret "backdowns," and warns of the risk of exhausting the bluff through overuse.

multi-modal) of the propensity to commit a crime, the crime-minimizing policing scheme involves dividing the population into no more than two groups, not necessarily of equal size. Our theory allows for ctizens to be heterogeneous in their propensity to commit the crime. The example also highlights an important maintained assumption of our theory: for crackdowns to be effective, it is important that criminals cannot easily arbitrage between crackdown and non-crackdown groups. In the example, citizens are assumed to be unable to change their eye color.

Let us now return to the example to consider how crackdowns make it possible to estimate the deterrence effect of policing.

Using crackdowns to identify the deterrence effect of policing *Consider now an increase in police manpower to 51 checks. How does the optimal policing scheme change? It can be shown that the optimal policing scheme involves moving one person from the non-crackdown group to the crackdown group. That is, police would pick a blue-eyed citizen, force him to wear brown contact lenses, and then check with probability 1 all those who appear to have brown eyes. The remaining citizens (with blue eyes) are never checked. We can calculate the expected decrease in the crime rate that follows from an increase in manpower of 1 check: it is the decrease in crime that obtains from moving a random citizen from the group that is not cracked down upon to the group that is cracked down upon. Because the average crime rate in both groups is observed, we can readily compute the expected decrease in the crime rate — in this case, the expected crime rate goes from 25 percent to 24.5 percent.*

In this paper, we develop a model of policing in which a police chief is given an incentive to reduce crime and a certain amount of resources. Under a variety of assumptions about police goals and constraints, the optimal monitoring strategy can take the form of random crackdowns. As in the above example, our analysis provides a methodology for estimating the deterrence effect of policing.

We apply our policing model to analyze the effectiveness of resources spent on speeding interdiction. Although the decision to speed is rarely studied by economists,[4] it has

---

[4]With some notable exceptions which will be discussed later: see Peltzman (1975), Levitt and Porter (2001), and Ashenfelter and Greenstone (2004).

great economic relevance, both in the U.S. and worldwide. According to data from the National Highway Traffic Safety Administration (NHTSA), speeding is a factor in 30 percent of all fatal crashes in the US.[5] In 2001, more than 12,000 people died in speed-related crashes on American roads, at an economic cost to society of more than $40 billion.[6] Worldwide, traffic injuries rank second to HIV/AIDS as the leading cause of ill-health and premature death among the 15-44 age group. Because the number of vehicles per capita is rapidly growing in developing countries, traffic injuries are projected to be one of the leading public health issues over the next few decades.[7]

To deter speeding, police in several countries have adopted programs of announced radar controls that publicize the location and approximate time of operation of radar controls.[8] The data analyzed in this paper were gathered in the Belgian province of the Eastern Flanders during the years 2000-2003. We have observations on all radar controls in that time period affecting 6.5 million cars and resulting in 206,146 tickets issued. The announced controls in the data are observed to rotate in a fairly mechanical fashion across different sections of the roads and time periods. The police plan the announcement schedule at least one month in advance, and the timing of announcements does not appear to be systematically related to speeding propensity. We interpret the announced controls as crackdowns on particular groups of motorists, those traveling on the announced section of the road at the announced time. We measure the deterrence effect of the increased probability of detection by comparing decisions to speed within the crackdown and non-crackdown groups. Using implications of the theoretical model, we are able to calculate the effect of increasing the level of resources devoted to speeding interdiction. In conjunction with value-of-life estimates, our results indicate that at the current level of interdiction, the marginal benefit, in terms of statistical lives saved, is close to the marginal cost of interdiction. Importantly, our analysis demonstrates that the rational theory of optimal interdiction can have not only *normative* implications (how the police *should* behave), but can even generative *positive* predictions about actual

---

[5] See *Traffic Safety Facts 2001.*

[6] Over 80 percent of the economic cost is attributable to lost workplace and household productivity. See Blincoe et al (2002).

[7] See the WHO publication on traffic safety: http://www.who.int/world-health-day/ 2004/en/traffic_facts_en.pdf.

[8] For example, the Netherlands, Belgium, Germany and Australia.

police behavior (how some police actually behave).

The paper develops as follows. Section 2 presents a theoretical model that we use to study the conditions under which crackdowns emerge as an optimal policing strategy when the goal of police is to minimize crime subject to a resource constraint. The model allows for unobserved heterogeneity in the benefits citizens get from breaking the law, under the assumption that the police knows the distribution of the unobserved variable. The model is extended to the case where the police is constrained in terms of the number of successful interdictions, which is the case in our application to speeding. Section 3 applies the model developed in Section 2 to data that we obtained from the Belgian police department. Section 4 provides a discussion of alternative theories of deterrence and some further extensions. Section 5 concludes.

1.1 Related Literature

The idea that deterrence may be improved by focusing interdiction on arbitrary subsets of the population is present in the literature on racial profiling (see Persico, 2001). Recently, Lazear (2006) develops a related idea in the context of designing educational tests, where the question is how much of the test content to reveal to the test-takers ahead of the test.[9] In the nutrition literature, there is a related idea in connection with nutrition curves. For example, Pratap and Sharma (2002) argue that, in the presence of limited amounts of food, maximization of family survival and resources may entail an unequal distribution of nutritional resources, i.e., focusing resources on a subset of the family. Relative to these strands of literature, the contribution of this paper is (a) to pose a general policing problem and to characterize the optimal policing strategy; (b) to point out that crackdowns (in our terminology) allow the researcher to infer the deterrence effect of policing; and, (c) to empirically illustrate the methodology within a policy-relevant application, speeding.

Our work is also part of the literature on bureaucratic incentives. The most relevant papers in the economics literature are Prendergast (2001) and Shi (2005). These papers look at the effect of bureaucratic oversight on policing, with special reference to race

---

[9]Coincidentally, Lazear (2006) also cites speeding as a potential application in developing his argument, but his and our work are independent. We discuss his paper in detail below in section 4.1.

disparities. Our paper also compares the strategies adopted by police under different incentive schemes, although that is not the main focus.

More broadly, the issue of deterrence in a traffic context is a subset of the vast literature on crime.[10] Of direct relevance to this paper is the literature concerned with traffic enforcement (speeding, drunk driving, and seat-belt wearing), which is reviewed by Zaal (1994). Much of the literature on speeding attempts to quantify the effects of a change in the speed limit on accidents.[11] There is also a literature directly concerned with police enforcement and with estimating the deterrence effect of increased policing and of greater penalties on speeding.[12] Parallel literatures deal with the deterrence effect of increased policing and greater penalties on drunk driving and on seat-belt wearing.[13] Also related is the literature that studies the connection between risk-taking behavior (speeding, drunk driving) and accidents (see Levitt and Porter (2001)). A paper by Ashenfelter and Greenstone (2004) uses changes in the speed limit across US states to estimate the value of a statistical life. They also provide a rich summary of economics papers in this area.

We acknowledge that our use of the term "crackdown" is somewhat different from the way the term is occasionally used to refer to increases in interdiction that are not deliberate randomizations, but rather may be considered exogenous increases in resources in the sense that they are caused by events unrelated to the crime that is the object of study. For example, Di Tella and Schargrodsky (2002, 2004), study the effect of crackdowns on bureaucratic corruption and on crime. Our work complements this line of inquiry by pointing out that "random" crackdowns can be expected to arise endogenously as part of the optimal policing strategy.

---

[10] See Becker's seminal (1968) paper. Also, see Levitt (1997) for a recent study of deterrence in law enforcement.

[11] See for instance Balkin and Ord (2001), Lave (1985, 1989) and the literature cited therein, and Ledolter and Chan (1996).

[12] See e.g. Redelmeier et al (2003), Homel (1988a), Ross and Gonzalez (1988).

[13] See e.g. Ross (1984) for drunk driving and Campbell (1988), Peltzmann (1975) for seatbelt use.

2 Policing Model

In this section, we characterize optimal monitoring strategies in a model in which the police minimize crime subject to a budget constraint. One may think of this problem as originating from an agency relationship in which a principal (a politician, a bureaucrat, or a high level police administrator) is faced with the problem of giving incentives for an agent (the police chief) to allocate resources effectively towards achieving some socially desirable outcome, such as a lower crime rate.[14] Citizens, who differ unobservably in their propensity to commit a crime, choose whether to commit a crime. We assume that the agent knows the distribution of the propensity to commit crimes across the population, but the principal only observes the realized crime rate (which depends on police behavior) and therefore gives the police chief incentives to minimize crime subject to a resource constraint. We also assume that the police chief receives some amount of resources (e.g. manpower), and that he can commit to choosing which citizens to police and with what intensity.

In Section 2.3, we explore several variants of this problem, including one in which the agent takes into account individual benefits from crime.

2.1 The model

There is population of size 1 that is heterogeneous in the benefit x from committing a crime, assumed to be unobservable by the police. Let x be distributed across this population according to a c.d.f. F, and let p denote the probability that the citizen is monitored. If a citizen commits a crime and is monitored, he is caught and receives penalty T.[15] We assume that $p \in [0, \overline{p}]$, which implies that a citizen can be monitored with probability no greater than some $\overline{p} \leq 1$.

---

[14]Besides crime reduction, the principal might have other objectives. For example, a principal might want to allow some citizens who get the highest value to commit crimes. We will explore this point in Section 5.

[15]Here we assume that the citizen's utility functions are linear and that committing the crime is a discrete decision, but in Section 4 we relax these assumptions. Also, for the moment we take T as given. We will return to this issue in Section 4.

A citizen with benefit x commits a crime if

$$x - pT > 0. \tag{1}$$

If a group of citizens is policed with intensity p, the fraction of criminals is

$$1 - F(pT).$$

The police minimize the crime rate. One possible policing strategy is to monitor every citizen with the same probability. Alternatively, the police can divide the population into subgroups and police them at different intensities. Of course, this division only matters if the citizens know that they are policed with different intensities, so we will assume that each citizen knows the intensity.[16] We denote by $\mu(p)$ the size of the group policed at intensity p. Because the total size of the population is 1, it must be $\int_0^{\overline{p}} \mu(p)\,dp = 1$.

The contribution of group $\mu(p)$ to total crime is $\mu(p)(1 - F(pT))$, and aggregating over all groups gives the total number of criminals:

$$\int_0^{\overline{p}} \mu(p)(1 - F(pT))\,dp. \tag{2}$$

In this section, we assume that the police's goal is to minimize the number of criminals. Alternative specifications of the objective function are studied in Section 2.3, where we also look at the social planner's problem.

Let us now turn to the resource constraint. The police chief is assigned some amount of resources (such as officer-hours). Monitoring a group of size μ with intensity p is assumed to require police resources in the amount of $\mu \cdot p$. If total resources of P per capita are available for policing, the aggregate police resource constraint is

$$\int_0^{\overline{p}} \mu(p)\,p\,dp \leq P. \tag{3}$$

We refer to this constraint as a *time constraint*. An alternative way of specifying the constraint on police resources is explored in Section 2.3. The police chooses a probability measure μ to minimize the number of criminals (2) subject to the resource constraint (3).

---

[16]In practice, this means that the police must inform citizens of the intensity with which they are policed.

2.2   Analysis

We next provide an intuitive characterization of the properties of the solution to the police problem previously described. These properties are summarized in Propositions 1 and 2. The propositions are a consequence of Theorem 1, which is stated and proved in Appendix A.

Let us start by supposing that the solution to the police problem entails policing all citizens with the same intensity. By the resource constraint, this intensity must equal P. In terms of our model, this policing strategy corresponds to $\mu(p)$ equal 1 if $p = P$, and equal zero otherwise. Substituting this choice of μ into the objective function (2) we see that the number of criminals equals $S = 1 - F(PT)$. This situation is depicted in the left hand panel of Figure 1.
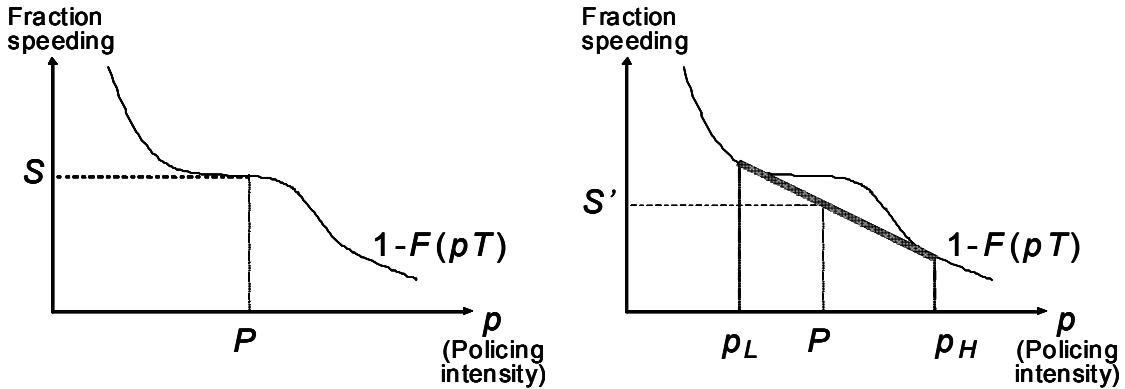

Figure 1. A crackdown is optimal.

