Profiling for crime reduction under severely uncertain elasticities [*]

Lior Davidovitch [†]   Yakov Ben-Haim [‡]

June 5, 2008

Abstract

The economic theory of crime views criminals as rational decision makers, implying elastic response to law enforcement. Group-dependent elasticities can be exploited for efficient allocation of enforcement resources. However, profiling can augment both number of arrests and total crime, since non-profiled groups will increase their criminality. Elasticities are highly uncertain, so prediction is difficult and uncertainty must be accounted for in designing a profiling strategy. We use info-gap theory for satisficing (not minimizing) total crime rate. Using an empirical example, based on running red lights, we demonstrate the trade-off between robustness to uncertainty and total crime rate.

JEL keywords: D81, K42

1   Introduction

The modern economic view of crime is traditionally traced back to Becker (1968). In his seminal paper, Becker notes that potential offenders come from various backgrounds, and therefore have different responses to the probability of conviction and to the expected punishment. He suggested that parameters such as premeditation, sanity and age may be used as proxies for the offenders' elasticities of response to punishment. The combination of elasticities for different groups, and the ability to statistically predict the elasticities via proxies, is the basis for *statistical discrimination* (Arrow 1973), or profiling[1].

[†]Faculty of Mechanical Engineering, Technion – Israel Institute of Technology, Haifa 32000 Israel. Email: liordav@technion.ac.il.

[‡]Yitzhak Moda'i Chair in Technology and Economics, Faculty of Mechanical Engineering, Technion – Israel Institute of Technology, Haifa 32000 Israel. Email: yakov@technion.ac.il. Corresponding author: Tel. 972-4-8293262, Fax 972-4-8295711.

[1]Of course, the use of actuarial tools in the context of criminology predates Becker. It is related to Burgess (1928), but he did not refer to profiling as a result of an economic model, or suggest an economic model to utilize the differences between the groups.

1

Although profiling has been shown to be potentially beneficial (as a tool for minimizing crime rate, or maximizing some abstract social benefit), it has been argued that the inequality that is the essence of profiling is, in fact, unjust. After Lamberth (1994) showed that there is a discriminatory policy, either official or de facto, against African American drivers in the context of drug interdiction, much research discussed to what extent profiling can be justified as an economic result, rather than a racial bias (Knowles et al. 2001; Borooah 2001; Hernández and Knowles 2004), and what is the trade-off between equality and efficiency of law enforcement (Farmer and Terrell 2001; Persico 2002; Blumkin and Margalioth 2005)[2].

Setting aside ethical considerations, profiling has been criticized for being *inefficient*. Harcourt (2007[3]) demonstrates that, under a set budget, targeting groups with higher crime rates may cause the total crime rate to *increase*. This is because shifting enforcement resources to a minority group will cause the remaining majority to increase its participation in crime. The net effect can be an increase in both total arrests as well as total crime. Bearing that in mind, a policy maker who wishes to use profiling as means for reducing the total crime rate must take into account not only the current crime rates of the different groups, but also the groups' responsivenesses (or elasticities) to policing.

However, it is extremely difficult to estimate the responsiveness of crime to policing, even for the general population. In fact, many researchers find that the correlation between policing efforts and crime is either non-existent or positive (more policing means more crime)[4]. Although it has been argued that the main reason for this is simultaneity problems[5], it is still non-trivial to estimate the responsiveness to policing[6]. Over the years, researchers have tried to optimize the social-welfare (or some proxy of the social-welfare) using the responsivenesses of the different groups within the population[7] (Becker 1968; Benson and Bowmaker 2005), the utility to the offender from the illegal act (Malik 1990; Polinsky and Shavell 2000), the dis-utility from disrepute due to conviction (Polinsky and Shavell 2000; Pradiptyo 2007), the knowledge of the criminals regarding their probability of conviction and expected punishment (Polinsky and Shavell 2000), and so on. The huge uncertainties of the models involved were often overlooked. Bar-Ilan and Sacerdote (2004) point out that estimating the elasticity to fine increases for running red lights may be compounded by "other costs to receiving a ticket, including increased insurance premiums, time costs, and feelings of guilt." Such effects may be difficult to quantify and can vastly change the estimated elasticities.

---

[2]Heaton (2006) shows the decreased efficiency of policing (resulted in an increase in crime rates) due to "anti-profiling" policy implemented in New-Jersey.

[3]Similar argument is presented in Harcourt (2006).

[4]See Levitt (1997) for examples of empirical research. Tsebelis (1990) uses game theoretic reasoning to prove that the crime rate is independent of the severity of the punishment, though it might be influenced by the probability of detection. See also Ehrlich and Liu (1999), as an example for the debate in the question of deterrence. On the other hand, Levitt (1998) reports a strong negative correlation between arrest rates and reported crime rates.

[5]Crime and law enforcement affect each other simultaneously, since high crime rates lead to further investment in law enforcement.

[6]Levitt (1997) uses the assumption that the growth in police size during election years (be it mayoral or gubernatorial) is not related to the level of crime, to compare the level of crime in election years to non-election years. Klick and Tabarrok (2005) utilize the increased police presence in periods of high-alert to show that the crime rate is mostly reduced in the area of the National Mall in Washington DC, which is supposed to have a higher presence of police in periods of high-alert (this district hosts the White House, Congress, Supreme Court, and so forth).

[7]Or the supply of offences.

In this paper we suggest the use of info-gap theory (Ben-Haim 2006) in order to *satisfice* the total crime rate, rather than to *optimize it*. By "satisficing" we mean keeping the value of a loss function (like total crime rate) below an acceptable level. Satisficing is to be distinguished from optimizing which entails minimizing the loss. The motivation for satisficing (rather than optimizing) derives from the great uncertainty associated with estimates of the responsiveness to policing. We will demonstrate the irrevocable trade-off between robustness to this uncertainty on the one hand, and reduction of the total crime rate on the other. An allocation which attempts to minimize total crime is an allocation with zero robustness to uncertainty in the responsiveness function. Under a fixed budget, elastic response to profiling can result in an increase in total crime. Hence, knowledge of the elasticity is critical. When this knowledge is highly uncertain, it is necessary to choose an allocation which is robust to this uncertainty while at the same time aiming at adequate reduction in total crime. Allocation must aim to reliably achieve acceptable reduction — rather than minimization — of the total crime rate. The quantitative analysis of this trade-off underlies the choice of an allocation.

We will demonstrate the profiling of two groups with uncertain responsiveness functions, and show how to choose an allocation of police resource which will be robust to errors in the estimation of responsiveness functions. We will give a numerical example, based on research that estimated the elasticities of different groups to policing in the context of driving through red lights (Bar-Ilan and Sacerdote 2004).

