

A Rational Theory of Random Crackdowns*

Jan Eeckhout, Nicola Persico, and Petra Todd

University of Pennsylvania[†]

July 2005

Abstract

This paper develops an incentives-based theory of policing that can explain the phenomenon of random “crackdowns,” which are intermittent periods of especially high interdiction/surveillance. We show that, when police minimize the crime rate, random crackdowns can emerge as part of an optimal policing strategy. We consider several variations of the basic policing model that would apply in different monitoring situations, such as speeding or 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 several implications of the crackdown theory using traffic data gathered by the Police Department in Belgium, and we use the model to estimate the deterrence effect of additional resources spent on speeding interdiction.

*We are grateful to Larry Samuelson for useful discussions. This paper was improved by comments received at the University of Pennsylvania IER crime conference, at the Northwestern NSF conference on profiling, and at seminars at UBC, Columbia, Duke, and Brown. We are also grateful to the Belgian police department for the data and for their assistance in interpreting the data. This project was supported by the National Science Foundation Grant SES-0422862.

[†]Department of Economics, 3718 Locust Walk, Philadelphia, PA 19104. Emails: eeckhout@ssc.upenn.edu, persico@ssc.upenn.edu, petra@athena.sas.upenn.edu.

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 to higher intensity police monitoring 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. Second, they are *publicized*, i.e., those who are subjected to it are informed *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

Being arbitrary and publicized, crackdowns may seem an inefficient deployment of police resources; potential criminal activity could merely be displaced to non-crackdown periods or locations. Criminologists 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 the potential criminals to

¹Our 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.

²For 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.

overestimate the risk of detection during non-crackdown periods.³ This view relies on the potential criminals' expectations being systematically wrong, and so is inconsistent with rationality.

In this paper we take a different approach and develop a model in which potential criminals are not fooled 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, *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.

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. Eye color is distributed independently of the propensity to commit a crime, so it is arbitrary for police to treat citizens differently according to eye color. 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

³Sherman (1990), Ross (1984). Sherman, p. 11, recommends that crackdowns be highly publicized and be followed by secret "backdowns," and warns of the risk of exhausting the bluff through overuse.

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 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. 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 crackdowns and non-crackdowns groups. In the example, citizens are assumed to be unable to disguise 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. In addition, it yields some testable implications that can be used to validate the model.

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 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 occasionally 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 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 random fashion across different stretches of the roads and time periods. We interpret the announced controls as crackdowns on particular groups of motorists, those travelling on the announced stretch of the road at the announced time. We measure of 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 theory, we are able to calculate the effect of more interdiction on speeding. In conjunction with value-of-life estimates, our results suggest that at the current level of interdiction, the marginal benefit, in terms of statistical lives saved, is close to the marginal cost of interdiction.

The paper develops as follows. Section 2 presents a theoretical model that we

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

⁵See *Traffic Safety Facts 2001*.

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

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

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

use to study the conditions under which crackdowns emerge as an optimal policing strategy when the police chief is given incentives to minimize crime. The model allows for unobserved heterogeneity in the benefits citizens get from breaking the law, under the assumption that the police chief knows the distribution of the unobserved variable. Section 3 extends the model to consider alternative police incentives and constraints including the following: the police chief minimizes undetected crime instead of overall crime; crime is carried out in teams; the police chief is constrained in terms of the number of successful interdictions; the police chief maximizes a social welfare function taking into account the benefits of breaking the law. We discuss potential applications of these variations of the model in drug or firearms policing and in screening to deter terrorism. Section 4 of the paper applies the model developed in section two to data that we obtained from the Belgian police department. Section 5 provides a discussion and some further extensions. Section 6 concludes.

1.1 Related Literature

The idea that deterrence may be improved by focussing interdiction on arbitrary subsets of the population is present in the literature on racial profiling (see Persico, 2001). Recently, Lazear (2004) 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 another 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., focussing 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

⁹Coincidentally, Lazear (2004) also uses speeding as a potential application in developing his argument, but his and our work are independent.

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 disparities. Our paper also compares the strategies adopted by the police under different incentive schemes, although it 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 policing¹² and of greater penalties¹³ on speeding. Parallel literatures deal with the deterrence effect of increased policing and greater penalties on drunk driving¹⁴ and on seat-belt wearing.¹⁵ There is also a literature studying the connection between risk-taking behavior (speeding, drunk driving) and accidents; see Levitt and Porter (2001) for a good example in the context of drunk driving. Ashenfelter and Greenstone (2004) use changes in the speed limit across US states to estimate the value of a statistical life. That paper also provides 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 Schargrotsky (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.

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

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

¹²See e.g. Redelmeier et al (2003) and the literature cited therein.

¹³See Homel (1988a), Ross and Gonzalez (1988).

¹⁴See e.g. Ross (1984).

¹⁵See e.g. Campbell (1988), Peltzmann (1975).

2 Benchmark Policing Model

In this section, we consider 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.¹⁶ 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). Because the principal only observes the crime rate, it is natural for the principal to give the police chief incentives to minimize crime. 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 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 .¹⁷ We assume that $p \in [0, \bar{p}]$, which implies that a citizen can be monitored with probability no greater than some $\bar{p} \leq 1$.

¹⁶Besides 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 3.4.

¹⁷Here we assume that the citizen's utility functions are linear and that committing the crime is a discrete decision, but in Section 5 we relax these assumptions. Also, for the moment we take T as given. We will return to this issue in Section 5.

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 with which he/she is policed.¹⁸ 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 $\int_0^{\bar{p}} \mu(p) dp = 1$.

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:

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

In this section, we assume that the police's goal is to minimize the total number of criminals. Alternative specifications of the objective function are studied in Section 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 devoted to crime interdiction). 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

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

We refer to this constraint as a *time constraint*. An alternative way of specifying the constraint on police resources is explored in Section 3.3. The police chooses a probability measure μ to minimize the number of criminals (2) subject to the resource constraint (3).

¹⁸In practice, this means that the police must inform citizens of the intensity with which they are policed.

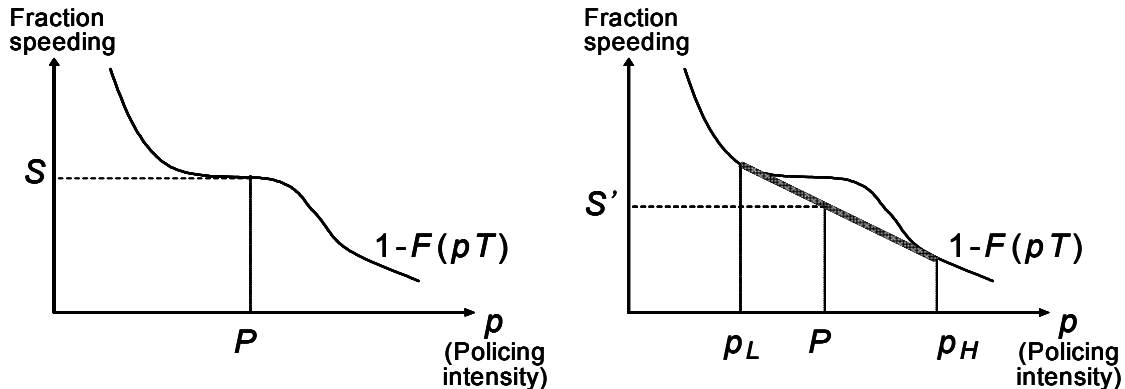


Figure 1: A crackdown is optimal.

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. A formal proof of these propositions is provided 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.

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

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

Of course, crackdowns are not always part of the optimal policing strategy. If, for example, the function $1 - F(pT)$ 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.