As seen in the right panel of the Figure 1, the number of criminals can be reduced if resources are allocated differently. If some citizens were policed with intensity p_L and the rest were policed with intensity p_H, then it would be possible to bring the number of criminals down to $S' < S$.

Citizens who are policed with intensity p_H are said to be subject to a crackdown. Figure 1 shows that crackdowns help "iron out" the inward bumps of the function $1 - F(pT)$, thus enabling the police to maintain policing intensity on the efficient frontier. More precisely, by using crackdowns the police elicit a response function from citizens that corresponds to the *convex hull* of the epigraph of (i.e., of the area above) the function $1 - F(pT)$.

More formally, this crackdown strategy corresponds to choosing $\mu(p) > 0$ if $p =$

11

p_L, p_H, and equal to zero otherwise. The fraction $\mu\left(p_H\right)$ is optimally chosen to be the largest possible compatible with satisfying the resource constraint (3), which therefore reads

$$\left(1 - \mu_H\right) p_L + \mu_H p_H = P.$$

Of course, crackdowns are not always part of the optimal policing strategy. If, for example, the function $1 - F\left(pT\right)$ is globally convex, as depicted in Figure 2, then crackdowns would not be optimal. Even in Figure 1, if P were smaller than p_L or larger than p_H, crackdowns would not be optimal. In those cases, the most efficient use of resources is to police every citizen with the same intensity.
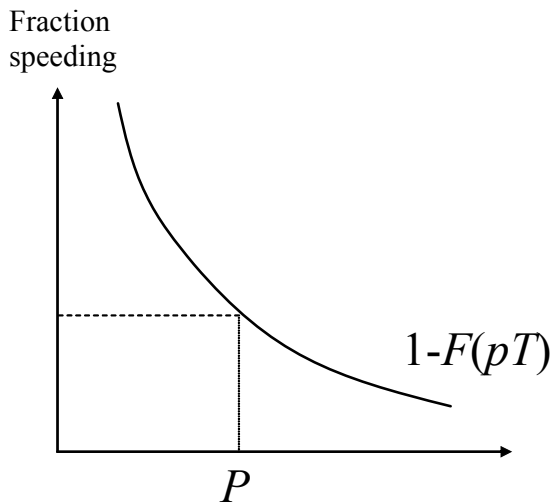

Figure 2. Crackdowns are not optimal.

When crackdowns are optimal, it is because the function $1 - F\left(pT\right)$ is not convex. Given that crackdowns play a "convexifying" role, there is no additional gain in dividing the population in *more than two* groups. In fact, given any function $1 - F\left(pT\right)$, any point in its convex hull can be achieved as a convex combination of at most two points in its epigraph. A three group crackdown, therefore, which would entail three different policing intensities, can achieve nothing more than a two-group crackdown. The following proposition actually takes this logic a bit further in stating that "generically," three group crackdowns are *strictly* suboptimal.

Proposition 1 *Given a homogeneous population with a generic distribution of propensity to commit a crime, the optimal policing strategy involves either monitoring everyone*

12

at the same rate, or dividing the population into at most two groups to be monitored with different intensities.

Proof. This proposition follows from Theorem 1 , which is proved in Appendix A. ∎[17]

An extreme form of crackdowns arises when $1 - F$ is globally concave. In this case, the convex hull is given by the segment that connects the points $(0, 1 - F(0))$ and $(\overline{p}, 1 - F(\overline{p}T))$, which means that for any P we have $p_L = 0, p_H = \overline{p}$. Thus, the optimal policy entails the use of extreme crackdowns: one group of citizens will be monitored as intensely as possible, the rest will not be monitored at all.[18] This observation gives rise to the following remark.

Remark 1 *If F is convex on its domain, then for any $P \in (0, \overline{p})$ the optimal policing strategy involves monitoring one group of citizens with maximal intensity, and not monitoring the others at all.*

It is worth pointing out that crackdowns are generally optimal for some P unless f is monotonically decreasing on its support.

Remark 2 *Unless F is concave on $(0, \overline{p}T)$, there exists some P such that the optimal policing strategy involves random crackdowns.*

We now turn to the comparative static result that deals with increases in the police budget in the presence of crackdowns. The intuition behind the result can be explained using Figure 1. Suppose that P, the amount of resources available to the police, is increased slightly. Although the fraction of citizens who are subjected to a crackdown is now higher because more resources are available, the optimal police strategy still entails a crackdown with intensities p_H and p_L. In other words, only the sizes of the two groups change, but not the intensity with which they are monitored. This simple but important point, which we note in Proposition 2, is important to being able to use crackdowns to infer the effect of relaxing the resource constraint.

---

[17]Theorem 1 in Appendix A is a general formulation of the problem we solve here. It allows for different variations, including those we explore further in the paper. By proving the result for the general case, all our results follow for each of the specific cases considered.

[18]This is the case described in the example in the introduction.

Proposition 2 *Given total police resources of P, suppose the optimal policing strategy involves dividing the population into a crackdown group of size μ_H monitored with intensity p_H and a non-crackdown group of size μ_L monitored with intensity p_L. Consider an increase in total police resources to $\widetilde{P} \in (P, p_H)$. In the new optimal strategy the crackdown group is larger than before, (i.e., $\widetilde{\mu}_H > \mu_H$ and thus $\widetilde{\mu}_L < \mu_L$, the non-crackdown group is smaller), but the intensities with which the two groups are monitored remain unchanged (they are still p_H and p_L).*

Proof. This proposition follows from Theorem 1 , which is proved in Appendix A. ∎

Proposition 2 provides a way of forecasting the deterrence effect of an increase in police resources. Crucially, the approach does not require knowledge of the shape of the function $1 - F(pT)$. Refer again to Figure 1. Graphically, increasing P results in the crime rate S' sliding down along the shaded segment. The slope of the shaded segment, therefore, determines the degree to which crime decreases as resources increase. This slope can be calculated based on the formula

$$\frac{[1 - F(p_H T)] - [1 - F(p_L T)]}{p_H - p_L}.$$

Multiplying this slope by $\widetilde{P} - P$ provides a way of estimating the expected decrease in crime due to a hypothetical increase in police resources from P to $\widetilde{P}$. Thus,

$$\frac{\Delta \text{Crime}}{\Delta P} = \frac{(\text{crime rate}|p_H) - (\text{crime rate}|p_L)}{p_H - p_L}.$$

The terms in the numerator on the right-hand side (the crime rates with and without crackdown) as well as those in the denominator (the intensity of monitoring) would be observable in most applied settings in which crackdowns are observed.

To see why observing crackdowns is necessary to carry out this computation, consider the no-crackdown primitives depicted in Figure 2. We are interested in forecasting $1 - F\left(\widetilde{P}T\right)$, the crime rate after the increase in the budget. Absent any information on the shape of the function $1 - F(pT)$, there is no way to compute $1 - F\left(\widetilde{P}T\right)$ based on the available information, which is only the knowledge of $1 - F(PT)$, the initial crime rate. This is why most of the literature on deterrence focuses on identifying sources of exogenous variation in P, which allows one to trace out (or at least locally approximate)

the function $1 - F(PT)$ as P varies. However, in the presence of crackdowns there is no need to identify sources of exogenous variation in P to identify the deterrence effect. One can think of crackdowns as being a case where exogenous variation in P arises as part of the optimal policing strategy.

If, in addition to observing crackdowns, one also has access to exogenous variation in total police resources P, then Proposition 2 yields a testable implication of the model. The implication is that, as P increases (between p_L and p_H), the optimal monitoring intensity should not change, but the size of the group subjected to crackdowns should increase. In Section 3.4, we consider this implication in the context of speed interdiction.

2.3   Constraint on successful interdictions

In the empirical application of section three, the Belgian police are given a constraint on the total amount of tickets that they are allowed to write, rather than a constraint on manpower. In this Section we adapt the model to the problem faced by the Belgian police.

As before, we assume that the goal of the police is to minimize the crime rate subject to a constraint. The resource constraint is now given in terms of the number of *successful* interdictions rather than in terms of time resources. That is, monitoring *criminals* has a cost to the police, but monitoring honest citizens is costless. This captures environments in which interdiction is cheap relative to the cost of processing violations. This happens to be the case in our speeding application of Section 3, where the police are administratively restricted in the number of tickets that they can issue in a year.[19]

The term $(1 - F(pT)) \cdot p$ represents the number of successful interdictions from a population that is policed with intensity p. To capture the constraint on successful interdictions we modify the constraint (3) to read

$$\int_0^{\overline{p}} \mu(p)(1 - F(pT)) p \, dp \leq C. \tag{4}$$

The police minimizes expression (2) subject to the constraint (4).

---

[19]This makes sense because detecting speeders with automatic radar machines is almost costless relative to processing a traffic ticket. More on this in Section 3.

To rule out trivial cases where the resource constraint is not binding, we assume that C is such that the police could not afford to monitor everyone with maximal probability. This assumption, which will be maintained throughout, is

Assumption $C < (1 - F(\overline{p}T))\overline{p}$.

The present problem shares a key formal similarity with the benchmark model: both programming problems are linear in μ. As a consequence, even though constraint (4) and constraint (3) are quite different, Propositions 1 and 2 continue to apply. In constraint (4), for example, the kernel of the integral is not necessarily increasing in p: higher monitoring intensity does not necessarily entail more successful interdictions. Yet, at the optimal solution, one can show that more resources (successful interdictions) must be expended per capita on the crackdown group than on the other group. One can also show that whenever crackdowns are optimal in the benchmark model for all values of P, then they are also optimal when police are ticket constrained. These results are collected in the following proposition.

Proposition 3 *Consider a monitoring problem in which police minimize crime under the constraint that successful interdictions not exceed C. Then:*

a) The optimal monitoring strategy involves either monitoring everyone at the same rate or dividing the population into at most two groups, which are monitored at different intensities.

b) Given C, suppose the optimal policing strategy involves dividing the population into a crackdown group of size μ_H monitored with intensity p_H and a non-crackdown group of size μ_L monitored with intensity p_L. Consider an increase in C to $\widetilde{C} \in (C, (1 - F(p_H T)) \cdot p_H)$. In the new optimal strategy the crackdown group is larger than before, (i.e., $\tilde{\mu}_H > \mu_H$ and thus $\tilde{\mu}_L < \mu_L$, the non-crackdown group is smaller), but the intensities with which the two groups are monitored remain unchanged (they are still p_H and p_L).

c) At the optimal monitoring strategy, the expected number of successful interdictions per capita (the probability that there will be monitoring multiplied by the fraction of motorists who speed) is larger in the crackdown group than in the non-crackdown group.

d) Crackdowns are optimal for all values of C if they are optimal in the benchmark model for all values of P. The converse is not true.

Proof. a), b): see Theorem 1.

c) Suppose not. Then if one perturbed the optimal strategy by shifting a small mass of citizens from the non-crackdown group to the crackdown group, the resource constraint would continue to be satisfied and the crime rate would decrease. This contradicts optimality of the original strategy.

d) Let P denote the maximal feasible policing intensity when all motorists are policed with the same probability. This is the monitoring intensity that minimizes crime among all feasible non-crackdown strategies. For future reference, observe that feasibility implies $C = P(1 - F(PT))$. Consider now the ancillary problem which is to minimize crime subject to the constraint (2). Let μ_L, p_L, μ_H, p_H denote the optimal crackdown probabilities in the ancillary problem. By definition, this crackdown policy generates a lower crime rate than if all citizens were policed with intensity P. We now show that the same crackdown strategy is feasible in the original problem. This will prove that equal policing is dominated by crackdowns in the original problem.