The paper is organized as follows. Section 2 briefly describes how info-gap theory is used to robustly satisfice a requirement. Section 3 exemplifies the use of info-gap theory in the case of profiling traffic violators. Section 4 discusses the similarity and difference between robust-satisficing and the min-max strategy. A concluding discussion appears in Section 5. Mathematical definitions and derivations appear in appendices.

2   Info-Gap Theory: An Intuitive Discussion

In this section we present an intuitive description of info-gap models of uncertainty, and how info-gap models can be used for deriving robust decisions. A mathematical description is available in Appendix A.

Decision making may be viewed as choosing a decision q from a set $\mathcal{Q}$ of feasible decisions. The outcome of the decision is expressed as a loss, $L(q, u)$, where u is the value of parameters or functions which are unknown or uncertain to the decision maker when the decision was made. u may be, for instance, the parameters of a model, or a functional relationship between variables, or a probability distribution of random variables, or sets of such entities. In this paper u is the uncertain responsiveness to policing. We have a best estimate of u, denoted $\widetilde{u}$, but our uncertainty about u is non-probabilistic. That is, we do not know a probability distribution which describes the uncertainty of u. In many cases, the uncertainty about u is unbounded: we cannot identify a worst case. Our analysis will be based on info-gap decision theory (Ben-Haim 2006).

Info-gap models are used to quantify non-probabilistic Knightian uncertainty (Ben-Haim 2006). An info-gap model is an unbounded family of nested sets. At any level of uncertainty, a set contains

possible realizations of u. As the horizon of uncertainty gets larger, the sets become more inclusive. The info-gap model expresses the decision maker's beliefs about uncertain variation of u around $\widetilde{u}$.

Info-gap models of uncertainty obeys two axioms:

1. **Contraction:** $\widetilde{u}$ is the only possibility when there is no uncertainty.

2. **Nesting:** the range of possible realizations increases as the level of uncertainty increases.

Suppose the decision maker wishes to reduce the loss, and has some notion of a critical loss L_c, whose exceedence cannot be tolerated. The *robustness* of a decision q, denoted $\widehat{\alpha}(q, L_c)$, is the greatest level of uncertainty which still guarantees a loss no greater than L_c. *Robust-satisficing* decision making maximizes the robustness and keeps the loss less than the value L_c, without specifying a limit on the level of uncertainty. That is, given a critical loss, the decision maker will choose the decision $\widehat{q}$ with greatest robustness to uncertainty. Under non-probabilistic Knightian uncertainty, this is an attempt to maximize the confidence in achieving no more than an acceptable loss.

It can readily be shown that there is an inherent trade-off between robustness and performance. Since robustness is the *immunity to failure*, the robustness decreases as the performance requirement L_c becomes more demanding. That is, $\widehat{\alpha}(q, L_c)$ gets smaller as L_c gets smaller. Another immediate result is that the estimated *optimal* result — the minimal loss under our best estimate $\widetilde{u}$ — has zero robustness, meaning that a slight deviation from our estimation $\widetilde{u}$ may result in exceeding L_c.

There are certain similarities between robust-satisficing and minimax. Section 4 presents the main similarities and differences between the two methods.

3   Case Study: Running Red Lights

Not often do we come across data that may be used to infer the responsiveness to policing of different groups within the population. However, we do have such data for driving through a red light. In the United States, roughly 2,000 deaths resulted in 1998 from drivers running red lights (Bar-Ilan and Sacerdote 2004). It should be noted that most of the enforcement of running red lights is done automatically, using cameras. This means that profiling, in its "natural" meaning of assigning different probabilities of detection to different groups within the population, is not easily implemented, but could be done by varying the density of detectors in different regions.

Drug interdiction on highways is much more relevant for profiling. Since a car search is initiated as a result of suspicion by a police officer, it is quite reasonable to assume that the suspicion is somewhat correlated to the group to which the driver of the car belongs, be it an ethnic group, a socioeconomic group, or a cultural group. Indeed, much research has examined the correlation between the ethnicity of the driver and the probability of his car being searched (Lamberth 1994; Knowles et al. 2001; Borooah 2001; among many others).

Nonetheless, running red lights is one of the rare cases where information regarding responsivenesses of different groups within the population can be found, while similar information regarding drug

interdiction is scarce. Therefore, in order to demonstrate the practical use of info-gap theory, we assume that running red lights could be profiled in a similar fashion as drug interdiction. Namely, that policing resources could be allocated arbitrarily between different groups, thus affecting the probability of catching a driver running through a red light. We will then use the data gathered on running red lights to demonstrate the robust-satisficing methodology described in section 2.

3.1 Responsiveness to Policing

Bar-Ilan and Sacerdote (2004) study running through red lights, and use incidents of change in the probability of detection (when traffic cameras are added) and changes in the punishment in the case of detection (when the fine for driving through red lights is increased) to show that the responsiveness to these two factors is quite similar, which suggests some degree of risk-neutrality of the drivers. In particular, Bar-Ilan and Sacerdote use very detailed data collected in Israel to compare the responsiveness and crime rates of different groups within the population, after Israel raised the fine for driving through red lights from 400 shekels ($122) to 1,000 shekels ($305) in December of 1996. As assumed by Harcourt (2007), the responsivenesses (or elasticities) of the different groups are not necessarily similar: young drivers have a higher rate of violations, but also have an elasticity which is significantly higher than the general population; drivers convicted of property crimes have a higher rate of red-light violations, but an elasticity which is similar to that of the general population; non-jewish drivers have a much lower elasticity than the general population.

In order to calculate the entire curve of responsiveness from limited data, we must assume the general shape of the curve. We will use an info-gap model to represent uncertainty in the shape of this curve. Our best guess is that the responsiveness curve of the ith group has the following form:

$$\widetilde{C}_i = \exp\left(-\gamma_i \frac{b_i}{\pi_i} - \delta_i\right) \tag{1}$$

$\widetilde{C}_i$ is the average crime rate of the ith group: the number of red light incidents per person in the group per time period of 14 quarters (the length of the period examined by Bar-Ilan and Sacerdote). γ and δ_i are parameters which characterize the responsiveness of the ith group. π_i is the fraction of group i within the general population. b_i is the fraction of the budget allocated to police group i. Thus, π_i and b_i are both between zero and one. This is the basis of the profiling: by setting $b_i > \pi_i$, we target group i (the fraction of policing resources allocated to group i is greater than its fraction within the general population). Appendix B gives the intuition for this model, and calculates its parameters, based on the research of Bar-Ilan and Sacerdote (2004).

3.2 Satisficing the Crime Rate

A good group to target is a group which constitutes a considerable fraction of the population of drivers, has a high value of γ_i (high elasticity), and of course, can be easily recognized by a police officer. The group of drivers between the ages 17 and 30 meets all the above criteria. Therefore, we shall concentrate our efforts on profiling this group, where the goal is to reliably reduce the total crime rate.