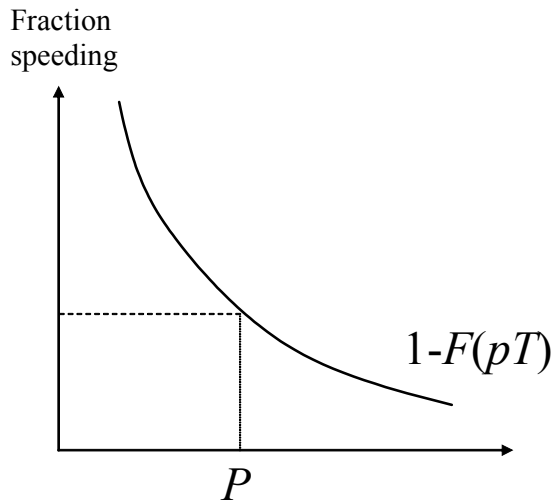


Figure 2: Crackdowns are not optimal.

When crackdowns are optimal, it is because the function $1 - F(pT)$ 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 $1 - F(pT)$, 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 at the same rate, or dividing the population into at most two groups to be monitored with different intensities.*

Proof. Theorem 1 provides a formal proof of this result. ■

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 $(\bar{p}, 1 - F(\bar{p}T))$, which means that for any P we have $p_L = 0, p_H = \bar{p}$. 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.¹⁹ This observation gives rise to the following remark.

Remark 1 *If F is convex on its domain, then for any $P \in (0, \bar{p})$ 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 $(0, \bar{p}T)$, 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 is noted in the following proposition.

¹⁹This 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 $\tilde{P} \in (P, p_H)$. 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).*

Proof. Theorem 1 provides a formal proof of this result. ■

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 $\tilde{P} - P$ provides a way of estimating the expected decrease in crime due to a hypothetical increase in police resources from P to \tilde{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 $1 - F(\tilde{P}T)$, 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 $1 - F(\tilde{P}T)$ 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 focusses 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.²⁰

If, in addition to observing crackdowns, if one also has access to exogenous variation in total police resources P , then Proposition 2 yields a testable implication of the model we developed. 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 4.3, we verify this implication in the context of speed interdiction. As noted in the introduction, a key assumption is that the citizens that are subject to increased interdiction as a result of crackdown cannot avoid being monitored. For our application of speed interdiction, we concentrate attention on highways, where avoiding the monitoring would imply an extra time cost that which does not compensate for the time gains from speeding.

3 Crackdowns under alternative policing situations

In the previous section, we assumed that the police chief is given a certain amount of resources along with incentives (explicit or implicit) to minimize the crime rate. In this section we explore alternative formulations of the agency problem between the police chief and his principal. Our goal is to provide a more precise fit for applications in which the police objective function or the resource constraint is not adequately captured by the benchmark model. For example, if the principal cannot observe the crime rate, the police must be given incentives based on some other measure of performance. For each variant of the model, we will mention potential applications.

Our discussion focusses mainly on comparing alternative models in terms of the likelihood that crackdowns arise as part of the policing strategy.²¹ Key to our analysis is that

²⁰One can think of crackdowns as being a case where exogenous variation in P arises as part of the optimal policing strategy.

²¹In comparing the models in Sections 2.1, 3.1, and 3.3, we will take the distribution of the propensity to commit a crime as given. These comparisons, therefore, are of interest (a) if we are comparing alternative systems of police incentives in how they reduce levels of the same crime, or (b) we are willing to accept that the distribution of the propensity to commit a crime is somewhat similar across different types of crime and policing situations, which could be the case if one variable (income, or education, say) determined the propensity to commit various types of crimes.

the models in Sections 2.1, 3.1, 3.2, 3.3, and 3.4 share a common “structural” feature: all five models give rise to programming problems that are linear in μ . This linearity explains why crackdowns arise in all the models.

3.1 First variant: minimizing undetected crime

Suppose the principal cares about reducing the crime of drug production, but the principal can only observe the drugs that make it to the market without being intercepted. In that case, police performance will have to be evaluated based on undetected crime. Sometimes, minimization of undetected crime may even arise as a first-best option. For example, a principal may find it optimal to give incentives to the police based on *undetected* crime when detection removes the social cost of the crime. This is the case for example with illegal firearms, where if a firearm is intercepted it is taken off the street.

When a group is monitored with intensity p , the fraction of crime in the group that goes undetected is $(1 - p)(1 - F(pT))$. Given a policing strategy μ , undetected crime is given by

$$\int_0^{\bar{p}} \mu(p) (1 - p) (1 - F(pT)) dp. \quad (4)$$

The police chooses a policing strategy μ to minimize expression (4) subject to the budget constraint (3).

This programming problem is very similar to the one studied in Section 2.1; there as well as here, the objective function is decreasing in p . This was the only property of the objective function that was used in Section 2.1, so it is immediate that Propositions 1 and 2 continue to hold in this setting.

Whether crackdowns are optimal depends, as before, on the convexity of the objective function. In the present case, it is the convexity of undetected crime that matters. If undetected crime is convex in p then crackdowns are never optimal (see Remark 1.) It is a simple to verify that undetected crime is “more convex” than crime, in the sense that if $(1 - F(pT))$ is convex then $(1 - p)(1 - F(pT))$ is also convex. Therefore, if F is such that police minimizing overall crime rates never finds it optimal to engage in crackdowns, then crackdowns are also not optimal if the objective is to minimize undetected crime. These observations are collected in the following proposition.

Proposition 3 *Suppose the police minimizes undetected crime. 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 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 $\tilde{P} \in (P, p_H)$. 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) There are no crackdowns for any P if the same is true when the police minimizes crime. The converse is not true.

Part (c) of the proposition suggests that crackdowns have drawbacks when the police minimizes undetected crime, and are therefore less likely to be part of the optimal strategy. A simple example provides some intuition for this result. Suppose $T = \bar{p} = 1$, and $P = 1/2$. Suppose further that F is a Uniform on $[0,1]$. When everyone is policed with intensity $1/2$ (no crackdown), half the citizens commit a crime and half of those criminals are undetected, so undetected crime equals $1/4$. Consider now an extreme crackdown in which half the citizen know that they are monitored for sure, and the rest know that they are never monitored. The latter half of the citizens all commit a crime, and undetected crime is equal to $1/2$. Undetected crime, therefore, increases due to the crackdown. This example shows the drawbacks of crackdown when the objectives is to minimize undetected crime: the group that is monitored less intensely commits a lot of crime, and that crime is more likely to go undetected.

The same example helps illustrates the “converse not true” statement in part (c). In the example, the crime rate is $(1 - p)$, which is linear in p . This means that any policing strategy including crackdowns is optimal. Undetected crime is $(1 - p)^2$ which is strictly convex, so crackdowns are strictly suboptimal. In this example, then, crackdowns are suboptimal if police minimizes undetected crime, but they are optimal (albeit weakly)

when police minimizes crime.²²

3.2 Second variant: detecting a criminal team

In this section we consider the problem of deterring a team of criminals instead of only one criminal. We assume that the team is detected as soon as one of its members is detected, and then the crime is averted. We call this a *team detection* setting, and we are interested in the prevalence of crackdowns as the team size changes. We will show that as the team size increases, crackdowns are less likely to be optimal.

As a potential application, consider the problem of airport screening for terrorists. Suppose there is a mastermind who has to decide whether to send an armed team of $n \geq 1$ terrorists on a plane. If all n terrorists pass the pre-flight screening then the terrorist act is carried out and the mastermind receives a utility v . If at least one is detected, then the boarding process is interrupted and the act cannot be carried out, whereupon the mastermind receives a utility of $-T$. The utility v from carrying out the act is unknown and is distributed according to a distribution F .

In the airport screening setting, we could think of a crackdown as a preannouncement that passengers on certain flights will be subject to high intensity monitoring.²³ We assume that such announcements could be made shortly before the boarding process, so that the mastermind cannot redirect the team to a non-crackdown flight without arising suspicion. The only choice that the mastermind faces is whether to send the armed team or to abort the mission.²⁴

Let p be the probability that each member of the team is detected during the pre-flight screening. If a team is sent, it will be successful with probability $(1 - p)^n$. A mastermind with value x will send the team if and only if

$$x(1 - p)^n - T(1 - (1 - p)^n) > 0,$$

²²Slightly tweaking the function F would ensure that crackdowns are *strictly* optimal when the police minimizes crime.