To verify feasibility in the original problem, write the following chain of inequalities:

$$
\begin{aligned}
C &= P(1 - F(PT)) \\
&= [1 - F((\mu_L p_L + \mu_H p_H)T)](\mu_L p_L + \mu_H p_H) \\
&\geq \mu_L p_L (1 - F(p_L T)) + \mu_H p_H (1 - F(p_H T)),
\end{aligned}
$$

where the inequality reflects the concavity of the function $x[1 - F(xT)]$. This function is concave because F is convex, which we know because crackdowns are optimal for all values of P in the ancillary problem (refer to Remark 1). ∎

Part d) of the above proposition suggests that crackdowns can be optimal when the police are ticket constrained even in cases where they are not optimal in the benchmark model. The intuition for crackdowns when the police faces constraint (4) is as follows. In a crackdown, the high interdiction group commits little crime, while the group that is more prone to committing crime is rarely policed. This tends to reduce the number of tickets that are written relative to the case in which both groups are policed at the same rate. Thus, besides helping satisfy the objective function, engaging in crackdowns has beneficial effects on constraint (4). The second effect was not present in the benchmark problem.

Part (b) of Proposition 3 yields a useful formula for computing the effect of an increase in police resources on the crime rate. Because this formula will be used in the empirical work in section three, we derive it here. The crime rate given $\widetilde{C}$ equals

$$\widetilde{\mu}\left(p_L\right) \cdot \left(1 - F\left(p_L T\right)\right) + \widetilde{\mu}\left(p_H\right) \cdot \left(1 - F\left(p_H T\right)\right)$$

(note that there is no tilde over the p's in light of Proposition 3 part (b)). To obtain the change in crime, subtract from this expression the analogous expression when resources equal C. This yields

$$
\begin{aligned}
\Delta \text{Crime} \quad & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad (5) \\
= \quad & \left[\widetilde{\mu}\left(p_L\right) - \mu\left(p_L\right)\right]\left(1 - F\left(p_L T\right)\right) + \left[\widetilde{\mu}\left(p_H\right) - \mu\left(p_H\right)\right]\left(1 - F\left(p_H T\right)\right) \\
= \quad & \left[\widetilde{\mu}\left(p_H\right) - \mu\left(p_H\right)\right]\left(F\left(p_L T\right) - F\left(p_H T\right)\right).
\end{aligned}
$$

The optimal policing strategy $\widetilde{\mu}$ must meet the budget constraint, and so

$$\widetilde{C} = \left(1 - \widetilde{\mu}_H\right)\left(1 - F\left(p_L T\right)\right) p_L + \widetilde{\mu}_H\left(1 - F\left(p_H T\right)\right) p_H.$$

Isolating $\widetilde{\mu}_H$ yields

$$\widetilde{\mu}_H = \frac{\widetilde{C} - \left(1 - F\left(p_L T\right)\right) p_L}{\left(1 - F\left(p_H T\right)\right) p_H - \left(1 - F\left(p_L T\right)\right) p_L}.$$

The optimal policing strategy before the increase in resources must also meet the budget constraint, and so

$$\mu_H = \frac{C - \left(1 - F\left(p_L T\right)\right) p_L}{\left(1 - F\left(p_H T\right)\right) p_H - \left(1 - F\left(p_L T\right)\right) p_L}.$$

Substituting into (5) we get

$$
\begin{aligned}
\Delta \text{Crime} \quad = \quad & \left(\widetilde{C} - C\right)\left[\frac{F\left(p_L T\right) - F\left(p_H T\right)}{\left(1 - F\left(p_H T\right)\right) p_H - \left(1 - F\left(p_L T\right)\right) p_L}\right] \qquad (6) \\
= \quad & \Delta C \cdot \left[\frac{\left(\text{crime rate}|p_H\right) - \left(\text{crime rate}|p_L\right)}{\left(\text{crime rate}|p_H\right) \cdot p_H - \left(\text{crime rate}|p_L\right) \cdot p_L}\right].
\end{aligned}
$$

All the terms in the right-hand side brackets are observable when the resource level equals C. Thus, the decrease in crime rate due to an increase in resources can be computed even without observing any variation in the data in the level of police resources. In Section 3 this slope is calculated in the context of highway speeding interdiction.

3   Empirical Application: Speeding Interdiction

In this section we apply our theoretical model of policing to study speeding interdiction in Belgium. As discussed in section one, speeding interdiction is an important policy issue because traffic accidents are a leading cause of death and disability worldwide.[20] Our goal in this section is to estimate the deterrence effect of resources devoted to speeding interdiction by directly applying the theoretical framework developed in section two.[21] Then, in conjunction with value-of-life estimates taken from the literature, we consider whether the marginal benefit, in terms of statistical lives saved, warrants the marginal cost. Our estimates indicate that the current level of interdiction is close to socially optimal.

3.1   The environment and the data

Our data come from the administrative records of the Belgian police department. In three Belgian provinces (Eastern Flanders, Liege and Luxembourg), the police puts extensive effort into publicizing announced radar controls, through different media that include newspapers, radio, internet, local stores and restaurants.[22] Our analysis samples cover the province of Eastern Flanders, which has two major highways and one minor highway, each of them connecting to the city of Gent.[23] The two major highways are divided into four *sections*: A14-North, A14-South, A10-East, and A10-West. In total there are 5

---

[20]In 1990, for example, traffic accidents represented the fourth leading cause of loss of DALYs (disability adjusted life years) in developed countries. Worldwide, accidents were the third cause of loss of DALYs for ages 15-44. By comparison, war was only the seventh leading cause for those ages.

[21]The application of the theoretical model to speeding interdiction has the advantage that the issue of incapacitation does not play a significant role. Speeders do not often receive prison sentences, although they may temporarily get their license revoked. Crime rates can be reduced by deterring potential criminals or by incapacitating them, and the models we examined in section two dealt only with deterrence.

[22]The controls are announced, for example, on the website: http://www.federalepolitie.be.

[23]The East-West highway A10 connects Gent with Brussels to the East and with Bruges to the West and the North-South highway A14 connects Gent with Antwerp and the Dutch border to the North, and with the French border to the South. Both highways are of approximately equal length and cover around 60-65km within the province of Eastern Flanders. The third highway, R4, is a short section of highway connecting Gent with its port to the North.

sections, each of which is of roughly equal length (30-40km). The province has two radar control machines that can be placed along roads or highways to record the speed of drivers passing along that road and to take photographs of cars that are speeding, which are then issued tickets.[24]

An announcement merely says that a section of one highway (for example, A14-North) will be subjected to increased monitoring in a time period (for example, between the hours of 6am-12). The announcement does not specify the direction of the road on which the machine is placed, nor of course its exact location. In fact, the police generally hide the position the radar machine, so as to avoid the possibility that drivers may slow down in the proximity of the machine and then pick up their speed again.[25] The announcement also does not specify the length of time for which monitoring will take place, which is typically less than the 6 hours long interval. The announced controls in the data are observed to rotate across different sections of the roads and time (more on this later). On any given day, the police either make no announcement or, when they announce, they keep one of the machines in reserve to possibly monitor some other section in an unannounced way.

The data we use for the empirical analysis record the date, time of day, and location of the machine and whether the radar control was announced as well as information directly recorded by the machine, such as the number of vehicles passing by the machine, the fraction of cars and trucks that were driving in excess of the speed limit (the limit differs for cars and trucks) and the fraction of vehicles exceeding the speed limit by 15 km/h.[26]

---

[24]Radar control machines record speeds and take photographs of speeding vehicles. The license information obtained from the photo and information recorded on the speed is used to issue the tickets. If the driver passes a radar machine while exceeding the speed limit by a certain threshold the probability is close to one of receiving a ticket. It is not equal to one, because, rarely, sun glare makes the photo unreadable. The radar control machines are mobile and are typically moved to several locations throughout the day.

[25]Only in rare cases (less than 1% of our data) is the radar machine not hidden. Even then, it does not seem to be easily detectable by drivers: we verified that the probability of speeding does not decrease in those cases. The same finding was independently reported to us by the police, based on experiments they have done positioning a second radar machine several kilometers after the first radar.

[26]A major speeding violation is defined in the law as traveling at a speed of 10km/h or more over the speed limit. In addition, the margin of error of the radar is ±3%, or 4km/h at the maximum speed on highways, which is 120km/h. It is police discretion when to issue a ticket for major violations, and

In what follows, we will use the following terminology. A *monitoring event* refers to each of the possible time-space combinations in which the police sets out the radar. There are a total of 5475 monitoring events: 3 time periods (6am-12, 12-6pm, and 6pm-midnight) per day for 365 days, and 5 comparable, equidistant sections of road (2 on A10, 2 on A14, and 1 on R4). Our choice of the time period is natural given the police announces controls within those 6 hour time periods, and for comparability, we use the same unit of account for unannounced controls. We refer to a *monitoring interval* as the actual duration of monitoring during any 6 hour monitoring event. In the data there is variation in the monitoring intervals, with an average of 3 hours.

During announcement events, drivers know that monitoring can happen on the specified section during part of the 6 hour time period and in either of the two directions. Within our model, this allows the driver to form beliefs about the likelihood of being monitored. When no announcement is made, the driver knows that there can nonetheless be monitoring and, within our model, forms a different belief about the probability of monitoring.

Police objective and constraints. The police department explicitly states that its goal in issuing tickets is to deter speeding. Maximizing the revenue from traffic tickets is not an objective, and in fact, the police do not get to keep the revenue from the tickets they write. Through conversations with the police, we learned that they face a binding constraint on the total number of tickets. The primary cost of issuing a speeding ticket is the administrative processing cost, which is estimated to be about US $0.50. The police are given a total budget at the beginning of the year (allocated by the legislature), so they know how many tickets can be issued during the year within this budget.[27] The

---

currently, police considers a major violation traveling at a speed of 136km/hr or higher on highways. During the time period covered by our data, tickets were not issued systematically for minor violations (over the speed limit, but less than 136km/h) as a result of the radar data collected.

[27]This institutional arrangement raises at least two questions. 1. Why not let the police choose the number of tickets and allow them to keep a fraction of the proceeds? The likely reason is that politicians and the police want to avoid the appearance of a monetary incentive to write more tickets because the public does not want police who are revenue maximizers. In fact, the statutory objective of the traffic police is to minimize speeding, and not to maximize fine revenues. 2. Why not relax the constraint on the number of tickets? Our calculation (Section 3.5 below) does not indicate a clear welfare reason for wanting to relax the constraints, so it is not immediately clear that the politicians are imposing an

budget determines in large part the total number of ticketed speeders per year, that is reported in Table 1a. To avoid issuing too many tickets, police do not make use of the radar control machines every day. On days when no announcement is made, police may or may not use the machines. On days with announcements, police make use of at least one machine on the announced road and may also use the other machine on the same or another road unannounced.

Following organizational reforms within the police, in 2002 and 2003 there was a sharp increase in the number of tickets that the police were allowed to issue on highways.[28] The size of the budget constraint more than doubled, from 33,951 tickets in 2000 to 78,136 tickets in 2002. Based on conversation with the police, this reallocation of funds was not triggered by any perceived change in the motorists' propensity to speed, but rather resulted from broader organizational changes.[29] We will therefore treat this change in the police budget as exogenous, and we will use this change as an additional way of examining support for the model.

Monitoring Policy. The monitoring policy is determined at least one month in advance of the actual radar control. For this reason, the officer who schedules the times and locations of announced and unannounced controls does not react to short term changes in circumstances, such as unpredictable weather conditions. The planned radar controls appear to be always implemented. Table 1a shows the number and percentage of vehicles subject to announced and unannounced radar control on the three major highways for years 2000-2003. Table 1b shows the number of drivers issued speeding tickets. Table 1c reports the number of monitoring events on each road, where a monitoring event is defined as the machine being placed on a road and recording information for some time interval. Highway A14 has the highest level of monitoring, followed by A10 and then the shorter highway, R4. Table B1 in the Appendix tabulates the number of announced

---

inefficiency by limiting the number of tickets.

[28]Prior to the reforms, the state police monitored both highway and nonhighway roads. After the reforms, there was a change in the jurisdiction so that state police monitored only highway roads and local police monitored nonhighway roads, and the number of tickets that the state police could issue on highways increased. We do not know what happened to the level of resources on non-highway roads, as they are not recorded in our dataset, which comes from the state police department.