Following is an intuitive review of the process of robustly satisficing the crime rate. Appendix C gives the mathematical definitions and results.

3.2.1 Info-Gap Model of Uncertainty

The exponential model representing the responsivenesses of the different groups to policing, eq. (1), is only a rough estimate; the shape of the curve may be different. The crime rate and responsiveness has been measured by Bar-Ilan and Sacerdote for a specific allocation, which we shall denote b^0. We may be fairly confident of the crime rate for b^0. However, it is reasonable to suppose that the uncertainty in the responsiveness function grows as the difference between the actual allocation, b, and the reference allocation, b^0, increases.

Since we will be examining *reallocation* of fixed total policing resources, we shall describe an allocation using the fraction of resources allocated to each group. That is, b_i is the fraction, between zero and one, of the resources allocated to group i, rather than the absolute amount of resources.

Let $\widetilde{C}$ be a vector of responsiveness functions, representing our best estimate of the responsiveness functions of the different groups. $\widetilde{C}_i$ will be based on the exponential model, eq. (1). We will refer to $\widetilde{C}$ as the *nominal model*, and represent the uncertainty surrounding the actual responsiveness functions using an info-gap model. Let C denote the vector of actual responsiveness functions, which may differ in functional form from the nominal vector $\widetilde{C}$. Our info-gap model, which is defined in Appendix C, assumes that the maximal error in our estimation of the responsiveness functions increases as the allocation deviates from b^0. Figure 1 illustrates an uncertainty envelope for this info-gap model. At the horizon of uncertainty shown in Figure 1, all functions $C_i(b_i)$ within this envelope are allowed. The *shape* of the envelope (dashed curves) is specified, but the true *magnitude* of deviation (distance between dashed and solid) is unknown. The info-gap model is an unbounded family of such envelopes.

3.2.2 Robustness

Let b_y denote the allocation of surveillance resources to the "young" population of 17 to 30 years old, and let $b_{\bar{y}}$ denote the allocation to the complementary group. Given some critical crime rate L_c, we can calculate the robustness of any given allocation b. Since $b_y + b_{\bar{y}} = 1$, we can represent an allocation through b_y.

In choosing an allocation, we wish to know how wrong the estimated response functions could be, and the allocation would still result in acceptable total crime rate. The robustness, $\widehat{\alpha}(b_y, L_c)$, of an allocation b_y, is the greatest horizon of uncertainty up to which all realizations of the responsiveness functions, C_i, result in total crime rate not exceeding the critical value L_c. Figure 2 illustrates the robustness curves for four different allocations of the policing resources: the *current* allocation ($b_y = 0.145$, which is equal to the fraction of "young" in the population, and is the allocation b_y^0 measured by Bar-Ilan and Sacerdote [2004]), the *nominal optimal* allocation (in the sense that it yields the minimal crime rate under the nominal model, $b_y = 0.268$), and two other allocations ($b_y = 0.2$ and $b_y = 0.12$). As expected, all robustness curves are monotonic: robustness increases as the critical crime rate
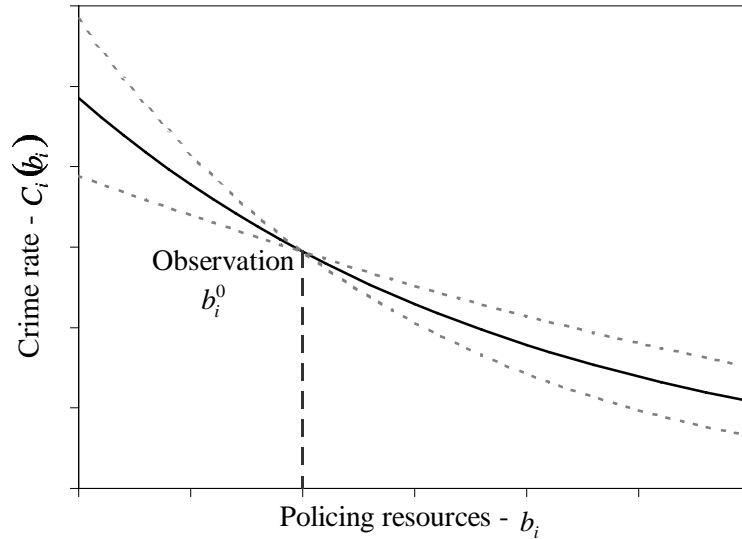
Figure 1: An envelope of possible responsiveness functions within the info-gap model when the maximal fractional error is proportional to the distance of the allocation from some observed allocation. This is the info-gap model defined by eq. (C1).
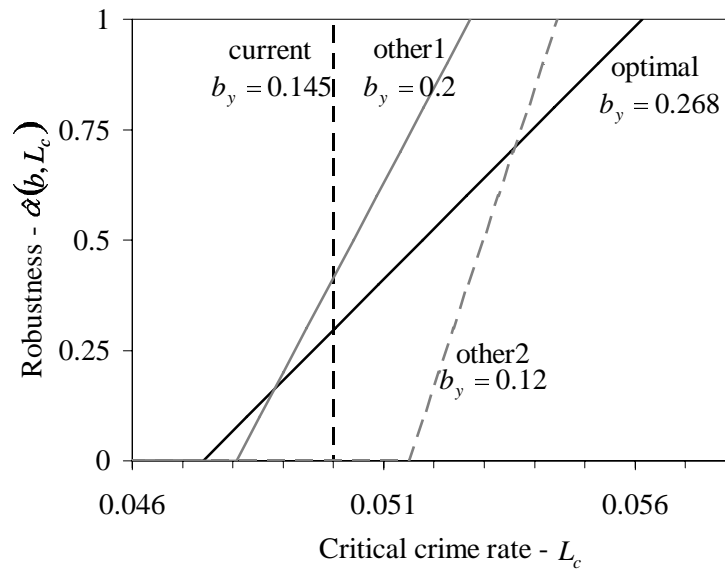

Figure 2: Robustness curves of four allocations: the current allocation ($b_y = 0.145$), the optimal allocation ($b_y = 0.268$), and some other allocations ($b_y = 0.2$ and $b_y = 0.12$).

increases (a *weaker* requirement is *more robustly* achieved). Also, each curve crosses the horizontal axis at the crime rate yielded by the corresponding allocation under the nominal model.