²³In is not the current practice to make such announcements; however, our analysis suggests that such announcements could be an optimal monitoring strategy, perhaps in conjunction with other passenger profiling practices.

²⁴Aborting may mean sending the team through unarmed so as not to arouse suspicion by canceling.

or, equivalently, if $x > T \left(\frac{1}{(1-p)^n} - 1 \right)$. From the police chief's viewpoint, the probability that the team is sent is

$$1 - F \left(T \left(\frac{1}{(1-p)^n} - 1 \right) \right).$$

The probability that a team is sent and is not detected is given by

$$(1-p)^n \cdot \left[1 - F \left(T \left(\frac{1}{(1-p)^n} - 1 \right) \right) \right].$$

It is useful to rewrite this probability as

$$g(h_n(p)),$$

where we have denoted

$$\begin{aligned} g(h) &= h \left[1 - F \left(T \left(\frac{1}{h} - 1 \right) \right) \right] \\ h_n(p) &= (1-p)^n. \end{aligned}$$

The police chief chooses $\mu(p)$, the probability that a flight is screened with intensity p , to minimize

$$\int_0^{\bar{p}} \mu(p) \cdot g(h_n(p)) \, dp.$$

under the budget constraint given by (3). The budget constraint reflects the fact that there may not be enough resources to subject all flights to a thorough screening process.

The programming problem is very similar to the ones studied in Sections 2 and 3.1. Since $g(\cdot)$ is an increasing function and $h_n(\cdot)$ a decreasing one for every n , the kernel of the objective function is decreasing in p . So Propositions 1 and 2 continue to hold in this setting. Crackdowns can be optimal only if there are portions of the objective function $g(h_n(p))$ that are concave in p . Since $g(\cdot)$ is increasing and $h_n(\cdot)$ is convex for every n , the composite function $g \circ h_n$ can have a concave portion only if the function $g(\cdot)$ has a concave portion. In particular, if $g(h_1(p))$ is everywhere convex, then crackdowns are never optimal for any n . This discussion is summarized in the following proposition.

Proposition 4 *Suppose the police minimizes undetected crime in a team detection setting. 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 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 $\tilde{P} \in (P, p_H)$. 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) *If there are no crackdowns for any P and for $n = 1$, then the same is true for any $n > 1$.*

We interpret part c) of the above proposition as indicating that crackdowns are not more likely to be effective when the setting is one of team detection.

Remark 3 *The logic behind Proposition 4 does not hinge on the specific functional form of g , that is, on whether the police minimizes undetected crime. Analogous statements can be proved for any objective function that is decreasing in the monitoring intensity p .*

3.3 Third variant: constraint on successful interdictions

In this variant of the model, as in Section 2.1, we assume that the goal of the police is to minimize the crime rate. The resource constraint, however, is given in terms of number of *successful* interdictions rather than in terms of time resources. That is, only monitoring *criminals* has a cost to the police; 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 4, where the police are administratively restricted in the number of tickets that they can write in a year.²⁵

²⁵This makes sense because detecting speeders with automatic radar machines is almost costless relative to processing a traffic ticket. More on this in Section 4.

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

$$\int_0^{\bar{p}} \mu(p) (1 - F(pT)) p dp \leq C. \quad (5)$$

The police minimizes expression (2) subject to the constraint (5).

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, reads

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

The present problem shares a key formal similarity with the benchmark model: both programming problems are linear in μ . As a consequence, Propositions 1 and 2 continue to apply.

True, constraint (5) and constraint (3) are quite different. In constraint (5), 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 5 *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 $\tilde{C} \in (C, (1 - F(p_H T)) \cdot p_H)$. 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 (5) 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 (5). The second effect was not present in the benchmark problem.

Part (b) of Proposition 5 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 later, we derive it here. The crime rate given \tilde{C} equals

$$\tilde{\mu}(p_L) \cdot (1 - F(p_L T)) + \tilde{\mu}(p_H) \cdot (1 - F(p_H T))$$

(note that there is no tilde over the p 's in light of Proposition 5 part (b)). To obtain the change in crime, subtract from this expression the analogous expression when resources equal C . This yields

$$\begin{aligned} \Delta \text{Crime} & & (6) \\ &= [\tilde{\mu}(p_L) - \mu(p_L)] (1 - F(p_L T)) + [\tilde{\mu}(p_H) - \mu(p_H)] (1 - F(p_H T)) \\ &= [\tilde{\mu}(p_H) - \mu(p_H)] (F(p_L T) - F(p_H T)). \end{aligned}$$

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

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

Isolating $\tilde{\mu}_H$ yields

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

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

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

Substituting into (6) we get

$$\begin{aligned}\Delta\text{Crime} &= (\tilde{C} - C) \left[\frac{F(p_L T) - F(p_H T)}{(1 - F(p_H T)) p_H - (1 - F(p_L T)) p_L} \right] \\ &= \Delta C \cdot \left[\frac{(\text{crime rate}|p_H) - (\text{crime rate}|p_L)}{(\text{crime rate}|p_H) \cdot p_H - (\text{crime rate}|p_L) \cdot p_L} \right].\end{aligned}\tag{7}$$

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 4 this slope is calculated in the context of highway speeding interdiction.

3.4 Fourth variant (the social planner’s problem): maximizing social welfare

Up to now the benefits of committing a crime have not been directly reflected in the police incentive scheme. We have thus implicitly assumed that if the principal has any regard for the citizens’ benefits from committing a crime, the trade-off between the benefits and costs of crime is resolved by choosing the appropriate size of police resources. This assumption seems realistic because, in practice, the police seldom appear to explicitly take into account the law-breakers’ utility when deciding their optimal strategy. (In our empirical application, for example, we know that the police view their mandate as simply to reduce speeding.) Nevertheless, taking into account the benefits of breaking the law makes theoretical sense, especially when we consider small stakes violations or “victimless” crimes such as drug consumption. In this section, we explore a model in which the presence of individual benefits of crime is directly reflected in police incentives.

Let us assume that the police chief’s incentives are to minimize a weighted sum of the crime rate and the loss to the deterred citizens of forgoing crime. We will call this the social planner’s problem and, in this section, refer to the police chief as the social planner. According to our model, if a group is policed with intensity p all citizens with $x < pT$ will forgo crime. The loss from forgone crime is therefore

$$L(p) \equiv \int_0^{pT} x dF(x).$$

Our social planner chooses μ to minimize

$$\int_0^{\bar{p}} \mu(p) [(1 - F(pT)) + \alpha L(p)] dp$$

subject to the budget constraint (3). The parameter $\alpha > 0$ represents the weight given by the social planner to the potential perpetrator's loss from forgone crime relative to the potential victims' benefit of reducing crime (the latter being normalized to one).

The social planner problem is more complicated than the one treated in the previous section, because the planner must weigh the cost of lost crime opportunities, and this cost is higher in the crackdown group (where more crime is forgone) than in the non-crackdown group. This consideration will affect the incentives to engage in crackdowns. When α is close to zero, i.e., the weight assigned by the social planner to the loss from forgone crime is small, the planner's objective function is similar to the one analyzed in Section 2 and so crackdowns are optimal under approximately the same conditions. When α is large, i.e., the social planner assigns large weight to the loss from forgone crime, the planner's objective function is no longer similar to the one treated in Section 2; for example, it is no longer necessarily decreasing in p , reflecting the fact that deterring crime now carries a social cost. For the purpose of assessing whether random crackdowns are optimal, what matters is the concavity of the planner's objective function. If the function exhibits concave portions in areas where the convex hull is decreasing in p , then crackdowns are optimal for some values of P . The presence of the term $L(p)$ may make it more or less likely that some portion of the objective function is concave. Indeed, it is easy to find distributions F such that the term $L(p)$ is convex when $(1 - F(pT))$ is concave, and vice versa. Depending on the specific shape of F , therefore, the solution to the social planner's problem can be less or more likely to exhibit random crackdowns.