[29]This change occurred, in part, because resources previously used to monitor both highways and some smaller roads were from 2002 on earmarked for highways only.

and unannounced monitoring events by month of year. There is no systematic pattern, except that monitoring is more frequent in the month of December. Table B2 tabulates monitoring and announcement events by day of the week and shows that monitoring is more frequent on weekend days in 2000 and 2001.

Driver's avoidance. One potential concern in applying our model to the data is whether drivers who hear the radar control announcements can select an alternate route to avoid detection, which would mean that the speeding response of people who choose to remain on the announced road could no longer be compared to the response of drivers in the absence of the announcements. Because our data pertain only to major highways, the potential problem of route selectivity is mitigated. If a motorist wants to avoid a highway with announced radar controls, she will necessarily have to take country roads, with relatively low speed limits (between 50 and 90 km/h) and with traffic lights. On the basis of time cost, a driver should prefer to take the highway rather than a country road. Aside from spatial avoidance, one might be concerned about temporal avoidance: anticipating or postponing one's travel to avoid monitoring. However, the length of an announced monitoring interval is 6 hours and drivers do not know exactly when the monitoring will take place. A driver whose ideal driving time is in the middle of this interval would have to anticipate or postpone his travel by 3 hours. Although there will be some drivers who can engage in some avoidance with little change to their ideal driving schedule, that only applies to those who are ideally driving at the very beginning or very end of the 6 hour interval. In fact, we computed that if monitoring induces a driver to reduce his speed by 20 km/h, then the time loss on the 40 km length of our section is about 3 minutes. According to this calculation the concern about temporal avoidance is limited to a small fraction of the monitored population.

3.2   Constructing the variables and fit to the model

The theoretical model presented in Section 2 assumed that individuals are heterogeneous in their propensities to speed and that the heterogeneity is unobserved by the police. In reality, the population driving on the road may vary in some observable as well as unobservable ways. For example, police may be aware that speeding is more common during certain times of the year or on weekends We denote the set of characteristics that

are observable to police by Z and assume that the theoretical model applies conditional on a set of observables Z. Table 2 reports the summary statistics of the variables used in the analysis.

Estimation of the expected probability of monitoring. Recall that in the theory p_L and p_H represent drivers' perceived probabilities of being caught speeding on unannounced (low) and on announced (high) monitoring days. On announcement days, police monitor the announced road with probability one. A typical announcement specifies the section of road being monitored and the length of time (for example, one hour in the morning). The driver's perceived probability of being caught speeding is typically less than one, because police do not monitor the entire length of time of the announced control and they only monitor one of the two driving directions. We assume that drivers form expectations about the probability of being subject to monitoring, and that they may use characteristics of the day in forming their expectations. We therefore allow $p_L(Z)$ and $p_H(Z)$ to depend on Z. For example, a driver may know that the police tend to do more monitoring on weekends or on holidays.[30]

On days with announcements, monitoring is announced within three separate time intervals, corresponding roughly to morning (6am-12pm), midday (12pm-6pm) and evening (6pm-12am). We therefore divide each day of the year into three potential monitoring intervals. For each year, we therefore have a total of 5475 monitoring events: 365 days, 3 time intervals per day, 5 sections of road (2 on A10 and A14 and 1 on R4). Let m be an indicator for whether any monitoring takes place during a monitoring interval and let a indicate whether the monitoring was announced. On announcement days, some monitoring definitely occurs during the stated interval, so $\Pr(m = 1|a = 1, Z) = 1$. On days without announcements, drivers form a belief about the probability that any monitoring takes place during a monitoring interval, $\Pr(m = 1|a = 0, Z)$, potentially taking into account factors such as the day of the week, the month of the year, whether the particular road was recently subject to monitoring, whether there has been an announcement on another road or whether it is a holiday. We estimate $\Pr(m = 1|a = 0, Z)$ by a logistic

---

[30]Because the police hide the location of the radar machine, a driver's belief about the probability of being monitored should be constant along a section of highway. In reality, not all motorists travel along the entire section—some make shorter trips. We will show below that, for our purposes, there is no loss in generality in treating shorter trips in the same way as longer trips.

regression, using as the estimation sample all potential monitoring events in which there was no announcement.[31]

On both announced and unannounced monitoring days, to arrive at a perceived probability of being monitored, drivers also have to form an expectation about the number of hours that the monitoring will take place. Using data on the actual length of monitoring during the event, we estimate a regression in which the dependent variable is numbers of hours spent monitoring. The independent variables indicate the day of week, month of year, holiday, and whether there were recent monitoring activities on that road. Using these ingredients, the predicted probability on announcement days, $p_H(Z)$, is obtained by

$$p_H(Z) = 0.5 \cdot E(\frac{H}{T_H}|m = 1, a = 1, Z),$$

where H represents the number of hours of monitoring and T_H is the total time of the announced monitoring event. The multiplication by 0.5 accounts for drivers' uncertainty about which direction of the road will be subject to the monitoring.

The corresponding probability on unannounced days, $p_L(Z)$, is

$$p_L(Z) = 0.5 \cdot \Pr(m = 1|a = 0, Z) \times E(\frac{H}{T_H}|m = 1, a = 0, Z).$$

$p_L(Z)$ is typically much lower than $p_H(Z)$, because $\Pr(m = 1|a = 0, Z)$ is much less than 1.[32]

We estimate the predicted probabilities of being monitored, $p_H(Z)$ and $p_L(Z)$, by estimating the components: $E(\frac{H}{T_H}|m = 1, a, Z)$ for $a = 0, 1$, and $\Pr(m = 1|a = 0, Z)$. For this estimation, we make use of data on the length of time spent monitoring within the monitoring event. Appendix B, Tables B3 and B4, report the estimated regression

---

[31]This estimation includes all potential monitoring events (365 days of the year times 3 intervals per day). The data only record events in which monitoring took place, but we could infer the characteristics of the events during which there was no monitoring (day of week, time of day, month of year, holiday, whether there was recent monitoring).

[32]For days without announcements, we set $T_H = 16$, because we do not observe in the data any monitoring during nighttime. For roads A14 and A10, we make an additional adjustment (multiply by 0.5) to the estimated probabilities $p_L(Z)$ to take into account drivers' uncertainty about which section of road is subject to the monitoring. On announcement days, drivers learn the section of road from the announcements but on unannounced days they do not know the section of road where monitoring will take place. On the shorter road R4, there is only one section so no further adjustment is necessary.

coefficients for the estimation of the predicted probability of monitoring on unannounced days and for the predicted length of time spent monitoring. As can be seen in Table B3, there are few systematic predictors of monitoring. It appears that the roads are more likely to be monitored unannounced if there was some unannounced monitoring in the previous week. Table B4 shows that few of the conditioning variables have any predictive power in explaining the length of time police spend monitoring. There is some indication that the time spent monitoring differs across quarters of the year, but there is no systematic pattern across all highways.

Finally, a note about the interpretation of p. In the theory, p represents the probability that a motorist traveling along an entire sector is monitored, equation (1) must be interpreted as a "per sector" equation. Thus, we take x to represent the "per sector" benefit of speeding.[33] A motorist who traveled only a fraction m of a sector would speed if $mx - mpT > 0$, or equivalently, if $x - pT > 0$. Thus, the motorist's decision problem is invariant to the fraction of the sector travelled. This formulation allows us to aggregate trips of different lengths, which is convenient because we do not observe the length of each individual trip.

Computing μ. $\mu(p_H)$ represents the fraction of monitoring effort devoted to high intensity interdiction (announced monitoring in our model). It is calculated as follows: in the denominator are all the 5475 possible monitoring events (365 days, 3 time blocks per day, 5 sections of road); in the numerator are the number of announced monitoring events (then $\mu(p_L) = 1 - \mu(p_H)$). The value of $\mu(p_H)$ are reported in Table 1c. For example, in 2000, $\mu(p_H) = 66/5475 = 1.21\%$.

Crime rate. Our theory requires us to compute $F(p_L T)$ and $F(p_H T)$, the fractions of speeders with and without a crackdown (see equation (6)), which we can obtain directly from the data.

3.3 The randomness of crackdown

One of the premises of our theory is that crackdowns are random. That is, the theory assumes that the difference in monitoring intensity between crackdowns and non-

---

[33] We can think of x as reflecting a time benefit from speeding over some interval, and we would expect value of time to differ across individuals.

crackdowns (p_H and p_L) reflects a non-convexity in F, rather than the police adapting their monitoring intensity to several convex F's, reflecting shifts in the propensity to speed that are observable to the police but unobservable to us. Figure 3 shows that the distributions of the p_H's and p_L's are bunched in proximity of their respective means and the two distributions have literally *no* overlap (there are no monitoring events with intensity between 0.05 and 0.15). The distribution of unobservables would have to be very peculiar to generate this "bimodal" pattern of monitoring intensities. That is, we would require the existence of a characteristic of drivers on a particular road segment that observed by the police but unobserved in our data, which sharply and discontinuously shifts their propensity to speed in a way that explains the bimodality. To the best of our knowledge from conversations with the police, there is no such characteristic and the bimodality is better explained by a non-connvexity in F.

There is, moreover, evidence that the police do not attempt to strategically condition announcements on factors that might influence the propensity to speed. First, the police commit to their announcement schedule at least 1 month in advance, and in so doing they forgo the opportunity to fine-tune their announcements to factors that might influence speeding (such as unseasonable weather, road conditions or traffic density). Second, the announcement schedule is a fairly mechanical rotation, and in personal communication, the police indicated that they were attempting to make the announcement schedule as random as possible.

There is thus ample anecdotal evidence that in our context crackdowns are random. In applications where such anecdotal evidence is absent (see Section 5 below for a description of alternative applications of random crackdowns) our theory can nonetheless offer a plausible foundation for empirical work provided one can establish support for the randomness of crackdown.

To further explore this question, Table B6 reports the estimated coefficients from a logistic regression of the probability of an announcement. The independent variables are characteristics of the day, such as whether it is a holiday and the quarter of the year, and information on recent monitoring events. We also include as a potential predictor variable the percentage speeding on the same road on an unannounced day, as a proxy for the underlying speeding propensities of motorists on that road.[34] Almost all the

---

[34] An anonymous referee pointed out this strategy for testing the randomness of crackdown. We are

variables, including the speeding measure, do not predict the announcements, which is generally supportive of the announcements being random.[35]

Even if there seems to be no clear evidence of targeted announcements our estimation strategy allows for potential systematic variation in monitoring behavior based on a set of observables Z that police and drivers might plausibly use.

3.4 Examining support for the model

Ideally, we would like to be able to estimate the empirical counterpart of the distribution of propensities to speed, as depicted in Figure 1, and look for a concave part in order to justify a crackdown. However, if the theory is correct, then we should observe only two points on the function $F(p_L T)$ and $F(p_H T)$, and there will not be any variation in the data to trace out the shape of the function. In the absence of a direct test on the shape of the distribution of propensities to speed, we turn to our model, and in particular to Proposition 3 for other kinds of predictions that can be used to examine the empirical support for the model.[36]

Implication #1: Partitioning into Two Groups One implication of the theory (Proposition 3 part (a)) is that the optimal policing scheme partitions the population in at most two groups. The optimal monitoring strategy involves either monitoring everyone at the same rate or dividing the population into at most two groups that are monitored at different intensities. The fact that police announce crackdowns on some roads on some days naturally gives rise to three different groups: (i) highway sectors with crackdowns; (ii) sectors without crackdowns on days in which crackdowns are announced on other sectors, and (iii) sectors on non-announcement days.

---

grateful for this suggestion.

[35]We expect the year effects to be statistically significant, given that the operating budgets are annual and the number of announcements increases over time.

[36]To estimate the empirical counterpart of the distribution of propensities to speed, one might look for variation in the day-to-day monitoring probabilities (the x-axis in Figure 1) that is (a) exogenous to the propensity to speed, and (b) anticipated by drivers. Unfortunately, this is difficult. Weather and traffic density, for example, are sources of variation, but they are likely to affect the drivers' underlying propensity to speed, thus shifting the curve in Figure 1. Thus, these are not sources of exogenous variation. In fact, we account for these factors in the empirical analysis by introducing them as conditioning variables.