7

The definition of robustness implies a vertical robustness curve for the reference allocation, $\hat{b}$. (We are not concerned with the statistical uncertainty of the observation. Rather, we focus on the uncertainty in the shape of the responsiveness functions as the allocation changes from the current value.) We can understand this as follows: we are certain of the crime rate under the current (observed) allocation, b^0. Therefore, the robustness of that allocation is zero for crime rates less than the current crime rate, and infinite for crime rates higher than that crime rate. The infinite robustness of the current allocation appears as a vertical curve at $L_c = 0.050$, the current crime rate. This means that the current allocation is the most robust (at the time of measurement) if the critical crime rate is at least the current crime rate.

The nominal optimal allocation, $b_y = 0.268$, yields the lowest total crime rate under the nominal model. This makes it more robust than any other allocation around the nominal optimal crime rate. However, the optimal crime rate is not a good choice for the critical value, since the robustness for the nominal optimal crime rate is zero. This means that the slightest deviation from the assumptions of the models may cause the crime rate to exceed the nominal optimal value.

Note that the nominal optimal robustness curve is crossed by other robustness curves. The crossing of the robustness curve of the nominal optimal allocation means that it is *not* the most robust allocation for all choices of the critical crime rate. For instance, for crime rates equal or higher than 0.049, the nominal optimal allocation is *less* robust than the allocation $b_y = 0.2$. Consequently, if a total crime rate of 0.049 (which is lower than the current rate of 0.050) is acceptable, then we would prefer the allocation $b_y = 0.2$ over the allocation $b_y = 0.268$, since the former is more robust than the latter, while satisficing the total crime rate at 0.049.

Figure 3 illustrates the correlation between the critical crime rate and the most robust allocation. The most robust allocation, $\hat{b}(L_c)$, maximizes the robustness and satisfices the total crime rate at the critical value L_c:

$$\hat{b}(L_c) = \arg\max_{b_y} \hat{\alpha}(b_y, L_c) \tag{2}$$

An important result is that, with the exception of the current allocation, most allocations are the most robust for only one critical value. The current allocation, which is most robust for any crime rate higher than the current crime rate, stands out as a single exception.

The importance of the above observation to the decision maker is that there is no "robust-dominant" decision, an allocation which is more robust than any other allocation for all critical crime rates. The most robust allocation is a function of the satisficing criterion, namely, of the crime rate which the policy maker seeks to achieve. In other words, the robust-satisficing allocation, $\hat{b}(L_c)$, depends on the decision maker's choice of the critical crime rate, L_c. In fact, as proved by proposition D.1, this is not a coincidental result of a specific choice of model and parameters.

Another interesting result is that some allocations are *robust-dominated*: for every critical crime rate there is some other allocation with greater robustness. This is important, since robust-dominated allocations should never be chosen. Sufficient conditions for an allocation to be robust-dominated can be derived, but will not be elaborated here.

The negative slope of figure 3a implies that as the critical crime rate decreases, the robust-optimal allocation requires an increased allocation to "young". This is not surprising since the "young" cohort
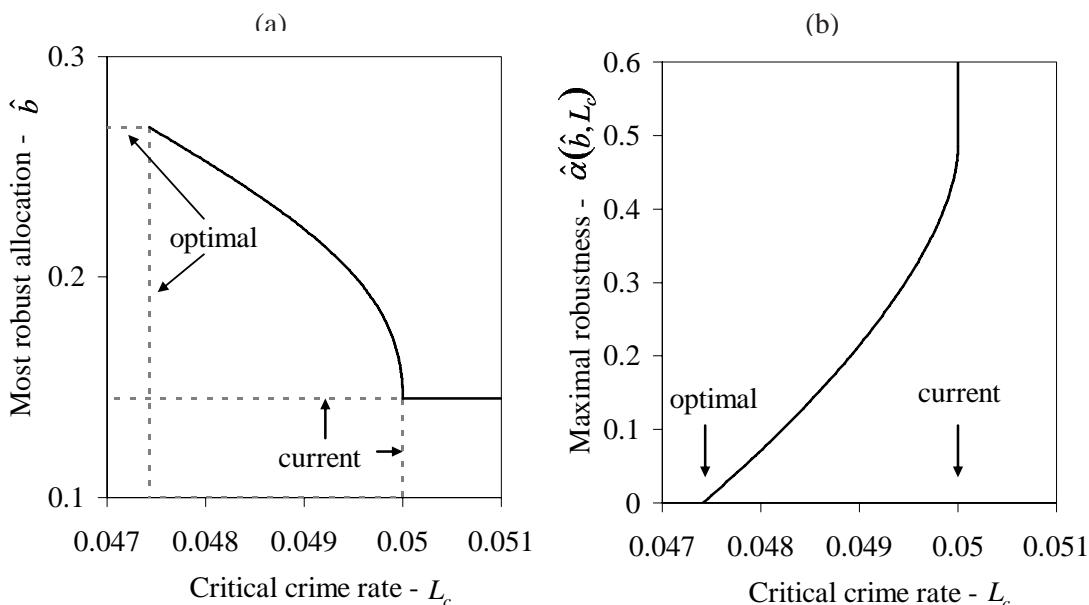
Figure 3: Figure 3a displays the correspondence between the critical crime rate and the most robust allocation. In other words, for any critical crime rate it shows the most robust allocation of policing resources. Proposition D.1 proves that for most allocations there can be only one critical crime rate for which the allocation will be most robust. The proposition also states the correlation between the allocation and the critical crime rate for which this allocation is most robust. Figure 3b illustrates the maximal robustness for any given critical crime rate. Note that as the critical crime rate increases (weaker requirement) the maximal robustness increases.

has higher participation in crime. However, the large negative slope near the "current" crime rate implies that a robust satisficing decision maker is unlikely to make a minor modification to the initial allocation. This is because small changes in the allocation are maximally robust only for negligible improvement in the crime rate. That is, there is a *threshold effect* for the robust satisficing decision maker: changes in the allocation are not robust-optimal for a meaningful reduction in the crime rate until the change exceeds a particular threshold. This threshold is determined by the "bend" in the curve in figure 3a, and occurs around $b_y \approx 0.17$.

The slope of figure 3b may be viewed as the tradeoff between critical crime rate and the maximal robustness. For instance, decreasing the critical crime rate by 0.001 entails a reduction in the maximal robustness by more than 0.1. That is, reducing the number of criminal incidents from 0.049 to 0.048 per person in a 14-month period, "costs" a substantial reduction in the robustness to uncertainty from 0.21 to 0.08. Near the current allocation the slope is very high (asymptotically infinite), implying that a small decrease in the critical crime rate has a great effect on the maximal robustness.