4 Empirical Application: Speeding Interdiction

In this section we apply our theoretical model of policing to study speeding interdiction in Belgium. Speeding interdiction is an important policy question in its own right, given that traffic accidents are a leading cause of death and disability worldwide.²⁶ In

²⁶In 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

addition, given that one of our goals is measuring deterrence, speeding interdiction has the advantage that the issue of incapacitation does not play a significant role.²⁷ That is, crime rates can be reduced by deterring potential criminals or by incapacitating them, and the models we examined only dealt only with the question of deterrence. Disentangling deterrence from incapacitation is a difficult task.

It will become clear that the variant of the model that best fits the application is the one developed in Section 3.3.

4.1 The environment and the data

The data set comes 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 including newspapers, radio, internet, local stores and restaurants.²⁸

Our data covers the province of Eastern Flanders, which has two major highways and one minor highway, each of them connecting the city of Gent (see the road map in the Appendix B).²⁹ The two major highways are divided into four *sectors*: A14-North, A14-South, A10-East, A10-West. The province has two radar control machines that are 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.³⁰ On any

DALYs for ages 15-44; by comparison, war was only the seventh leading cause for those ages.

²⁷Speeders are not usually incapacitated by prison terms. However, sometimes they do get their licenses revoked for lengthy periods, so incapacitation can play some role in reducing speeding.

²⁸The controls are announced, for example, on the website: <http://www.federalepolitie.be>. The logo that accompanies announced controls is in Appendix C.

²⁹The 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 approximately equidistant and cover around 60-65km within the province of Eastern Flanders. The third highway, R4, is a short stretch of highway connecting Gent with its port to the North.

³⁰Radar 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 (see below) 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 sometimes moved to several

given day, the police either make no announcement or they announce the location of at most one of the machines. An announcement covers exactly one section of one highway. For example, an announcement might reveal that the machine will be placed somewhere on a given section of a highway between the hours of 9am-12. The announcement does not specify the direction of the road on which the machine is placed, nor the exact location. 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.³¹

Our empirical analysis combines data from two sources. The first source is information concerning date, time, and location of the machine, whether the radar control was announced, the number of vehicles passing by the machine (as recorded 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.³² Our second source of information are police records on all vehicles that were issued speeding tickets, which includes information on the speed of the driver (as recorded by the machine), on the type of car, and on the place of residence of the driver.

Police objective and constraints. The police department explicitly states that its goal is to optimally deter speeding, given an upper bound on the number of speeding tickets can be issued each year. In particular, maximizing the revenue from traffic tickets is not part of the police objective function. In fact, the police do not even 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

locations throughout the day.

³¹Only 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.

³²A major speeding violation is defined in the law as travelling at a speed of 10km/h or more over the speed limit. In addition, the margin of error of the radar is $\pm 3\%$, 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 currently, police considers a major violation travelling at a speed of 136km/hr or higher on highways. Currently no tickets are issued systematically for minor violations (over the speed limit, but less than 136km/h) as a result of the radar data collected.

administrative cost of processing the ticket, estimated to be about US \$0.50; the police are given a total budget at the beginning of the year, so they know how many tickets can be issued during the year within this budget. This budget is the total number of ticketed speeders reported in Table 1a.³³ To avoid issuing too many tickets, police do not make use of the radar control machines every day of the year. On days when no announcement is made, police may or may not be using the machines to monitor the roads for speeders. On days with announcements, police make use of at least one machine on the announced road and may also use the other machine on another road unannounced. Based on this description of the police problem, the relevant variant of the model is the one developed in Section 3.3.

Following organizational reforms within the police, in 2002 and 2003 there was a sharp increase in the number of tickets that the police was allowed to issue on highways. The size of the budget constraint more than doubled, from 33,951 in 2000 to 78,136 in 2002. 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. 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 was a side effect of broader organizational changes. We will therefore treat this change in the police budget as *exogenous*, and we will use this source of variation to validate our model.

Monitoring Policy. Monitoring policy is decided several weeks before 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 the circumstances, such as weather. Once a radar control is planned, it is always executed. Table 1a shows the number of vehicles subject to announced and unannounced radar control on the three major highways for the years 2000-2003 as well as the number of drivers issued speeding tickets. Table 1b shows the number of monitoring events on each road. Highway A14 has the most monitoring, followed by A10 and then the shorter highway, R4. Table A1 tabulates the number of announced and unannounced monitoring events by month of year. There is no systematic pattern, except that monitoring is more frequent in the

³³The number of ticketed speeders does not *exactly* match the budget constraint because the police cannot *exactly* predict how many speeders will be ticketed during a monitoring event.

month of December. As seen in table A2, which tabulates monitoring and announcement events by day of the week, monitoring is also more frequent on weekend days in 2000 and 2001, for the reason described above

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. Note, however, that our data pertains only to major highways. The potential problem of route selectivity is much less serious on major highways, because 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, even with the announced controls.³⁴

4.2 Connecting the data to the theory

We now describe how we organize the data in light of the theoretical framework.

Monitoring intensity. Because the police hide the location of the radar machine, a driver’s belief about the probability of being monitored must be constant along an entire section of highway. Thus, p represents the probability that a motorist traveling along a section of the highway is monitored. 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.

Even on announced days, police monitor somewhere in the announced sector, but the probability of a driver being caught speeding in that sector 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.

³⁴This argument might fail, however, if some motorists derive pleasure from exceeding the route-specific speed limit for the sake of it. In that case, taking an alternate, slower route might deliver the pleasure of breaking a speed limit, even though the limit being broken would be lower than maximum speed that is legal on the highway.

The monitoring intensities are obtained as follows. For sectors without announcements, we estimate a logistic model for the probability of the sector being monitored, where we take into account other factors that potentially affect the motorists’ expected probability of being monitored, such as the day of the week, the month of the year, whether the particular road was recently subject to any monitoring (announced or unannounced), whether there has been an announcement on another road and whether it is a holiday. Table A3 presents the coefficients from the logistic regression, where the model is estimated separately by highway and the unit of observation is a day in which there was no announced monitoring of that particular highway. Similarly, we estimate a regression model for the duration of time spent monitoring on days without announcements.³⁵ To obtain p_L , we take the product of the predicted probability of monitoring and the expected time spent monitoring, which is this then multiplied by 0.5 to account for uncertainty about the direction of the road being monitored.

We compute p_H in the same way, except that we replace the conditional probability of monitoring by one (because there is always monitoring on sectors with announcements) and obtain the expected time spent monitoring by dividing the duration of the actual monitoring period by the duration specified in the announcement.³⁶

Crime rate. Our theory requires us to compute $F(p_L T)$ and $F(p_H T)$, the fractions of speeders on a regular day with and without a crackdown (see expression 7), which we can obtain directly from the data. Below, we describe how we also take into account other possible determinants of decisions to speed, such as weather and traffic conditions.

Heterogeneity in trip lengths. Because p is 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.³⁷ A motorist who travelled 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

³⁵The fraction of time spent monitoring is obtained by the expected number of hours divided by 16. There is no monitoring in the data during nighttime.

³⁶For example, if the announced crackdown duration is one hour within a specified three hour time period, then the expected probability of being monitored during the announced interval is 1/6.

³⁷We 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.

sector travelled. This formulation allows us to aggregate trips of different lengths. This is convenient because we do not observe the length of each individual trip.

Heterogeneity in driving population. The theoretical model presented above assumes that individuals are identical in the eyes of police. In reality, the population driving on the road also varies in observable ways. We might worry, for instance, that the slightly higher levels of weekend monitoring before 2002 might be evidence that the police viewed the weekend driving population as different than the weekday drivers. When we raised this concern, however, the police gave us a different explanation. They had anticipated the 2002 increase in resources prior to 2000 and wanted to “smooth the transition” to more frequent monitoring by providing higher frequency monitoring on weekends in 2000 and 2001. In our empirical analysis, we include day of the week as a conditioning variable.