[Figure 3]

Figure 3 plots the histogram of the estimated probabilities with which police monitors drivers, obtained by the procedure described in Section 3.2 (the plot refers to road A14; the histograms for A10 and R4 are similar and available upon request). The columns of the figures refer to year 2000, 2001 and 2002, and the mean of the probabilities in each years is shown in the x-axis label. For each year, the figure in the first row displays the distribution of the probabilities if there was an announced control (i.e. of $p_H(Z)$). The mean of this distribution is around 25% because, during an announced control, the police choose to keep the the machine out for about 3 hours on average within an announced 6 hour time window, on 1 of 2 possible driving directions, so $\frac{3}{6} \times \frac{1}{2} = 0.25$.

The middle row depicts the distribution of probabilities if there was no announcement on that road (A10), but there was an announcement simultaneously on another road. The bottom figure corresponds to the case without announcements on that road (A10) or on any other road. Groups (ii) and (iii) are monitored with similar intensities, and with sharply lower intensity than group (i), which is consistent with the prediction of the theoretical model. However, a statistical test of equality of the estimated probabilities in groups (ii) and (iii) rejects the hypothesis at 5% significance level. The probability of being monitored on a given road is slightly less when there is announced monitoring on another road.

Implication #2: Effect of an Exogenous Increase in Resources on Monitoring Probabilities The theory (Proposition 3 part (b)) predicts that with an increase in the number of tickets there will be an increase in announced controls (μ_H increases, with a corresponding decrease in μ_L) but no change in the monitoring probabilities. In 2002 and 2003, there were large increases in police resources that were arguably exogenous with respect to the speeding propensities of the drivers. The predictions of the model that the probabilities remain stable and that announcements occur more frequently seem to be borne out in the figures. Table 1c shows a very significant increase in μ_H in 2002, to 6.15%, more than five-fold compared to 2000. A simple comparison of the histograms in columns 1, 2, and 3 in Figure 3 reveals that as the number of tickets issued nearly doubled in 2002, the number of vehicles subject to monitoring increased dramatically but the monitoring probabilities remained roughly the same. The stability in the probability

of monitoring across years can also be seen in Table B3, where the estimated year effects are generally not significant determinants of the probability of monitoring, as well as in Table B5, which shows the average predicted probability of monitoring for each of the years.

Implication #3: **More tickets during crackdowns** The model predicts that the expected number of successful interdictions per capita is larger in the crackdown group than in the non-crackdown group (Proposition 3 (c)). The expected number of successful interdictions per capita equals the probability that there is monitoring multiplied by the fraction of motorists who speed. On days with announcements, the probability of monitoring increases substantially. If the fraction speeding went down enough to offset the increased probability of monitoring, then it is conceivable that the number of successful interdictions could go down. The data are, however, consistent with the model and the number of successful interdictions per capita increases with announcements. For example, on highway A10 the expected number of tickets per capita on announcement days is 0.505% whereas on unannounced days it is 0.026%.[37] Likewise on A14 (1.010% and 0.083% respectively) and R4 (1.041% and 0.038%). This test of the model is arguably weak, because the only way the test could fail is if the announcements had a great deterrence effect and motorists were about ten times less likely to speed during announced periods.

Overall, we conclude that the data broadly support the key predictions of Proposition 3. We next use the model to infer the deterrence effect of resources devoted to speeding interdiction.

3.5   The deterrence effect of announced controls

The main goal of our empirical analysis is to estimate the deterrence effect of a change in the number of tickets issued by the police. The simplest approach would be to measure deterrence by the change in total crime before and after a reform that exogenously changes the level of resources. We do not, however, have total crime in our data, which

---

[37]This calculation is based on estimates of p_L and p_H and of the percentage speeding with and without announcements that are estimated by the method described in the next section and are reported in Table 4 (the first column and the column labeled (3)).

only records information during monitoring events, so this simple approach is not feasible.[38] Our model provides an alternative way of estimating the deterrence effect that only requires knowing the speeding rates during crackdown and non-crackdown periods.

The theoretical model implied that relaxing the budget constraint should lead to an increase in the number of announced controls. Drivers who take to the roads on announcement days are subject to a higher probability of being caught speeding and are therefore viewed as the group subject to a crackdown. Here, the criteria by which the crackdown group is distinguished are time and day of traveling on the road.[39] To estimate the deterrence effect, we compare the speeding response on days with announcements (days with p_H) to the speeding response on no announcement days (with p_L). The maintained assumption is that announcements (crackdowns) are random conditional on the observed covariates. The theoretical model of section two assumed that monitoring intensity is the only observable factor affecting speeding decisions, but in reality other likely important factors are weather conditions, traffic density, day of the week, month of the year, time of the day and whether it is a holiday. We therefore include those additional variables as potential determinants of speeding decisions. To examine sensitivity to the set of included covariates (Z), we report results with and without conditioning on the additional covariates.

We estimate a discrete choice logistic model for drivers' decisions to speed, where the speeding decision is assumed to depend on the probability of monitoring and on the other

---

[38]Following this simple approach would require us to know *total* crime, i.e. the number of vehicles on the road and the speeding rates on all days and times of day when there was no monitoring, which is something we do not know. In the absence of this information, we could impute the number of vehicle and the speeding rates on the days with no monitoring. A difficulty with this imputation is that there are some observables that we do not know for days when there was no monitoring, such as traffic density and weather. With our model, we obtain the same objective, i.e. a way to get from the two crime rates to the deterrence effect. In addition to the fact that we do not need to know the total level of speeding vehicles, the model also allows us to examine policy changes that are not observed, i.e. any change in tickets different from the increase of 32,872 observed in 2002-2003.

[39]The observed policy of rotating announcements across space and time is not the unique optimal policy in our theoretical model. Alternatively the police could focus all resources on one particular road (for example always monitor one section of A10 intensely and always monitor all other sections with low intensity). We believe that a plausible reason (outside our model) for rotating announcements is to allocate the burden of interdiction "fairly".

covariates. Table 3 presents the estimated coefficients for three different specifications. In specification (1) the speeding decision is assumed to be a function solely of the monitoring probability. Specification (2) adds the conditioning variables that may also be relevant to the speeding decision: indicators for different levels of traffic density, an indicator for poor visibility on the road, an indicator for morning and evening rush hour weekday traffic, an indicator for whether the day is a holiday, and fixed effects for days of week, months of year, and year. Specification (3) includes the same set of conditioning variables, but the speeding decision is now assumed to depend only on whether there is an announcement, without taking into account the information contained in the length of the monitoring event. This would be the appropriate specification if driver's expectations about the probability of being monitored only depend on whether an announcement was made and not on other day-specific factors.[40]

As seen in Table 3, speeding decreases during announcement periods and is a decreasing function of the probability of monitoring. This result is robust to the inclusion of conditioning variables, although a comparison of the estimated coefficients across specifications shows that the estimated deterrence effect is smaller in the specifications that include the covariates. Controlling for covariates especially affects the estimated coefficient associated with the probability of monitoring on highway R4. As expected, individuals are more likely to speed when traffic density is lower. Speeding also tends to be higher during weekday rush hour times, on holidays, and on Sundays.

Using the estimated coefficients from Table 3, we estimate for each person the probability of speeding, evaluated at p_H and p_L values that are set at the average over all the observed values. The average p_H and p_L values for each road are reported in the first column of Table 4. Column (1) of Table 4 reports the average predicted decrease in speeding on each road that can be attributed to the announcements for the same three model specifications that were shown in Table 3.[41] As noted above, the estimated deterrence effects are smaller when additional covariates are included in the specification. We

---

[40]The results in the third column are robust to any potential misspecification in the model for the probability of monitoring.

[41]That is, we obtain a predicted decrease in speeding for each driver and then take the average over all drivers. The predicted decrease in driver-specific in the specifications where $p_L(Z)$ and $p_H(Z)$ depend on covariates Z, which are driver-specific (such as day and month of travel).

focus on the coefficients that include the covariates (reported in columns (2) and (3)), because they are likely to be important determinants of speeding decisions. On highway A10, the estimated coefficients imply that announcements reduce the fraction of drivers speeding on average by 5.0-15.5%. For highway A14, estimates range from 8.0%-18.7%, and, for highway R4, from 1.9%-8.6%.

These estimates can now be used to compute the deterrence effect of a hypothetical increase in police resources in the form of an increase in the number of tickets issued. The effect of an increase of $10,000$ tickets is reported in Table 4 for each of the highways. It is easiest to understand the computation in the context of a specific example. Consider model specification (1) for Highway A10. The expected number of tickets written per car on an unannounced day is given by the probability that a car speeds (0.031 from the first row, labeled (a)) times the probability that it is monitored ($p_L = 0.0084$):

$$1 * (.031) * (.0084) = 2.604 \times 10^{-4}.$$

On an announcement day, the probability that a car speeds is 6.5% lower (from the third row), and the probability that it is monitored is equal to 20.07%. The expected number of tickets written per car is therefore:

$$1 * (.031)(1 - 0.065) * (.2007) = 5.8203 \times 10^{-3}.$$

Therefore, an increase in the budget of $10,000$ tickets allows subjecting to announcement

$$\frac{10,000}{5.8203 \times 10^{-3} - 2.604 \times 10^{-4}} = 1.7986 \times 10^6$$

additional cars. Without an announcement, 3.1% would speed. Of these, 6.5% are deterred from speeding by the announcement. Thus,

$$1.7986 \times 10^6 * 0.031 * 0.065 = 3,597$$

drivers will be deterred as a result of writing $10,000$ more tickets which will be issued on additional announcement days. The same calculation is performed to obtain "Effect of additional $10,000$ tickets" in Table 4. Depending on the model specification, the reduction in the number of speeders ranges from 2,743 to 9,603 on highway A10, from 3,849 to 10,291 on highway A14, and from 763 to 3,817 on R4.

There is a large literature documenting the effect of speeding on accidents, injuries and traffic deaths.[42] We can use estimates from the literature of the impact of speeding on fatalities along with our estimates of how additional resources reduce speeding to evaluate whether the police optimally trade-off the costs and benefits of speeding interdiction. We take an average of the estimated deterrence effects of 10,000 tickets, found in Table 4 to be about 4,000 speeders. Assuming that each deterred motorists travels the length of a sector (about 40 km), 160,000 km are travelled by deterred motorists. The expected number of deaths on 160,000 travelled kilometers is around $\frac{1.3}{100000000} \cdot 160000 = 0.00208$.[43] In our data, deterred motorists reduce their speed by about 8 km/h.[44] Assuming the probability of injury and death increases by 5% per km/hour,[45] the additional interdiction is expected to reduce the number of deaths by 40%, or by $0.00208 * 0.4 = 8.32 \times 10^{-4}$.

On the cost side, writing 10,000 more tickets costs $5,000 in administrative costs and, we assume, wastes about 1 minute per deterred driver, or a total of about 67 hours. Given a wage of $10/h (the opportunity cost of time), the total costs of interdiction is $5,000 + 670$.

If the police were resolving optimally the trade-off between marginal cost of interdiction and marginal benefits, in terms of statistical lives saved, then the implied value of a statistical life is $\frac{5670}{8.32 \times 10^{-4}} = 6.8$ million dollars. This value is within the range of commonly used estimates of the value of a statistical life, indicating that the use of resources in policing may be close to efficient.[46]

---

[42] For the US, the *National Highway and Transportation Safety Authority* (NHTSA) provides estimates of speed-related crashes.

[43] We impute the expected number of deaths at 1.3 per 100 million km travelled (See DiGuiseppi et al (1998).) For comparison, in the US the expected number of deaths was 1.51 per 100 million of highway miles travelled in 2002 (see *Motor Vehicle Traffic Crash Fatality Counts and Injury Estimates for 2003*).

[44] Average speeds among speeders are 142 and 144 respectively for A10 and A14. Assuming that deterred motorists travel at the maximal non-ticketed speed (135km/h), deterred motorists reduce their speed by 7 - 9km/hr.

[45] This estimate is given in Finch *et al.* (1994), which presents a meta-analysis of before-after studies of speed limit changes in some western European countries. This estimate was used by the Belgian police in an internal memorandum to evaluate the impact of speeding on casualties.

[46] For example, Murphy and Topel (2003) report a range of $3 million to $7 million. The Environmental Protection Agency in the U.S. uses an estimate of 4.5 million (Murphy and Topel, 2003). If we do the same calculation using an estimated effect of 12,000 vehicles deterred per 10,000 tickets, which is at the

4 Discussion and Some Extensions

This section discusses some other existing theories of crackdowns, and mentions some extensions and variants of the base model.