At any point of the curve of figure 3b, its slope equals the slope of the maximal robustness curve for that critical crime rate. The difference between maximal robustness at L_c, and the robustness of the nominal-optimal allocation at L_c, is called the "robustness premium" for the former allocation. A large

robustness premium implies a strong preference for the robust-optimal allocation over the nominal-optimal allocation. The robustness premium is calculated as the difference between the curve and the tangent at the nominal optimum. The low curvature over most of figure 3b implies low robustness-premium for maximal robustness over this range of L_c values. For instance, at $L_c = 0.049$, the robustness premium is $\Delta_{\widehat{\alpha}} = 0.21 - 0.18 = 0.03$. The maximal robustness at $L_c = 0.049$ is 0.21, so the robustness premium is thus only about 15% of the maximal robustness. In other words, by choosing the nominal-optimal allocation for the critical crime rate $L_c = 0.049$, we loose approximately 15% of the robustness. Conversely, the high curvature of figure 3b near the current allocation implies large robustness premium for the robust-optimal allocation in that range. In summary, the policy implication of the curvature of figure 3b is that small reductions below the current crime rate have substantial robustness premium, while large reductions have small differences in robustness between the nominal and the robust-optimal allocations.

This is different from the threshold effect mentioned earlier. The large robustness premium, for small changes in the current allocation, corresponds to very small improvement in the critical crime rate (this is the threshold effect). Large robustness premium by itself does not motivate the policy maker to change the allocation. The policy maker will require large robustness for acceptable (not negligible) reduction in crime.

What if the fraction of "young" change? This can happen gradually, as a result of a demographic change, or suddenly, by applying the our model to a specific sub-population (for instance, when considering the police enforcement in regions with different fraction of "young" drivers. Figure 4 illustrates the most robust allocation and the maximal robustness as a function of the fraction of "young" in the population.

The positive slope of figure 4a implies that as the fraction of "young" in the population increases, a robust satisficing decision maker would increase the fraction of resources allocated to the "young" group. However, as figure 4b demonstrates, this does not mean that the robustness will also increase. The robustness has a turn-around effect — from some point, an increase in the fraction of "young" (the more responsive group) actually decreases the robustness. This is because, as the fraction of "young" increases, the most robust allocation tends to allocate more and more resources to the "young" group, thus moving further away from the observed allocation, which entails increased uncertainty in the responsiveness functions.

The slope of the two curves expresses the response of the allocation and the robustness to gradual changes in the composition of the population. For instance, the slope of figure 4a is only slightly less than unity. A change of 1% in the fraction of "young" within the population (relative to the current $\pi_y = 0.145$) will cause a change of only 0.9% in the robust-optimal allocation. Thus demographic changes are matched by similar changes in the robust-optimal allocation. Similarly, a 1% change in the fraction of "young" results in a change of approximately 0.001 in the maximal robustness (about 1.5%). Thus, both the robust-optimal allocation and the maximal robustness are will follow gradual changes in the fraction of the "young" drivers within the general population.
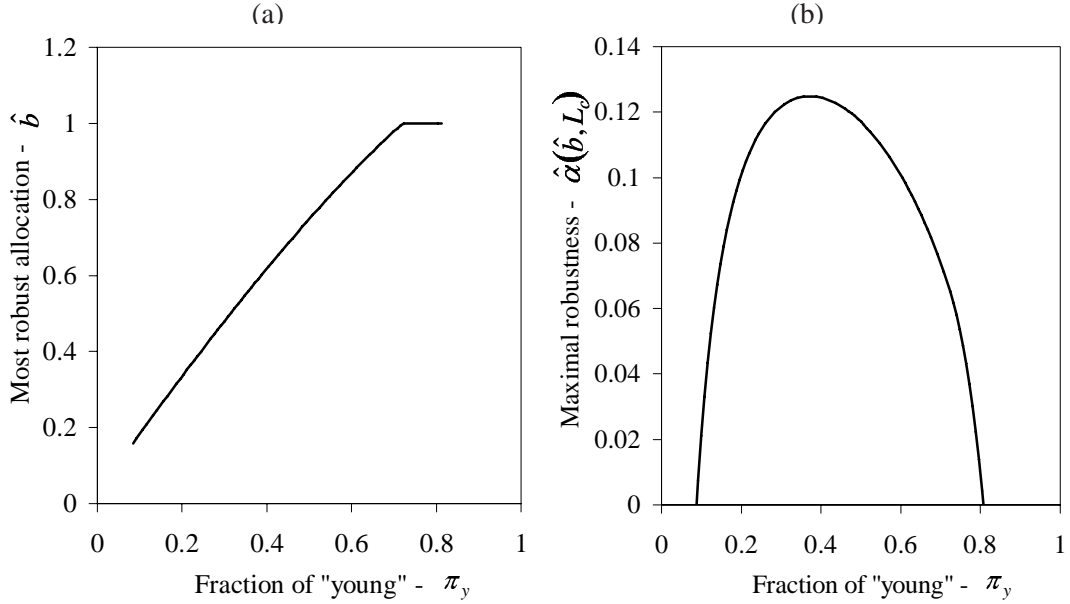
| | (a) | | | (b) | |

Figure 4: Figure 4a displays the correspondence between the fraction of "young" in the general population and the most robust allocation. In other words, for any composition of the population it shows the most robust allocation of policing resources. Figure 4b illustrates the maximal robustness for any given fraction of "young" within the general population. Both figures assume critical crime rate $L_c = 0.048$.

4   Robust-Satisficing vs. Minimax

One might get the impression that robust-satisficing is very similar (if not identical) to minimax (or maximin) or various worst case or robust control strategies. While there are similarities, there are crucial differences in policy-selection between the two methods. We will briefly describe the similarities and differences.

In both methods, we are facing a set $\mathcal{U}$ of possible states of the world, and have to choose a decision q from a set $\mathcal{Q}$ of possible decisions, where the consequence of q is expressed by a loss function $L(q, u)$. Here, $u \in \mathcal{U}$ is the *unknown* true state of the world. For convenient comparison we shall assume that there is an estimate $\widetilde{u} \in \mathcal{U}$ of the state of the world, and that $\mathcal{U}(\widetilde{u}, \alpha) \subseteq \mathcal{U}$ denotes the set of states accessible at horizon of uncertainty α. The horizon of uncertainty, of course, is unknown.

The technical difference between the two methods is the fixed parameter. When minimaxing, we assume that we know the horizon of uncertainty and that its value is α_m. Then, the minimax decision maker looks for the decision $q^\star$ that guarantees the minimal loss for any state of the world at uncertainty α_m. In other words, the minimax decision is:

$$q^\star = \arg\min_{q \in \mathcal{Q}} \max_{u \in \mathcal{U}(\alpha_m, \widetilde{u})} L(q, u) \tag{3}$$

11

On the other hand, when robust-satisficing, we choose a critical loss L_c, and we have no idea what is the worst case, since we do not know the true horizon of uncertainty. Now we are looking for the decision that has the greatest tolerance (robustness) for mistakes in the estimate of $\tilde{u}$. In other words, we are looking for a decision which guarantees that the critical loss will not be exceeded even when faced with a great deviation from the estimate $\tilde{u}$. The robust-satisficing decision is:

$$\hat{q} = \arg\max_{q \in \mathcal{Q}} \hat{\alpha}(q, L_c) \tag{4}$$

where $\hat{\alpha}(q, L_c)$ is the robustness, discussed in section 2 and defined formally in eq. (A3) of appendix A.