Aside from weekend monitoring, the only other significant predictors of announcements are whether it is a holiday and whether the same road was recently subject to monitoring (see appendix Table A5), which increases the likelihood that there will be an announcement.³⁸ Because our model is static, it is silent on the question of the temporal dynamics of monitoring.

4.3 Validating the model

Police behavior in our speeding application seems to fit well our definition of crackdown: the police randomly and publicly engage in high interdiction phases. Our model, and in particular the variant studied in Section 3.3, rationalizes this behavior as optimal. The model also has some additional predictions, summarized in Proposition 5, that can be used to validate the model.

First, the most direct implication of our theory (Proposition 5 part (a)) is that the

³⁸For this estimation, we can only use the limited set of conditioning variables that is available for all days of the year (not just for the monitoring days). For example, police may monitor less on days with high traffic density; but we do not have a measure of traffic density for days in which there was no monitoring. The police told us that they establish the monitoring schedule approximately one month in advance, so it is unlikely that the schedule depends on factors that could not be anticipated far in advance, such as daily weather fluctuations.

optimal policing scheme partitions the population in at most *two* different groups. 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. 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. We now show that the monitoring intensity is virtually the same in groups (ii) and (iii), which is consistent with the theoretical prediction of the model that there are only two different monitoring intensities.

To illustrate this point, figure 3 plots the histogram of the probabilities with which police monitors drivers (the plot refers to road A10; the histograms for A14 and R4 are similar and available upon request). This is reported below where the first column of 3 figures is for the year 2000, the second and third are for 2001 and 2002. For each year, the top figure displays the distribution of probabilities of monitoring if the police has announced the control. The middle figure 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 has the probability distribution if there is no announcement on that road (A10) or on any other road. For all three years, the monitoring intensity when there are announcements (the top figure for each year) is much different from when there is no announcement on that road (middle and bottom figures). More importantly, for each year, the monitoring intensities do not differ much between the bottom two figures. Thus, groups (ii) and (iii) appear to be monitored with similar intensities, and with sharply lower intensity than group (i). Table A5 suggests that having an announcement on another road has opposite effects on the intensity of unannounced monitoring on the two main roads, and, more importantly, that the effect is quantitatively negligible. Table A4, which reports the average predicted probabilities for all three roads, confirms that probabilities for the unannounced monitoring are very low (they range between 0.001 and 0.030) and are not appreciably different if there is an announcement on another road that day, whereas the announced monitoring probability p_H range between 0.2 and 0.3 for different roads.

Second, the theory (Proposition 5 part (b)) predicts that with an increase in the number of tickets, there will be an increase in the fraction of the population subject

to announced controls (μ_H increases and μ_L decreases) but no change in the monitoring probabilities p_H, p_L . In 2001 and 2002, there were large increases in police resources. The predictions of the model seem to be borne out in the figures. A simple comparison of the histograms in columns 1, 2, and 3 reveals no shift in the monitoring probabilities, even as the number of tickets issued nearly doubled in 2002, and an increase in the number of vehicles subject to monitoring. This pattern can also be seen in Table A3, where the estimated coefficients for the year effects are not significant, as well as in Table A4, which shows the predicted probability of monitoring over the years.

Third, 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 5 (c)). The expected number of successful interdictions per capita is equal to the probability that there is monitoring multiplied by the fraction of motorists who speed. The expected number of tickets issued per capita is systematically larger on announcement days than it is on unannounced days. For example, on highway A10 the expected number of tickets per capita on announcement days is 0.466% whereas on unannounced days it is 0.035%. Likewise on A14 (1.010% and 0.097% respectively) and R4 (0.994% and 0.039%).

Overall, all the predictions of Proposition 5 are supported by the data, which we take as evidence in support of the model as a reasonable approximation to police behavior.

4.4 The deterrence effect of announced controls

A main goal of our empirical analysis is to estimate the deterrence effect of announced controls. Drivers who take to the roads on announcement days are subject to a higher probability of being caught speeding and can therefore be viewed as the group subject to a crackdown. Here, the criteria by which the crackdown group is distinguished are time and day of travelling on the road. To estimate the deterrence effect, we compare the speeding response on p_H days (days with announcements) to the response on p_L days. Our estimation also allows for possible observed heterogeneity that may affect the police's decisions about when and where to monitor and drivers' decisions about whether to speed. The theoretical model of the last section assumed that monitoring intensity is the only observable factor affecting speeding, but in reality there are other relevant

factors, such as weather conditions and traffic density. Additionally, there is likely to be some variation in the population driving on the roads at different times, for example, weekend drivers may differ in utility derived from speeding from weekday drivers. Let Z denote the vector of observables, such as day of the week and month of the year, that are potential determinants of the monitoring probabilities, $P(Z)$.

We estimate a logistic model for individual drivers' decisions to speed, where the speeding decision is assumed to depend on the probability of monitoring ($p(Z)$) and possibly on some other factors, X .³⁹ Table 2 presents the estimated coefficients obtained from the logistic regression for three alternative specifications. In specification (1) the speeding decision is assumed to be a function solely of the predicted probability of monitoring, that was estimated using the method described above. Specification (2) adds the following set of conditioning variables that may 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 and months of year, and year. Specification (3) includes the same conditioning variables, but the speeding decision is now assumed to depend only on whether there is an announcement. This specification estimates the impact of having an announcement, but does not take into account the information on the length of the monitoring interval.

As seen in Table 2, 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 specifications that exclude and include the conditioning variables shows that the estimated deterrence effect is smaller in the specifications that include the covariates. Controlling for covariates has an especially large effect on the estimated coefficient associated with the probability of monitoring on highway R4. As expected, individuals are more likely to speed when there is low traffic density. Speeding also tends to be higher during weekday rush hour times, on holidays, and on Sundays.

Table 3 translates the estimated coefficients from Table 2 into an estimated average impact on the probability of speeding. That is, columns (1) and (2) present estimates

³⁹Some of the elements of X (such as day of week) coincide with elements of Z .

of the average predicted decrease in speeding on each road due to announcements.⁴⁰ As noted above, the estimated deterrence effects are smaller when additional conditioning variables are included in the specification. We focus on the coefficients that include the conditioning variables (reported in columns (2) and (3)), because they likely reflect additional determinants of driver’s decisions to speed that need to be taken into account. On highway A10, the estimated coefficients imply that the fraction of drivers speeding decreases on average by 8.4-19.1% due to the announcements. For highway A14, estimates range from 8.4%-10.2%, and, for highway R4, from 1.6%-3.6%.

We use these estimates to compute the change in the deterrence effect of an increase in resources on speeding interdiction. The key resource constraint the police face is the number of tickets issued. Therefore, we examine how the number of speeders varies as the number of tickets the police are allowed to issue increases, using equation (7) derived from the theory. The effect of an increase of 10, 000 tickets is reported in Table 3 for each of the highways. Depending on the model specification, the reduction in the number of speeders on highway A10 ranges from 4,746 to 12,123 from 3,954 to 10,291 on highway A14, and from 668 to 3,831 on R4.

There is a vast literature documenting the effect of speeding on accidents, injuries and traffic deaths.⁴¹ We next use our estimates of the impact of speeding on fatalities to evaluate whether the police optimally resolve the trade-off between costs and benefits of speeding interdiction. We take an average of the estimated deterrence effects of 10,000 tickets, observed in Table 3, to be about 4,000 speeders. Assuming 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 480,000 travelled kilometers is around $\frac{1.3}{100000000} \cdot 160000 = 0.00208$.⁴² In our data, deterred motorists reduce their speed by about 8 km/h;⁴³ assuming the probability of injury and death increases by 5% per

⁴⁰The table reports the average (over all drivers) decrease in speeding.

⁴¹For the US, the *National Highway and Transportation Safety Authority* (NHTSA) provides estimates of speed-related crashes.

⁴²We impute the expected number of deaths at 1.3 per 100 million km travelled (See DiGuseppi et al (1998).) For comparison, in the US, where the speed limit is lower, 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*).