4.1 Alternative theories of deterrence

As noted in the introduction, Lazear (2006) independently analyzes an auditing model which is close to the one analyzed here, and therefore deserves careful discussion. The main focus in Lazear's paper is on academic testing as an incentive for pupils to study a subject. The test can only include a given number of questions and so it only has a limited power to incent. This power can be spread thinly across the entire subject matter or focused on specific areas. Unlike us, Lazear (2006) does not investigate the optimal testing strategy. Instead, that work focuses on the following specific strategy: the students are told that a fraction $(1 - q)$ of the subject matter will not be tested *at all* and that the test will only cover the remaining q of the material. Note that if $q < 1$ this scheme induces, in our language, a special type of crackdown policy with $p_L = 0$.

An auditing policy of this type is, in general, suboptimal because it constrains p_L to be zero. The effects of this restriction can be seen in Figure 4 where for a given P the optimal crackdown intensities (marked by p_L and p_H) are compared to the constrained crackdown intensities (indicated by 0 and l_H). As expected, under the optimal crackdown the crime rate (S) is lower than under the restricted crackdown (S_l).

Observe that restricting the no-crackdown probability to zero also distorts the crackdown probability (l_H) relative to the optimal one. Indeed, it is easy to see that, regardless of the shape of F, it must be $l_H \leq p_H$. We may also note that in Lazear's treatment, crackdowns are inefficiently rare, in the sense that crackdowns are adopted for fewer values of P than is optimal (again, this follows by inspection of Figure 3). For values of P when crackdowns arise in Lazear's framework, the estimated deterrence effect as computed using Lazear's crackdowns exceeds the deterrence effect in the optimal scheme (in Figure 3, the slope of the "Lazear crackdown" is steeper than the slope of the optimal crackdown). Finally, under the optimal policing scheme the deterrence effect of addi-

---

upper range of our estimates, we get an implied value of a statistical life of 2.86 million dollars.

tional resources is subject to the law of diminishing returns: as P increases, its marginal effect on crime never increases (the deterrence effect is given by the slope of the lower envelope of the function $1 - F(pT)$). In Lazear's treatment, the law of diminishing returns does not hold: deterrence is generally non-monotonic in P.[47]
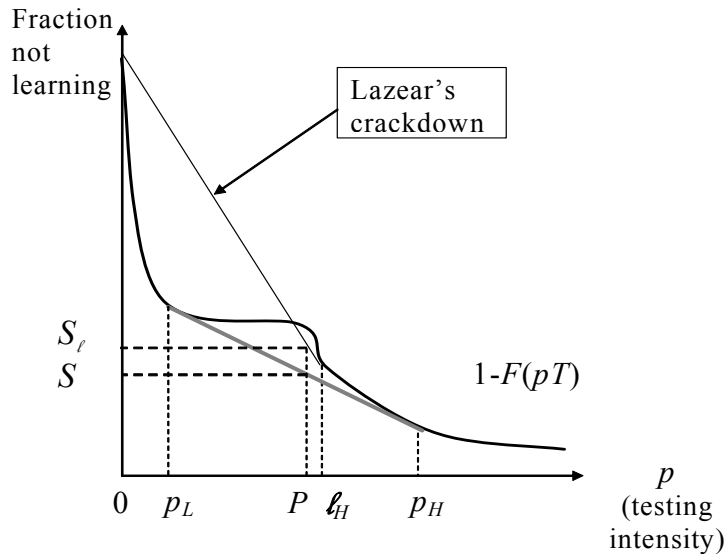

Figure 4. Optimal versus Lazear's Crackdown.

Despite the difference in approach, it is reassuring that in one respect the two analyses give the same prediction: as the budget constraint is relaxed, both approaches imply that the fraction q of the material subjected to high-intensity testing increases, and that the per-unit-of-material intensity of testing remains unchanged (even though the latter is suboptimal in Lazear, 2006).

Turning to other theories of deterrence, we acknowledge that crackdowns could be rationalized based on some form of non-standard (at least from an economist's viewpoint) estimate of probabilities. For example, suppose that absent a crackdown, the probability of monitoring is so small as to be ignored by the driver, and that crackdowns raise the driver's awareness of monitoring. If this increased awareness extends to periods

---

[47]Practical reasons do not often, in our view, dictate that we restrict attention to a scheme where $p_L = 0$. The Belgian police, for example, spend considerable resources on non-crackdown interdiction, thus making $p_L > 0$. In the context of academic testing, the optimal scheme could readily be implemented by telling students that a fraction $(1 - q)$ of the subject matter is *less likely* to be tested, but not necessarily out of bounds.

when a crackdown is not in force, crackdowns help reduce speeding.[48] Criminologists have justified crackdowns using a different theory of deterrence based on *subjective* risk assessment. This theory, developed in Ross (1984) and Sherman (1990), highlights the distinction between risk (which is accurately perceived by motorists) and uncertainty (which is not accurately perceived). According to this theory, an effective policing strategy would leave motorists in as much in doubt as possible as to the location and timing of the crackdowns, in order to maximize their uncertainty.

4.2   Extensions

The analysis of Section 2 extends to somewhat more general environments, namely: the case in which the utility from committing a crime is non-linear, and the case in which the decision to violate the law is not discrete (yes/no) but rather continuous (how much to speed). All the qualitative features of the solution to the original problem carry over to these extensions, as explained in a web appendix.

In the web appendix we also discuss two scenarios which are formally similar to the ones discussed in this paper. One is the case in which the social goal is the minimization of *undetected* crime.[49] The other is the case in which monitoring by an agency generates inefficiencies due to the exclusion of some consumers.[50]

---

[48] In our speeding application, this theory of deterrence would suggest that motorists on announcement days would be reminded of the possibility of being monitored and therefore slow down. According to this theory, speeding should also go down even on roads that are not monitored due to the increased alertness from announcements on other roads. This implication, however, is refuted by our data, because we find that the fraction of speeders on non-monitored roads does not decrease during monitoring days (see Table A4).

[49] Consider for example the problem of minimizing the amount of drugs, or illegal firearms, that make it to the retail market without being intercepted by patrol officers. Similarly, in tax auditing a plausible objective is to minimize undetected tax evasion.

[50] The analysis is developed within the real-world context of auditing by Radiotjänst, a private corporation that administers the Swedish TV-fee and checks that people are actually paying the fee. This agency implements widely publicized crackdowns, called *special fee-controls*, in 3-5 predetermined areas every year.

5 Conclusions

This paper introduced and analyzed a model of police interdiction. Within this model, we have fully characterized the optimal interdiction strategy. We have shown that even if all citizens look identical to the police, it may be rational for the police to divide the population into two (but no more) groups and monitor the groups at different intensities. For this division to be effective in curtailing crime, it is important that the group subjected to the crackdown be made aware when they are being monitored at the higher rate. This explains why police would announce when and where crackdowns will occur. Our analysis provided a rational choice explanation for pre-announced police crackdowns, which are regarded in the criminology literature as exploiting a non-rational perception of risk on the part of the citizens.

The model provides a behavioral setting in which monitoring intensities could differ across locations or groups in a way that is totally random, thus potentially providing a rigorous basis for using the observed variations in monitoring intensities in an observational data to estimate the deterrence effect without having to worry about endogeneity.

We applied our theoretical model to study speeding interdiction in Belgium. The data provide support for several implications of the model. Among these are, first, that the announcement strategy of the police indeed amounts to dividing the population into exactly two groups, and, second, that when police resources are increased, the frequency of crackdowns increases but the probability of being policed during a crackdown does not change. We used the model to estimate the deterrence effect of additional resources devoted to speeding interdiction in the form of 10,000 additional speeding tickets. Our calculations suggest that the marginal benefit, in terms of statistical lives saved, is close to the marginal cost of deterrence (it is exactly equal if we take the value of a statistical life to be 6.85 million dollars). Thus, the current level of speeding interdiction is arguably in line with socially optimal use of resources.

The standard theoretical approach we adopted—set up a stylized model and derive the optimal policy—turns out to be remarkably successful in explaining the policing behavior of a police department. What is remarkable is that the optimal interdiction strategy in our model has some features (crackdowns) which are unlikely to be put in place by chance, and which are found in the Belgian data. Thus, the rational theory of

optimal interdiction can have not only *normative* implications, but even *positive* corre-
lates in practical police work. If apparently abstract models can capture the behavior of
complex real-world institutions such as police departments, then this is good news for
the vast theoretical literature dealing with optimal enforcement.

Appendix A: Proofs

Theorem 1 *Let the function $f : [0, S] \rightarrow [0, 1]$ be continuous and strictly increasing. Let the function $g : [0, S] \rightarrow \mathbf{R}$ be continuous. Let $\mathcal{P}_g^*$ denote the set of probability distributions defined on the interval $[0, S]$ that solve the following linear problem*

$$\max_{\mu} \int_0^S f(p) \, \mu(p) \, dp$$

$$\text{s.t.} \quad \int_0^S g(p) \, \mu(p) \, dp \leq C. \tag{7}$$

For given f, let $\mathcal{F}_g$ denote the set of all functions f with the property that all $\mu^* \in \mathcal{P}_g^*$ place all the probability on one or two points in $[0, S]$. Then, the set $\mathcal{F}_g$ is dense in the set of all continuous functions g equipped with the supnorm.

If, moreover, the solution requires that probability mass be placed on two points in $[0, S]$, then the same two points receive all the probability when C is slightly increased.

Proof: Consider first the easy case in which the constraint is not binding at the optimal solution. In that case, a generic f will have exactly one strict maximum, and so the optimal μ^* will put mass one on exactly one point (the strict maximum).

Let us now consider the more difficult case in which the constraint is binding at the optimal solution. In that case, there exists a number $\lambda > 0$ such that μ^* maximizes the Lagrangean

$$\mathcal{L}(\mu, \lambda) = \int_0^S [f(p) - \lambda g(p)] \, \mu(p) \, dp + \lambda C.$$

We will show that, if $\mu^* \in \mathcal{P}_g^*$ puts positive mass on more than two points, then f is non-generic. To this end, let $\mathcal{A}$ denote the set of p's that is defined by

$$\mathcal{A} = \arg\max_p [f(p) - \lambda g(p)].$$

By definition of $\mathcal{A}$, there is a number M such that

$$f(p) - \lambda g(p) = M \quad \text{for } p \in \mathcal{A}$$
$$f(p) - \lambda g(p) < M \quad \text{for } p \notin \mathcal{A}$$

If μ^* puts positive mass on more than two points, then the cardinality of $\mathcal{A}$ would have to exceed 2. Consider the transformation $\varphi(p) = f^{-1}(p/S)$. The function φ is a one-to-one

mapping of $[0, S]$ onto itself. We can therefore write

$$f\left(\varphi\left(p\right)\right) - \lambda g\left(\varphi\left(p\right)\right) = M \quad \text{for } \varphi\left(p\right) \in \mathcal{A}$$
$$f\left(\varphi\left(p\right)\right) - \lambda g\left(\varphi\left(p\right)\right) < M \quad \text{for } \varphi\left(p\right) \notin \mathcal{A},$$

or, with the obvious meaning of symbols,

$$\frac{p}{S} - \lambda g\left(\varphi\left(p\right)\right) = M \quad \text{for } p \in \varphi^{-1}\left(\mathcal{A}\right)$$
$$\frac{p}{S} - \lambda g\left(\varphi\left(p\right)\right) < M \quad \text{for } p \notin \varphi^{-1}\left(\mathcal{A}\right).$$

Note that the set $\varphi^{-1}\left(\mathcal{A}\right)$ has the same cardinality of $\mathcal{A}$. Thus, if $\mathcal{A}$ has cardinality greater than 2, it means that the two numbers λ and M are such that the negatively-sloped straight line identified by $\frac{1}{\lambda}\left(\frac{p}{S} - M\right)$ is tangent to the function $g\left(f^{-1}\left(p/S\right)\right)$ at more than two points and never exceeds it. This means that there is a tangent hyperplanes to the set

$$Y = \left\{(p, y) : y \leq g\left(f^{-1}\left(p/S\right)\right)\right\}$$

which makes contact with the set Y at more than two points and has negative slope. We now show that, the set of f's such that this property does not hold is dense. To this end, and without loss of generality, let us assume that $S = 1$. Our task, then, is to show that if a negatively-sloped tangent hyperplanes to Y make contact with Y in more than two points, there is a function $\tilde{f}$ close to f with the property that no tangent hyperplane has more than two contact points. Let $\mathcal{H}$ denote the set of hyperplanes that have more than two contact points with Y. Elements of $\mathcal{H}$ are identified by their slope h. For every hyperplane $h \in \mathcal{H}$, take the sup and the inf of the first dimension of all its contact points and call those a_h and b_h. Consider now a continuous function $d_g\left(p\right)$ which is equal to 0 for every p unless $p \in \left(a_h, b_h\right)$ for some $h \in \mathcal{H}$, in which case $d_g\left(p\right)$ assumes values strictly between zero and 1. Let $\tilde{f}_\varepsilon\left(p\right) \equiv \left[1 + \varepsilon \cdot d_g\left(p\right)\right] \cdot f\left(p\right)$. For any $\varepsilon > 0$, the set $\tilde{Y}_\varepsilon = \left\{(p, y) : y \leq \tilde{f}_\varepsilon\left(p\right)\right\}$ has exactly the same set of tangent hyperplanes as Y. This follows from the fact that since the functions f and g are continuous, hyperplane h makes contact with Y at a_h and b_h. Moreover, by construction no hyperplane is tangent to $\tilde{Y}_\varepsilon$ at more than two points. Since the function $\tilde{f}_\varepsilon\left(p\right)$ can be made arbitrarily close to $f\left(p\right)$ in the supnorm by choosing ε to be small, the set $\mathcal{F}_g$ is dense.