$q^\star$ and $\hat{q}$ need not necessarily differ. If $\alpha_m = \hat{\alpha}(\hat{q}, L_c)$, then we have $q^\star = \hat{q}$. Similarly, if L_c equals the minimax loss, then, again, we have $q^\star = \hat{q}$.

However, from the decision maker's point of view, minimax and robust-satisficing can differ. First of all, if the decision maker has no estimate of the uncertainty, α_m, then the minimax strategy cannot be implemented. However, even if the decision maker agrees that the horizon of uncertainty can be as large as α_m, the critical loss, L_c, may be less than the corresponding minimax loss. In this case $\hat{q}$ and $q^\star$ differ, and the $\hat{q}$ is more robust to uncertainty than $q^\star$, as we now demonstrate.

The relation between minimax and robust-satisficing decision strategies can be illustrated with the help of figure 2, part of which is reproduced in figure 5. Consider the choice between the two options, $b_y = 0.268$ and $b_y = 0.2$. Suppose the decision maker believes the uncertainty is $\alpha_m = 0.4$, but desires (or is required) to keep the crime rate below $L_c = 0.0485$. The minimax decision at $\alpha_m = 0.4$ is $b_y = 0.2$, which is also the robust-satisficing decision for the corresponding critical loss, $L_\star = 0.050$. However, the decision maker needs to keep the crime rate below $L_c = 0.0485$. The robust-satisficing decision is $b_y = 0.268$, whose robustness is $\hat{\alpha}(0.268, 0.0485) = 0.13$, which is greater than the robustness of the minimax decision at this critical crime rate, $\hat{\alpha}(0.2, 0.0485) = 0.09$. The loss will not exceed 0.0485 for a wider range of contingencies with $b_y = 0.268$ than with $b_y = 0.2$. The robust-satisficing and minimax decisions differ. The decision maker who is concerned to keep the crime rate below 0.0485, regardless of whatever beliefs are held regarding the horizon of uncertainty, might justifiably prefer the robust-satisficing strategy over the minimax strategy.

5   Conclusion

The economic theory of crime views criminals as rational agents who adapt their behavior in response to costs and benefits. This implies that involvement in criminal activity will respond with negative elasticity to changes in penalties or probabilities of apprehension. Since different groups respond differently, knowledge of the elasticities (or the responsiveness functions) would enable efficient allocation of enforcement resources. However, under a set budget, differential allocation of fixed total resources—profiling—can augment both the number of arrests and the total crime rate, since non-profiled groups will increase their criminal activity. Specifically, profiling a minority can cause increased total arrests (mostly in the minority) but also increased total crime since the majority responds rationally to decreased enforcement by engaging in more crime.
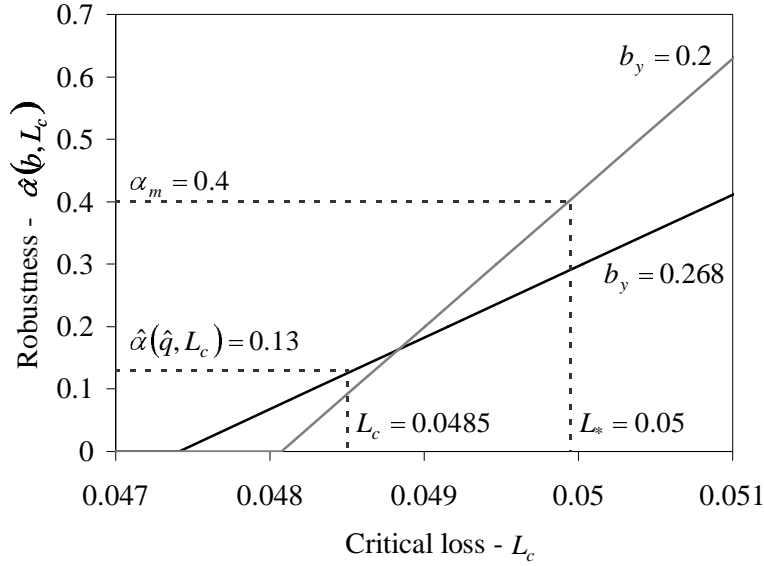
Figure 5: Relation between robust-satisficing and minimax.

We have focussed on the problem of formulating a profiling strategy in light of the great uncertainty accompanying estimates of responsiveness to law enforcement. Since elastic response to profiling can result in increased total crime, the advocate of profiling must choose a strategy which will not inadvertently result in this undesired outcome.

This paper has developed a robust-satisficing methodology for allocation of enforcement resources when the responsiveness functions are highly uncertain. We have used info-gap theory for satisficing (not minimizing) the total crime rate. We have demonstrated the trade-off between robustness to uncertainty on the one hand, and reduction of total crime on the other hand. Attempting to minimize total crime has zero robustness to uncertainty in the responsiveness to policing. Since the responsiveness to policing is highly uncertain, low robustness is undesirable. Positive robustness is obtained only by aiming at a crime rate which is larger than the estimated minimum. The robust-satisficing strategy chooses an allocation which guarantees an acceptable total crime rate (which usually will not be the estimated minimum), for the largest possible range of error in the estimated elasticities. The robustness analysis enables the decision maker to evaluate profiling options in terms of whether they promise adequate improvements in total crime, at plausible levels of immunity to error in the responsiveness functions.

We have presented an empirical example based on measurement of the responsiveness to enforcement of traffic laws. We demonstrated a "threshold effect": changes in the allocation are not robust-optimal for a meaningful reduction in the crime rate until the change exceeds a particular threshold. We have also seen the effect of changing demographics on the robust-optimal profiling strategy. While the allocation changes approximately in parallel to the changing composition of the population, the robustness changes non-linearly, showing a maximum at an intermediate fraction of "young" drivers. Since it is the robustness premium that motivates adopting the robust-satisficing allocation, this implies

that not all demographic changes should induce shifts in policy.

We have not addressed the ethical aspect of profiling. However, we note that arguments for profiling which are based on the utility of optimal profiling (rather than satisficing) based on best-estimates of the responsiveness functions, should be viewed skeptically. We have shown that optimal allocations have zero robustness to error and, since responsiveness functions are highly uncertain, the purported benefits of optimal allocations are highly unreliable. If profiling can be justified on utilitarian grounds, such justification must rest on showing that desirable reduction of total crime can be obtained with adequate robustness to the main source of uncertainty (the responsiveness functions). That is, the strategy of robust-satisficing is directly relevant to the ethical argument for (or against) profiling.