⁴³Average speeds among speeders are 142 and 144 respectively for A10 and A14. Assuming that

km/hour, in our case, 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 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 accepted estimates of the value of a statistical life, indicating that the use of resources in policing may be efficient.⁴⁴

5 Discussion and Some Extensions

This section compares our theory to some other existing theories of crackdowns. It also extends the model to capture an environment in which crime is not a binary decision but rather a continuous choice and the utility of the citizens is not necessarily linear (to allow for risk aversion). All the previous results carry over to these more general settings.

5.1 Alternative theories of deterrence

The model analyzed in this paper is a model of perfectly rational risk assessment, in which motorists form a Bayesian update of the probability of being monitored based on the available information. In this environment, we have shown that crackdowns may be part of the deterrence policy chosen by the police.

Alternatives to our theory exist that can rationalize crackdowns. These theories are typically based on some form of non-standard (at least from an economist's viewpoint) rationality. For example, an alternative model of the effect of crackdowns is the following.

deterred motorists travel at the maximal non-ticketed speed (135km/h), deterred motorists reduce their speed by 7 - 9km/hr.

⁴⁴For 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, we get an implied value of a statistical life of 2.86 million dollars.

Suppose that absent a crackdown, the probability of monitoring is so small as to be ignored by the driver. Crackdowns raise the speeder’s probability of detection to the point where it is not negligible, and in the process the motorist becomes alert to the detection risk which was previously unforeseen. If this increased alertness persists even when a crackdown is not in force, crackdowns help reduce speeding. 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).

Criminologists have justified crackdowns using an alternative 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). The idea is that crackdowns, because they are fleeting, generate doubt in the mind of the motorist about the interdiction intensity at any particular time, thus boosting the uncertainty component involved in the decision to speed. According to this theory, using crackdowns may magnify the deterrence effect obtained from a given amount of resources.⁴⁵ 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. This implication appears to contrast with the actual policy of liberal information dissemination observed in our application. Note that, consistent with the observed pattern of information dissemination, the optimal policy in our model is to inform motorists about the crackdowns.⁴⁶

⁴⁵Sherman (1990) goes further and argues that the beneficial effect of crackdowns on the motorists’ uncertainty about interdiction actually persists even after a crackdown period is over. He call this effect “residual deterrence,” and argues that residual deterrence is an important component in the effectiveness of the crackdowns.

⁴⁶Our model predicts the effectiveness of the crackdown is directly proportional to the fraction of motorists who are aware of it. If no motorist were aware of the crackdown, all motorists would expect the *average* amount of interdiction and there would be no effect of crackdowns on speeding.

5.2 Crackdowns persist if citizens' utility function is nonlinear

The model developed in this paper assumes that the utility function was linear in the benefit, x , from committing a crime. We next show that the linearity assumption can be relaxed. Suppose citizens have a utility function u that is increasing in x . Then they will commit a crime iff

$$(1 - p)u(x) + pu(x - T) > u(0).$$

Consider the set of values of x such that the inequality is satisfied, and denote by $H(p)$ the measure of this set. The function $H(p)$ represents the crime rate. Note that since u is increasing, the left hand side is increasing in x and, also, $u(x) > u(x - T)$ whereby the left hand side is decreasing in p . Therefore, the set of values of x such that the inequality is satisfied decreases as p increases. This means that $H(p)$ is decreasing in p . The analysis of Sections 2 and 3 can then be carried out replacing F with H .

5.3 Crackdowns persist if crime decision is continuous

Suppose that instead of a binary problem (committing a crime or not), each citizen solves a more complicated problem involving not only whether to commit a crime, but also the degree to which to commit it. For example, a motorist may choose whether to speed and how much to speed. Suppose that the penalty for driving at s miles per hour above the speed limit is a nondecreasing function $T(s)$ (which could be equal to zero below the speed limit) and that the agent's utility from exceeding the speed limit by s is an increasing function $x(s)$. We allow different individuals to have different functions $x(s)$. Given a certain level of interdiction p , an agent with a given function $x(\cdot)$ solves

$$\max_s x(s) - pT(s).$$

Denote with $s^*(p)$ the maximizer of this problem. Denote by $\tilde{F}(s|p)$ the fraction of individuals who choose to travel at or below speed s for given p . The quantity $\tilde{F}(s|p)$ will depend on the distribution of the functions $x(\cdot)$ that is present in the population. It is easy to see, however, that the optimal speed $s^*(p)$ is decreasing (or at least not increasing) in p : $s'(p) = T' / (x'' - pT'')$ is negative by the concavity of $x - pT$ at the maximum. This means that any motorist, regardless of his or her $x(\cdot)$, will decrease

his or her optimal speed as the probability of being monitored increases. The function $\tilde{F}(s|p)$ is therefore increasing in p .

If police cared not only about the fraction of people who exceed the speed limit, but also about their speeding levels, the police's objective function would be represented by the function

$$D(p) \equiv \int K(s) d\tilde{F}(s|p),$$

where $K(s)$ is some nondecreasing function. The function $K(s)$ represents the disutility that the police receives from having one motorist travel at speed s . Because $\tilde{F}(s|p)$ is increasing in p , the function $D(p)$ is decreasing in p . Now, rewrite problem (2) replacing $1 - F(p)$ with $D(p)$. This yields a mathematical formulation of the problem in which motorists can choose the amount of speeding and police care not only about the fraction of speeders but also about their speed. From a formal viewpoint this new formulation is similar to the original problem. Therefore, all the qualitative features of the solution to the original problem carry over, including the optimality of crackdowns when the function $D(p)$ exhibits non-convexities.

5.4 Size of the penalty

In considering the agency relationship between the principal and the police we have taken the size of the penalty (T in the formal model) to be exogenous. In actuality, the size of the penalty is an object of choice, sometimes on the part of the principal.⁴⁷ It should be pointed out that our theory remains meaningful irrespective of how or by whom the penalty is chosen, as long as deterrence is not perfect. Thus, the theory does not require us to explain how the penalty is determined.

One might ask, however, why would deterrence not be perfect. Or, a more meaningful question in our applied setting, why would penalties not always be set at their maximal value so that everyone is deterred with a minimum deterrence cost? In our application, for example, we might ask why speeders are not sent to jail? This point was raised by Becker (1968) who noted that, when interdiction is costly, increasing the size of the fine allows the same level of deterrence to be implemented while saving on interdiction costs.

⁴⁷In our speeding application the fines are chosen by the Belgian parliament.

Regardless of the objective to be implemented, then, penalties should always be set at their maximal value.

Why are observed penalties not always maximal? The literature has identified several arguments; here we mention those that are more directly applicable to our analysis, and refer to Polinsky and Shavell (2000) for an excellent overview of the others. First is the need to generate marginal deterrence: if all violations were punished with the same (maximal) intensity, there would be no incentive to choose a lesser over a larger (and more harmful) violation (see Stigler 1970, and more recently Shavell, 1992). If enforcement cannot be tailored perfectly to the magnitude of the violation, lower penalties will need to be applied to less serious crimes. A second reason is that, when a harsh penalty is imposed, those who enforce the law must be monitored lest they abuse their position, and opportunities for judicial appeal must be provided to redress enforcement mistakes.⁴⁸ Taking into account these ancillary (but very important) costs may explain why penalties are seldom set at their maximal possible value. A third consideration is risk aversion (see Polinsky and Shavell 1979): whenever it is socially optimal for some individuals to violate the law (for example, speeding may be socially optimal in some circumstances), increasing the penalty and decreasing the risk of apprehension increases the risk faced by these “optimal violators,” and thus reduces social welfare. This consideration places limits on the size of the penalty chosen by the social planner. While we do not explicitly model these considerations, they could easily be added to the model without affecting the fundamental force that generates crackdowns.

6 Conclusions

This paper presented a rational theory of crackdowns in police interdiction. Our analysis showed 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

⁴⁸So, for example, the degree of judicial protection is much larger (and the system therefore much more expensive) in death penalty cases than for traffic violations.

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.

We compared several variants of the policing model according to the likelihood that crackdowns arise as an optimal policy, and showed that crackdowns are prevalent under a variety of incentive configurations.