Let us now turn to the second part of the statement. For given C, suppose that the solution requires placing probability mass on two points $p_L < p_H$. Then, the constraint

must be binding. To see this, define

$$p_m \equiv \arg\min_{p\in\{p_H,p_L\}} f(p)$$

$$p_M \equiv \arg\max_{p\in\{p_H,p_L\}} f(p).$$

Since f is strictly monotone, $f(p_M) > f(p_m)$, and the only reason why it is optimal to place any probability mass on p_m is to help satisfy the constraint. It must therefore be $g(p_M) > C > g(p_m)$. At the optimal solution, moreover, it cannot be optimal to place anything but the smallest probability mass on p_m so that the constraint is just satisfied (with equality). Denote by λ^* the Lagrange multiplier associated to this programming problem. Since f is strictly monotone, $\lambda^* > 0$.

Suppose now that the constraint is relaxed slightly, by increasing C to $\widetilde{C} = C + \varepsilon$ with ε a small positive number. The solution to the programming problem is a saddle point $\left(\widetilde{\mu}, \widetilde{\lambda}\right)$ for the Lagrangean. We now proceed to construct this saddle point. We start by keeping the Lagrange multiplier unchanged, $\widetilde{\lambda} = \lambda^*$. Because of this choice, the $\widetilde{\mu}$ that maximizes the Lagrangean still places probability mass on p_M and p_m only, which is what we wanted to prove. To conclude the proof we need to finish the construction of the saddle point. To this end, observe that in order for $\widetilde{\lambda} = \lambda > 0$ to minimize the Lagrangean, the Lagrangean must be constant with respect to λ, which is equivalent to

$$g(p_m)\,\widetilde{\mu}(p_m) + g(p_M)\,\widetilde{\mu}(p_M) = \widetilde{C} \tag{8}$$

Since $g(p_M) > C > g(p_m)$, for ε sufficiently small also $g(p_M) > \widetilde{C} > g(p_m)$. Therefore, it is possible to choose $\widetilde{\mu}(p_m)$ and $\widetilde{\mu}(p_M) = 1 - \widetilde{\mu}(p_m)$ so that equation (8) is satisfied. Choosing $\widetilde{\mu}$ accordingly concludes the proof.

Corollary 2 *If f is increasing and the solution requires that probability mass be placed on two points in $[0, S]$, the probability mass placed on the largest point increases when C is slightly increased.*

Proof. >From the proof of Theorem 1 we know that the constraint (7) is binding both at C and at $C + \varepsilon$. This means that for $c = C, C + \varepsilon$, the probability mass μ_c placed on p_H must solve

$$g(p_H)\,\mu_c + g(p_L)\,(1 - \mu_c) = c.$$

Since f is increasing, $p_M = p_H$ and thus $g(p_H) > g(p_L)$. The result follows. ∎

References

Traffic Safety Facts 2001: Speeding. NHTSA Report, DOT HS 809 480. On file with the authors.

Motor Vehicle Traffic Crash Fatality Counts and Injury Estimates for 2003. NHTSA Report, DOT HS 809 755 August 2004.

Ashenfelter, Orley and Michael Greenstone. "Using Mandated Speed Limits to Measure the Value of a Statistical Life." *Journal of Political Economy,* Vol. 112, No. 1, pt. 2, 2004.

Balkin, Sandy, and J. Keith Ord. "Assessing the Impact of Speed-Limit Increases on Fatal Interstate Crashes."*Journal of Transportation and Statistics,* April 2001, pp. 1-26.

Becker, G. (1968) "Crime and Punishment: An Economic Approach." *Journal of Political Economy,* 76, p. 169-217.

Blincoe, Lawrence J., Angela G. Seay, Eduard Zaloshnja, Ted R. Miller, Eduardo O. Romano, Stephen Luchter, Rebecca S. Spicer. (2002). *The Economic Impact of Motor Vehicle Crashes 2000.* NHTSA Technical Report, available online at http://www.nhtsa.dot.gov/people/economic/EconImpact2000/summary.htm

Campbell, B.J. (1988) The association between enforcement and seat belt use. *Journal of Safety Research,* 19 (4), p. 159-163

Chander, Parkash, and Louis L. Wilde (1998): "A General Characterization of Optimal Income Tax Enforcement," *The Review of Economic Studies* 65(1), January, 165-83.

DiGuiseppi C, Li L, Roberts I.(1998). "Influence of travel patterns on mortality from injury among teenagers in England and Wales, 1985-95: trend analysis." *British Medical Journal,* Mar 21, 316(7135). pp. 904-5.

Di Tella, Rafael, and Ernesto Schargrodsky (2002). "Political and Economic Incentives during an Anti-Corruption Crackdown" Manuscript, Harvard Business School, 2002.

Di Tella, Rafael, and Ernesto Schargrodsky (2004), "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack", *American Economic Review* 94(1) 115-133.

Ehrlich, Isaac (1987) "Crime and Punishment," in *The New Palgrave: A Dictionary of Economic Theory and Doctrine,* John Eatwell, Murray Milgate, and Peter Newman, editors, London: Macmillan Press, 1987.

Finch D J, Kompfner P, Lockwood C R and Maycock G (1994). Speed, speed limits and accidents. Project Report PR58. Crowthorne: TRL Limited.

Homel, R.J. (1988a) *Policing and punishing the drinking driver: a study of general and specific deterrence.* New York, Springer-Verlag, 1988.

Knowles, John, Nicola Persico and Petra Todd (2001): "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, February.

Lazear, Edward P.(2004). "Speeding, Tax Fraud, and Teaching to the Test." NBER Working Paper No. 10932, November 2004.

Levitt, Steven D. (1997): "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *The American Economic Review* 87(3), 270-91.

Levitt, Steven D. and Jack Porter (2001): "How Dangerous are Drinking Drivers?" in *Journal of Political Economy*, vol. 109, no. 6, December, 2001, 1198-1237.

Murphy, K and Topel, R. (2002): "Diminishing Returns? The Costs and Benefits of Improving Health," University of Chicago and NBER, November 2002, working paper.

Peltzman, Sam (1975) "The Effects of Automobile Safety Regulation." *The Journal of Political Economy*, Vol. 83, No. 4. (Aug., 1975), pp. 677-726.

Persico, Nicola (2001). "Racial Profiling, Fairness, and Effectiveness of Policing." *The American Economic Review* 92(5), December 2002, pp. 1472-97.

Polinsky, A. Mitchell and Steven Shavell. (1979). "The Optimal Trade-off between the Probability and Magnitude of Fines." *The American Economic Review*, 69, pp. 880–91.

Polinsky, Mitchell, and Steven Shavell. "The Economic Theory of Public Enforcement of Law." *Journal of Economic Literature* Vol. XXXVIII (March 2000) pp. 45–76

Pratap, Sangeeta, and Tridib Sharma (2002). "The Nutrition Curve and Intra Household Allocation of Calories," Manuscript, ITAM.

Prendergast, Canice (2001), "Selection and Oversight in the Public Sector, With the Los Angeles Police Department as an Example," NBER Working Paper 8664.

Redelmeier DA, Tibshirani RJ, Evans L (2003). Traffic-law enforcement and risk of death from motor-vehicle crashes: case-crossover study, *Lancet*; 361(9376): 2177-2182.

Ross, H. Lawrence. *Deterring the Drinking Driver: Legal Policy and Social Control.* Lexington, MA, Lexington Books, 1984.

Ross, H.L. & Gonzales, P. (1988). Effects of Licence Revocation on Drunk-Driving Offenders. *Accident Analysis and Prevention,* 20,379-391.

Shavell, Steven. "A Note on Marginal Deterrence." *International Review of Law and Economics* (1992) 12, 345-355

Sherman, Lawrence W. "Police Crackdowns: Initial and Residual Deterrence." In M. Tonry and N. Morris (eds.) *Crime and Justice: A Review of Research*, vol. 12 (1990), Chicago: University of Chicago Press.

Sherman, Lawrence W, and David Weisburd "General Deterrent Effects of Police Patrol in Crime Hot Spots': a Randomized, Controlled Trial." *Justice Quarterly*, vol. 12(4), December 1995, pp. 625-48.

Shi, Lan "Does Oversight Reduce Policing? Evidence from the Cincinnati Police Department After the April 2001 Riot." Working Paper, University of Washington, Seattle, January 28, 2005.

Stigler, George (1970). "The Optimum Enforcement of Laws." *Journal of Political Economy* 78:526-36.

Taylor, M C, A Baruya and J V Kennedy.(2002) "The relationship between speed and accidents on rural single-carriageway roads." Report Prepared for Road Safety Division, Department for Transport, Local Government and the Regions. TRL Report TRL511. On file with the authors.

Zaal, Dominic "Traffic Law Enforcement: a Review of the Literature." April 1994 Report No. 53 Monash University Accident Research Centre.

Table 1a
Number and Percentage of Vehicles Subject to
Announced or Unannounced Monitoring by Year§

	2000	2001	2002	2003 (first half of year)
Announced	266,240 (39.5%)	394,540 (55.2%)	1,746,340 (76.8%)	1,777,977 (60.9%)
Unannounced	406,941 (60.5%)	319,650 (44.8%)	526,422 (23.2%)	1,139,428 (39.1%)
Total	673,181	714,190	2,272,762	2,917,405

§Based on daily observations on the number of vehicles, as recorded by the radar control machines.

Table 1b
Number of Ticketed Speeders by Year and Announced or Unannounced Monitoring
(percentage of total ticketed each year shown in parentheses)§

		2000	2001	2002	2003 (first half of year)
Announced	Number ticketed	12677 (37.3%)	20039 (44.3%)	55804 (71.4%)	32079 (69.4%)
Unannounced	Number ticketed	21274 (62.7%)	25225 (55.7%)	22332 (28.6%)	14140 (30.6%)

§ Based on daily observations on vehicles passing by the radar control machines and on their speeding status, as recorded by the machine.

Table 1c
Number of Announced (Ann) and Total Monitoring Events
by Highway and Year

	2000		2001		2002		2003 (first half of year)	
	Ann	Total	Ann	Total	Ann	Total	Ann	Total
A10	18	46	23	52	181	214	125	158
A14	38	138	51	105	156	244	150	218
R4	10	34	1	24	0	5	0	0
Total	66	218	75	181	337	463	275	376
μ_h†	0.012		0.014		0.062		0.100	

† μ_h is calculated as the number of announced monitoring events divided by the total number of potential monitoring events (5475 in 2000-2002 and 2715 in the first half of year 2003), which includes intervals when monitoring took place as well as intervals in which no monitoring took place). 5475 is calculated as number of road segments (5) times the number of intervals per day (3) times the number of days per year (365).