We have studied the profiling of two groups with uncertain responsiveness to policing, and illustrated our results with estimated responsiveness to policing of running red lights. The extension of our results to multi-group profiling is straightforward. An additional important extension is to study the dynamic interaction between enforcement and criminal activity, in which each side learns about the other.

A   Info-Gap Theory: A Mathematical Précis

Let $\widetilde{u}$ denote our best estimate of u, a parameter, vector, function, and the like, which is used to estimate the loss $L(q, u)$ due to decision $q \in \mathcal{Q}$. An info-gap model is an unbounded family of nested sets, $\mathcal{U}(\alpha, \widetilde{u})$, of u-values. As α gets larger, the sets become more inclusive. The info-gap model expresses the decision maker's beliefs about uncertain variation of u around $\widetilde{u}$.

Info-gap models obeys two axioms:

$$\text{Contraction:} \qquad \mathcal{U}(0, \widetilde{u}) = \{\widetilde{u}\} \tag{A1}$$
$$\text{Nesting:} \qquad \alpha < \alpha' \text{ implies } \mathcal{U}(\alpha, \widetilde{u}) \subseteq \mathcal{U}(\alpha', \widetilde{u}) \tag{A2}$$

Contraction asserts that $\widetilde{u}$ is the only possibility in the absence of uncertainty, $\alpha = 0$. Nesting asserts that the sets become more inclusive as α gets larger.

Given a critical loss L_c, the *robustness function* $\widehat{\alpha}(q, L_c)$ is the greatest level of uncertainty α which still guarantees a loss no greater than L_c:

$$\widehat{\alpha}(q, L_c) = \max \left\{ \alpha : \left(\max_{u \in \mathcal{U}(\alpha, \widetilde{u})} L(q, u) \right) \leq L_c \right\} \tag{A3}$$

Robust-satisficing decision making maximizes the robustness and satisfices the loss at the value L_c, without specifying a limit on the level of uncertainty:

$$\widehat{q} = \arg \max_{q \in \mathcal{Q}} \widehat{\alpha}(q, L_c) \tag{A4}$$

B Estimating the Responsiveness Function, $\widetilde{C}$

Our best, but highly uncertain, guess of the responsiveness function is:

$$\widetilde{C}_i = \exp\left(-\mu_i p_i f - \delta_i\right) \tag{B1}$$

Here, p_i is the probability of a driver from the ith group being caught after driving through a red light, and f is the fine for running a red light. Notice that this model assumes *risk neutrality* — a utility maximizing driver will be indifferent between an increase to the fine and an increase in the probability of detection (as long as the factor of increase is identical). μ_i and δ_i are parameters which characterize the responsiveness of the ith group. This is precisely eq. (1).

The probability of detection is proportional to the fraction of policing resources which are allocated to the ith group, b_i:

$$p_i = \frac{b_i}{\pi_i}p \tag{B2}$$

where π_i is the fraction of the ith group within the general population, and p is the probability of detection under a fair allocation, $b_i = \pi_i$. Assuming p is sufficiently small[1], we have $0 < p_i < 1$. We can now rewrite eq. (B1) in the following way:

$$\widetilde{C}_i = \exp\left(-\mu_i \frac{b_i}{\pi_i}pf - \delta_i\right) \tag{B3}$$

We have two measurements of the crime rate for each group, due to Bar-Ilan and Sacerdote (2004). These two measurements relate to two (known) fines: f^{before}, and f^{after}. We shall denote the two measured crime rates by C_i^{before} and C_i^{after} for group i. We will assume that for both measurements the allocation was "fair" ($b_i = \pi_i$). (This is justified since the traffic violations were detected by automatic sensors.) Thus eq. (B3) becomes, for 'before' and 'after':

$$C_i^{\text{before}} = \exp(-\mu_i p f^{\text{before}} - \delta_i) \tag{B4}$$
$$C_i^{\text{after}} = \exp(-\mu_i p f^{\text{after}} - \delta_i) \tag{B5}$$

We can now use measured values of C_i^{before} and C_i^{after} to calculate the values of δ_i and of $\mu_i p$ for each group.

We will define the following quantity for the ith group, whose value is known based on the estimates of $\mu_i p$:

$$\gamma_i = \mu_i p f^{\text{after}} \tag{B6}$$

Now the responsiveness to allocation of policing, eq. (B3), evaluated for the increased fine, f^{after}, can be expressed succinctly as:

$$\widetilde{C}_i(b_i) = \exp\left(-\gamma_i \frac{b_i}{\pi_i} - \delta_i\right) \tag{B7}$$

---

[1] Israeli police (2007) estimates it as 0.5%.

15

Table B1: MEASURED CRIME RATES

| Group | Crime Rate | | | Responsiveness Parameters | |
	Before Increase	After Increase	Fraction	γ_i	δ_i
Age 17-30 (y)	0.123	0.056	0.145	1.31	1.57
Age 31+ $(\bar{y})$	0.065	0.049	0.855	0.47	2.54

NOTE.—"Crime rate" is the mean number of tickets during the 14 quarter period before the fine increase and the 14 quarter period after the fine increase (Bar-Ilan and Sacerdote 2004). "Fraction" is the group's relative fraction within the general population of Israeli drivers, based on a random sample of 1% of the Israeli drivers (Bar-Ilan and Sacerdote 2004).

$\widetilde{C}_i(b_i)$ is the estimate of the average number of violations, per unit of time, per individual in the ith group, given an allocation b_i of policing resources to this group. This is precisely eq. (1).

Table B displays the responsiveness parameters, γ_i and δ_i, for the two groups profiled in our example. Also appearing in the table are the data needed to calculate the responsiveness parameters: the crime rate before and after the increase to the fine, C_i^{before} and C_i^{after}, and the fraction of the different groups within the general population.

C   Satisficing Crime Rate

C.1   Info-Gap Model

In Appendix B we derived an estimate to the approximated responsiveness functions of the different groups. The estimation was based on two measurements of the crime rate, given two levels of fine. Let b^0 denote the allocation of policing resources during the periods analyzed by Bar-Ilan and Sacerdote. As mentioned above, we assume that the allocation was fair, $b_i^0 = \pi_i$.

We may be fairly confident in the observed crime rate for b^0 at the time of measurement. However, it is reasonable to suppose that the uncertainty in the responsiveness function grows as the difference between the sampled resource allocation and the current resource allocation grows.