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 suggests 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.

We believe our theory lends itself to investigating other situations where disparate treatment of identical groups may be an efficient way of allocating resources. Disparate treatment of an arbitrarily chosen subgroup of the population can be applied to other contexts, such as the auditing of firms for tax purposes or the security screening of passengers at airports. For example, the crackdown theory would imply that instead of auditing firms which are observably similar in the same way, it can be optimal to divide firms arbitrarily into groups, one of which is audited more intensely (e.g. subjected to more frequent inspections) and the other less intensely.⁴⁹ In the case of airport security, this may lead to publicly announcing that specific flights will be screened more intensely than others, rather than screening passengers on all flights with the same average intensity.

⁴⁹This assumes that the goal of auditors is to minimize tax evasion subject to a constraint on the amount of auditing resources, see Chander and Wilde (1998).

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

	2000	2001	2002	2003 (first half of year)
Announced	266,240	394,540	1,746,340	1,777,977
unannounced	406,941	319,650	526,422	1,139,428
Total	673,181	714,190	2,272,762	2,917,405
number of ticketed speeders	33,951	45,264	78,136	48,795

Table 1b
Number of Announced/Total Monitoring Observations
by Highway and Year

	2000	2001	2002	2003 (first half of year)
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	0/2
Total	66/218	75/181	337/463	275/376

Table 2
 Estimated coefficients from logistic regression of probability of speeding
 (Standard errors in parentheses)

Variable	(1)	(2)	(3)
Intercept	-3.38 (0.52)	-3.43 (0.01)	-3.26 (0.02)
indicator for A14	0.52 (0.08)	0.23 (0.008)	0.04 (0.007)
indicator for R4	0.30 (0.02)	-0.21 (0.02)	-0.42 (0.02)
announcement on highway A10	-0.36 (0.008)
announcement on highway A14	-0.12 (0.006)
announcement on highway R4	-0.04 (0.05)
probability of monitoring – A10	-1.17 (0.04)	-0.47 (0.04)	...
probability of monitoring – A14	-0.97 (0.02)	-0.40 (0.02)	...
probability of monitoring – R4	-0.38 (0.15)	-0.07 (0.15)	...
traffic density 3	...	-0.27 (0.06)	-0.09 (0.05)
traffic density 4	...	0.30 (0.03)	0.48 (0.03)
traffic density 5	...	0.24 (0.01)	0.29 (0.01)
morning rush hour*weekday	...	-0.47 (0.02)	-0.47 (0.02)
evening rush hour*weekday	...	0.02 (0.006)	-0.007 (0.006)
holiday	...	0.51 (0.007)	0.49 (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
p-value from joint test that all coefficients equal 0	<0.0001	<0.0001	<0.0001

Table 3
Decrease in speeding due to crackdown and the deterrence effect
of increasing the number of tickets

		Model Specification		
Highway		(1)	(2)	(3)
A10	Predicted % speeding above threshold on unannounced days	3.3	3.0	3.2
$p_H=0.2007$ $p_L=0.0084$	Decrease in speeding on announcement days implied by estimated coefficients	0.20	0.25	0.60
	Decrease as a % of speeding on unannounced days	6.1%	8.4%	19.1%
	Slope of 1-F	-1.04%	-1.30%	-3.12%
	Effect of additional 10,000 tickets	-3,364	-4,746	-12,123
A14	Average % speeding on unannounced days	5.4	5.0	5.0
$p_H=0.2441$ $p_L=0.0167$	Decrease in speeding on announcement days implied by estimated coefficients	1.01	0.41	0.50
	Decrease as a % of speeding on unannounced days	18.9%	8.4%	10.2%
	slope of 1-F	-4.44%	-1.80%	-2.20%
	Effect of additional 10,000 tickets	-10,291	-3,954	-4,926
R4	Average % speeding on unannounced days	4.4	4.3	4.3
$p_H=0.2576$ $p_L=0.0091$	Decrease in speeding on announcement days implied by estimated coefficients	0.38	0.07	0.16
	Decrease as a % of speeding on unannounced days	8.6%	1.6%	3.6%
	slope of 1-F	-1.53%	-0.28%	-0.64%
	Effect of additional 10,000 tickets	-3,831	-668	-1,563

* Total number of vehicles on each highway for the entire year (Driver Pop) was estimated from the data collecting on monitoring events. The estimation was performed by regressing the number of vehicles per hour on conditioning variables (quarter of the year, day of the week, time of day (morning, afternoon, evening) holiday indicator, and an indicator for holiday*weekend). The fitted regression was used to impute numbers of vehicles for days/times when there was no monitoring.

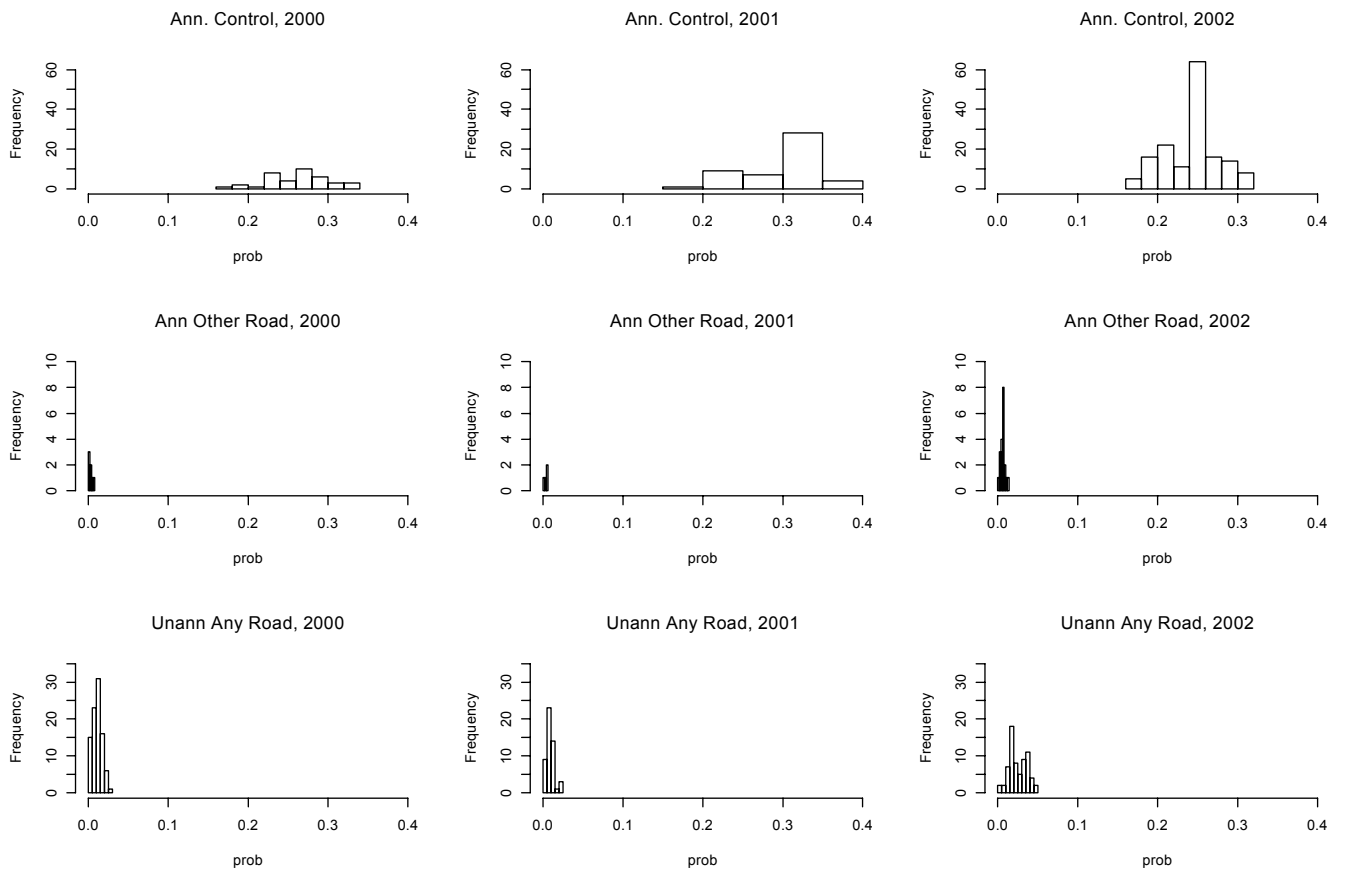


Figure 3: Probability of being monitored, highway A10.