Table 2
Means of Variables†

Variable	Mean (Std. err)
Fraction speeding	0.04 (0.004)
Fraction vehicles traveling on highway A10	0.43 (0.010)
Fraction vehicles traveling on highway A14	0.56 (0.01)
Fraction vehicles traveling on highway R4	0.02 (0.003)
Quarter 1 (Jan, Feb, Mar)	0.31 (0.009)
Quarter 2 (Apr, May, Jun)	0.34 (0.010)
Quarter 3 (Jul, Aug, Sep)	0.15 (0.007)
Quarter 4 (Oct, Nov, Dec)	0.21 (0.008)
Year 2000	0.11 (0.006)
Year 2001	0.11 (0.006)
Year 2002	0.40 (0.01)
Year 2003	0.37 (0.01)
Heavy traffic density	0.02 (0.003)
Medium traffic density	0.001 (0.0007)
Moderate traffic density	0.006 (0.002)
Light traffic density	0.97 (0.003)
Weekday morning rush hour	0.02 (0.003)
Weekday evening rush hour	0.29 (0.009)
Holiday	0.08 (0.005)

† Observations are monitoring intervals. Means are weighted by the number of vehicles in each interval. There are 1238 total monitoring intervals (see Table 1c for breakdown by year)

Table 3
Estimated Coefficients from Logistic Regression of Probability of Speeding§
Dep. Var: Indicator for Whether Speeding
(Standard errors shown in parentheses)

	Model Specification		
Variable	(1)	(2)	(3)
Intercept	-3.43	-3.46	-3.41
	(0.007)	(0.02)	(0.02)
Indicator for highway A14	0.58	0.27	0.19
	(0.009)	(0.008)	(0.008)
Indicator for highway R4	0.35	-0.16	-0.23
	(0.02)	(0.02)	(0.02)
Announcement on highway A10	…	…	-0.18
			(0.009)
Announcement on highway A14	…	…	-0.11
			(0.006)
Announcement on highway R4	…	…	-0.04
			(0.05)
Probability of monitoring – A10‡	-0.93	-0.28	…
	(0.04)	(0.04)	
Probability of monitoring – A14‡	-0.96	-0.39	…
	(0.02)	(0.02)	
Probability of monitoring – R4‡	-0.38	-0.08	…
	(0.15)	(0.15)	
Medium traffic density†	…	-0.28	-0.25
		(0.06)	(0.06)
Moderate traffic density†	…	0.24	0.27
		(0.03)	(0.03)
Light traffic density†	…	0.22	0.24
		(0.01)	(0.01)
Indicator for weekday morning rush hour	…	-0.45	-0.44
		(0.02)	(0.02)
Indicator for weekday evening rush hour	…	0.02	-0.02
		(0.006)	(0.006)
Indicator for holiday	…	0.51	0.49
		(0.007)	(0.007)
Includes fixed effects for days of week	No	Yes	Yes
Includes fixed effects for months of year	No	Yes	Yes
Includes fixed effects for year	No	Yes	Yes
Number of vehicle observations	5,641,522	5,641,522	5,641,522
% correctly classified under the model	44.0%	48.02%	48.3%

§ The unit of observation is a vehicle passing by the radar control machine. The speeding status is recorded by the machine.

† The omitted category is high traffic density.

‡ The probability of monitoring was obtained using the procedure described in the text, which takes into account the expected probability of monitoring, the expected length of time spent monitoring, and uncertainty about the direction of the road being monitored.

Table 4
Decrease in Speeding Attributable to a Crackdown and the Deterrence Effect
of Increasing the Number of Tickets

Avg monitoring probability§		Model Specification		
	Highway A10	(1)	(2)	(3)
p_H=0.2007 p_L=0.0084	Predicted % speeding on days without announcements (a)	3.1	3.0	3.1
	Decrease in % speeding due to the announcement (b) †	0.20	0.15	0.48
	Percentage decrease in rate of speeding *	6.5%	5.0%	15.5%
	Implied slope of 1-F ¶	-1.04%	-0.78%	-2.50%
	Effect of additional 10,000 tickets on number of speeders £	-3,597	-2,743	-9,603
	Highway A14			
p_H=0.2441 p_L=0.0167	Predicted % speeding on days without announcements (a)	5.4	5.0	5.0
	Decrease in % speeding due to the announcement (b) †	1.01	0.40	0.48
	Percentage decrease in rate of speeding *	18.7%	8.0%	9.6%
	Implied slope of 1-F ¶	-4.44%	-1.76%	-2.11%
	Effect of additional 10,000 tickets on number of speeders £	-10,291	-3,849	-4,707
	Highway R4			
p_H=0.2576 p_L=0.0091	Predicted % speeding on days without announcements (a)	4.4	4.3	4.2
	Decrease in % speeding due to the announcement (b) †	0.38	0.08	0.16
	Percentage decrease in rate of speeding *	8.6%	1.9%	3.8%
	Implied slope of 1-F ¶	-1.53%	-0.32%	-0.64%
	Effect of additional 10,000 tickets on number of speeders £	-3,817	-763	-1,596

§ The monitoring probabilities p_H and p_L are derived following the procedure described in the text, which takes into account the probability of monitoring, the expected length of time spent monitoring, and uncertainty about the direction of the road being monitored. p_H is the average probability of being monitored during an announcement period and p_L is the probability of being monitored when there was no announcement.

† Change in speeding rate implied by the estimated coefficients from the logistic regression model (Table 3) when the monitoring probability increases from p_L to p_H.

* (b)/(a)*100

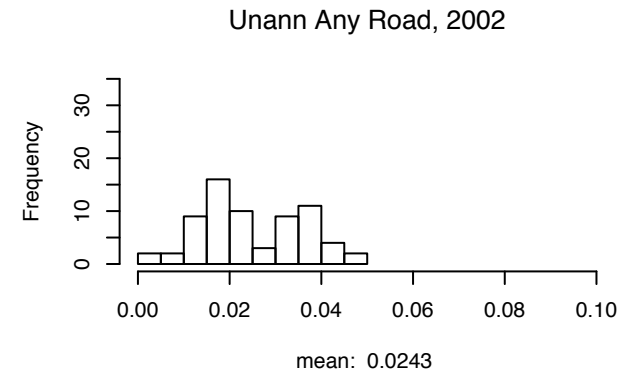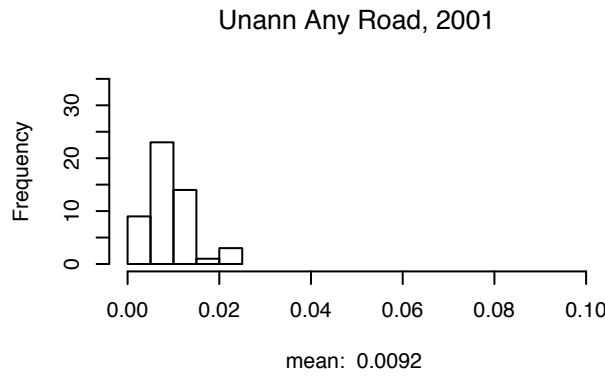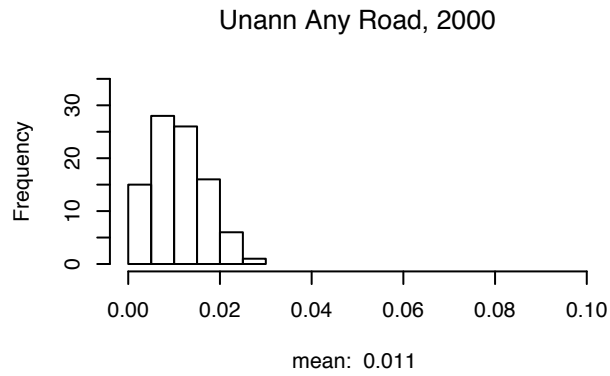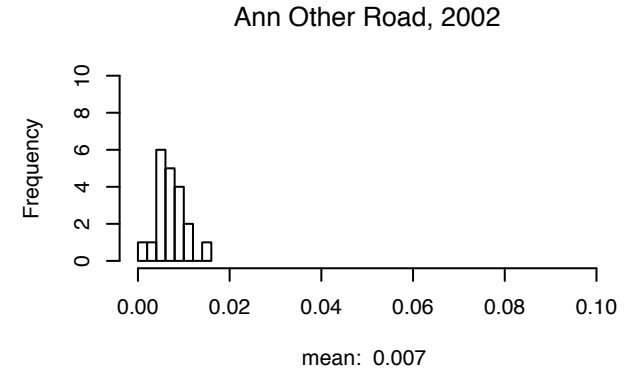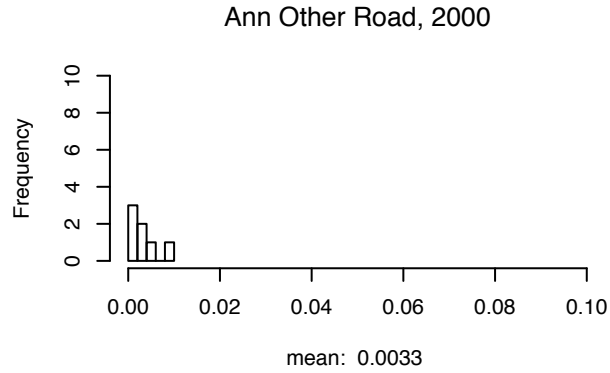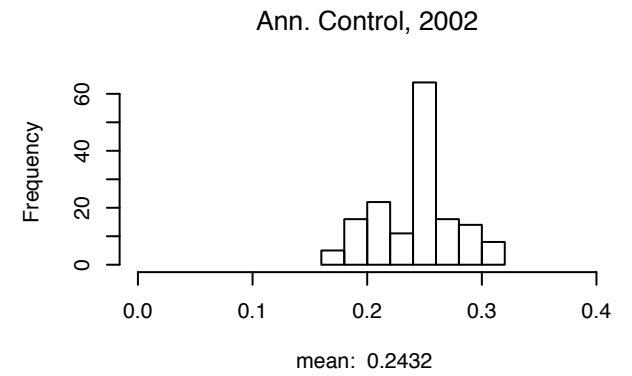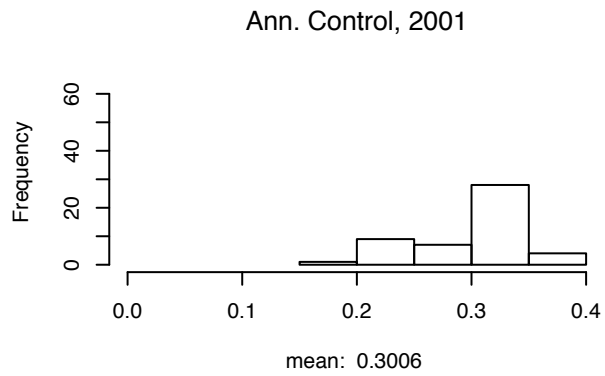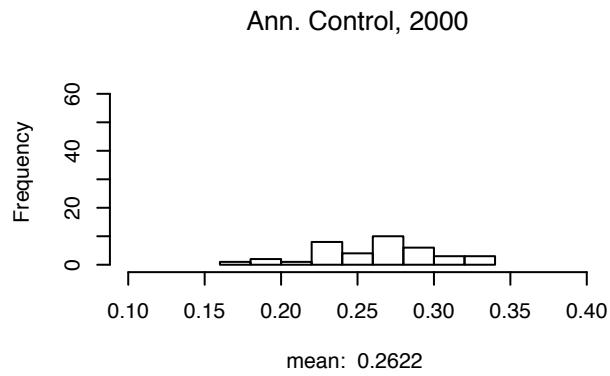¶ Slope = (b)/(p_L - p_H)

£ As described in the text.
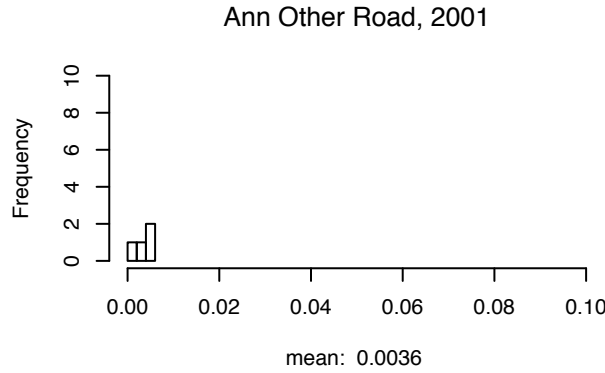
Figure 3: Probability of being monitored, highway A14