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*

$$\begin{aligned} \max_{\mu} \quad & \int_0^S f(p) \mu(p) dp \\ \text{s.t.} \quad & \int_0^S g(p) \mu(p) dp \leq C. \end{aligned} \tag{8}$$

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

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

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

$$\begin{aligned} f(\varphi(p)) - \lambda g(\varphi(p)) &= M \quad \text{for } \varphi(p) \in \mathcal{A} \\ f(\varphi(p)) - \lambda g(\varphi(p)) &< M \quad \text{for } \varphi(p) \notin \mathcal{A}, \end{aligned}$$

or, with the obvious meaning of symbols,

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

Note that the set $\varphi^{-1}(\mathcal{A})$ 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 $\frac{1}{\lambda}(\frac{p}{S} - M)$ is tangent to the function $g(f^{-1}(p/S))$ at more than two points and never exceeds it. This means that there is a tangent hyperplanes to the set

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

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 $d_g(p)$ which is equal to 0 for every p unless $p \in (a_h, b_h)$ for some $h \in \mathcal{H}$, in which case $d_g(p)$ assumes values strictly between zero and 1. Let $\tilde{f}_\varepsilon(p) \equiv [1 + \varepsilon \cdot d_g(p)] \cdot f(p)$. For any $\varepsilon > 0$, the set $\tilde{Y}_\varepsilon = \{(p, y) : y \leq \tilde{f}_\varepsilon(p)\}$ 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}_ε at more than two points. Since the function $\tilde{f}_\varepsilon(p)$ can be made arbitrarily close to $f(p)$ 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 $\tilde{C} = C + \varepsilon$ with ε a small positive number. The solution to the programming problem is a saddle point $(\tilde{\mu}, \tilde{\lambda})$ for the Lagrangean. We now proceed to construct this saddle point. We start by keeping the Lagrange multiplier unchanged, $\tilde{\lambda} = \lambda^*$. Because of this choice, the $\tilde{\mu}$ 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 $\tilde{\lambda} = \lambda > 0$ to minimize the Lagrangean, the Lagrangean must be constant with respect to λ , which is equivalent to

$$g(p_m) \tilde{\mu}(p_m) + g(p_M) \tilde{\mu}(p_M) = \tilde{C} \quad (9)$$

Since $g(p_M) > C > g(p_m)$, for ε sufficiently small also $g(p_M) > \tilde{C} > g(p_m)$. Therefore, it is possible to choose $\tilde{\mu}(p_m)$ and $\tilde{\mu}(p_M) = 1 - \tilde{\mu}(p_m)$ so that equation (9) is satisfied. Choosing $\tilde{\mu}$ 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 (8) 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. ■

Appendix B: Additional Tables

Table A1
Number of Monitoring and Announcement
Observations by Month of Year on all Highways

	2000		2001		2002		2003 (first half of year)	
	Ann	Unann	Ann	Unann	Ann	Unann	Ann	Unann
January	3	7	10	12	15	17	41	3
February	10	12	7	9	17	9	45	9
March	5	6	5	3	19	26	32	24
April	3	18	6	4	24	21	36	24
May	3	10	6	8	29	11	64	21
June	4	14	4	11	27	8	57	20
July	7	13	6	10	31	6	*	*
August	8	7	7	8	32	1	*	*
September	4	9	6	7	29	6	*	*
October	3	18	5	5	41	12	*	*
November	3	17	6	13	36	6	*	*
December	13	21	7	16	37	3	*	*
Total	66	152	75	106	337	126	275	101

Table A2
Number of Monitoring and Announcement
Observations by Day of Week on all Highways

	2000		2001		2002		2003 (first half of year)	
	Ann	Unann	Ann	Unann	Ann	Unann	Ann	Unann
Saturday	8	38	23	27	40	19	42	8
Sunday	17	29	10	34	64	31	31	24
Monday	8	17	6	11	24	12	50	11
Tuesday	7	15	12	8	68	17	40	17
Wednesday	10	16	13	9	58	14	41	20
Thursday	6	19	7	9	39	22	47	11
Friday	10	18	4	8	44	11	42	18
Total	66	152	75	106	337	126	293	109

Table A3
 Estimated Logistic Model for the Probability of Monitoring
 when no announcement was made by Year and by Road

	Highway		
	A10	A14	R4
Intercept	-2.50 (0.36)	-2.48 (0.36)	-4.36 (0.58)
quarter 1	0.37 (0.25)	-0.11 (0.44)	0.46 (0.45)
quarter 2	-0.19 (0.28)	0.44 (0.20)	0.55 (0.43)
quarter 3	-1.91 (0.49)	-0.48 (0.22)	0.10 (0.44)
announced last week	0.15 (0.35)	-0.12 (0.20)	0.04 (0.49)
announced yesterday	-0.22 (0.41)	-0.27 (0.23)	...
monitored last week	0.79 (0.40)	2.04 (0.37)	2.10 (0.34)
monitored yesterday	0.37 (0.36)	0.46 (0.19)	-1.21 (0.64)
some announcement same day on any road	0.64 (0.23)	-0.96 (0.21)	-1.36 (0.49)
year 2001	-0.16 (0.29)	-0.67 (0.20)	0.27 (0.34)
year 2002	-0.08 (0.32)	0.30 (0.20)	-0.27 (0.54)
year 2003	-0.01 (0.31)	0.51 (0.25)	...

* All specifications also include fixed effects for day of week. The variable "holiday" was not included in the above specifications because of too few observations.

Table A4
 Average Predicted Probability of Monitoring

Year	highway	no-announcement	no-announcement this sector, announced other sector	announcement
2000	A10	0.004	0.009	0.27
	A14	0.011	0.003	0.26
	R4	0.009	0.002	0.29
2001	A10	0.004	0.007	0.30
	A14	0.011	0.003	0.30
	R4	0.014	*	0.35
2002	A10	0.008	0.010	0.20
	A14	0.027	0.007	0.25
	R4	0.001	0.001	*
2003	A10	0.011	0.018	0.19
	A14	0.030	0.020	0.23
	R4	*	*	*

* Too few observations in the cell.

Table A5
Estimated Logistic Model for the Probability of Announcement by Highway

	Highway		
	A10	A14	R4
intercept	-3.76 (0.36)	-4.94 (0.40)	-3.83 (0.82)
quarter 1	0.08 (0.21)	-0.004 (0.20)	-0.89 (1.18)
quarter 2	0.22 (0.21)	0.02 (0.20)	-0.03 (0.83)
quarter 3	-0.23 (0.22)	-0.01 (0.20)	0.10 (0.79)
holiday	2.23 (0.47)	0.98 (0.27)	...
announced last week	0.34 (0.36)	-0.18 (0.21)	0.13 (0.95)
announced yesterday	-0.52 (0.29)	0.19 (0.21)	...
monitored last week	1.38 (0.41)	2.65 (0.38)	0.88 (0.75)
monitored yesterday	0.25 (0.27)	-0.31 (0.19)	...
year 2001	0.06 (0.03)	0.42 (0.23)	-1.68 (1.08)
year 2002	2.13 (0.30)	1.55 (0.22)	-12.13 (214)
year 2003	2.12 (0.30)	2.16 (0.24)	...
p-value from test of joint significance of all covariates, except year indicators	<0.0001	<0.0001	0.7923

*All specifications include fixed effects for days of week. Some days of week indicators are significant for A10 and A14 in 2000 and for A14 in 2001.

**There is only one announcement day during 2001 on R4.

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