Let $\widetilde{C}$ be a vector of responsiveness functions, representing our best estimate of the responsiveness functions of the different groups. $\widetilde{C}_i$ will be the estimated exponential model, depicted in eqs. (1) and (B7). We will refer to $\widetilde{C}$ as the nominal model, and represent the uncertainty surrounding the actual responsiveness functions using the following info-gap model:

$$\mathcal{U}(\alpha, \widetilde{C}) = \left\{ C : C_i(b_i) \geq 0, \left| \frac{C_i(b_i) - \widetilde{C}_i(b_i)}{\widetilde{C}_i(b_i)} \right| \leq \alpha \left| \frac{b_i - b_i^0}{b_i^0} \right| \right\}, \quad \alpha \geq 0 \qquad \text{(C1)}$$

At any horizon of uncertainty α, the set $\mathcal{U}(\alpha, \widetilde{C})$ contains all non-negative responsiveness functions $C_i(b_i)$ which deviate from the nominal function by no more than $\alpha \left| \frac{b_i - b_i^0}{b_i^0} \right|$. Since α is unbounded, this is an unbounded family of nested sets of responsiveness functions.

The weight on the horizon of uncertainty (the absolute value term on the righthand side of the inequality) means that for any given horizon of uncertainty, the uncertainty regarding the responsiveness grows as the allocation, b, deviates from the measured reference allocation, b^0. An uncertainty envelope for this info-gap model is illustrated in figure 1.

C.2 Robustness

Combining eqs. (C1) and (A3), one can readily show that, for any allocation which distributes the policing resources between the groups y and $\bar{y}$, the robustness is:

$$
\begin{aligned}
\widehat{\alpha}(b, L_c) &= \max\left\{\alpha : \left(\max_{C \in \mathcal{U}(\alpha, \widetilde{C})} \sum_{i \in \{y, \bar{y}\}} \pi_i C_i(b_i)\right) \leq L_c\right\} \\
&= \frac{L_c - \sum_{i \in \{y, \bar{y}\}} \pi_i \widetilde{C}_i(b_i)}{\sum_{i \in \{y, \bar{y}\}} \pi_i \widetilde{C}_i(b_i) \left|\frac{b_i - b_i^0}{b_i^0}\right|}
\end{aligned}
\tag{C2}
$$

D Proposition and Proof

For the following proposition we will assume that $\widetilde{C}(b)$ is a vector of differentiable, positive, monotonic decreasing and strictly convex functions. y and $\bar{y}$ will denote two disjoint and complementary groups. $\widetilde{C}_i$ represents our best estimate of the responsiveness of the ith group to policing. We will refer to $\widetilde{C}$ as *the nominal model*. Note that $\widetilde{C}_i$ need not necessarily be exponential, as assumed in eq. (1). Thus eq. (1) is a special case.

b^0 is the *observed allocation*, the only allocation of police resources for which the responsivenesses of the groups are known. b^{opt} will denote the *optimal allocation*, the allocation which yields the lowest crime rate under the nominal model.

For later simplicity, we will introduce the following definitions:

$$
\xi = \sum_{i \in \{y, \bar{y}\}} \pi_i \widetilde{C}_i(b_i)
\tag{D1}
$$

$$
\xi' = \frac{\partial \xi}{\partial b_y} - \frac{\partial \xi}{\partial b_{\bar{y}}}
\tag{D2}
$$

$$
\zeta = \sum_{i \in \{y, \bar{y}\}} \pi_i \widetilde{C}_i(b_i) \left|\frac{b_i - b_i^0}{b_i^0}\right|
\tag{D3}
$$

$$
\zeta' = \frac{\partial \zeta}{\partial b_y} - \frac{\partial \zeta}{\partial b_{\bar{y}}}
\tag{D4}
$$

The gist of our proposition is that, for most resource allocations, there can be at most one critical crime rate for which the allocation is most robust. In fact, there are only four potential exceptions:

17

the current (observed) allocation, the optimal allocation, and the two extreme allocations, $b_y = 0$ and $b_y = 1$.

Proposition D.1 *Nearly all profiling allocations are maximally robust for at most a single critical crime rate.*

Let b be an allocation such that $b \neq b^0$, $b_y > 0$ and $b_{\bar{y}} > 0$. If $\zeta' \neq 0$, or $b \neq b^{opt}$, then there is at most one critical crime rate L_c for which b is most robust. When there is such a critical crime rate L_c, it is

$$L_c = \xi - \frac{\xi' \zeta}{\zeta'} \tag{D5}$$

Proof: We will separate the above proposition into two claims.

1. *If $\zeta' \neq 0$, then there is at most one critical crime rate L_c for which b is most robust, given by eq. (D5).*

   Since $b_y > 0$ and $b_{\bar{y}} > 0$ (meaning b is not on the "edge" of the valid values), then b being most robust for some critical value L_c must mean

   $$\left(\frac{\partial}{\partial b_y} - \frac{\partial}{\partial b_{\bar{y}}} \right) \widehat{\alpha}(b_y, L_c) = 0 \tag{D6}$$

   In other words, increasing the policing resources of the y group by an infinitesimal amount, at the expense of the $\bar{y}$ group (or vice versa), will not change the robustness. Recall that $b_y + b_{\bar{y}} = 1$.

   By differentiating eq. (C2) we have

   $$\left(\frac{\partial}{\partial b_y} - \frac{\partial}{\partial b_{\bar{y}}} \right) \widehat{\alpha}(b_y, L_c) = \frac{-\xi' \zeta + \zeta' \xi - \zeta' L_c}{\zeta^2} \tag{D7}$$

   Thus, a necessary condition for b to maximize the robustness is that the term on the righthand side of eq. (D7) equals zero. At fixed b, this can hold only for a single value of L_c, since ζ' is non-zero. Thus, if b maximizes the robustness, it does so for only a single L_c value, given by eq. (D5).

2. *If $b \neq b^{opt}$, then there is at most one critical crime rate L_c for which b is most robust.*

   We will assume that b is the most robust allocation for more than one critical crime rate, and arrive at a contradiction.

   If b is the most robust allocation for some critical value L_c, then, as shown in eqs. (D6) and (D7), it must hold that

   $$\frac{-\xi' \zeta + \zeta' \xi - \zeta' L_c}{\zeta^2} = 0 \tag{D8}$$

18

In order for b to be most robust for some other critical value $L'_c \neq L_c$ it must hold that $\zeta' = 0$, since eq. (D8) must hold for both L_c and L'_c. This in turn means that $-\xi'/\zeta = 0$, and since $\zeta > 0$, this means that $\xi' = 0$. Recall that $\xi' = \left(\frac{\partial}{\partial b_y} - \frac{\partial}{\partial b_{\bar{y}}} \right) \sum_{i \in \{y, \bar{y}\}} \pi_i \widetilde{C}_i(b_i)$, which is the change in the total crime rate when trading an infinitesimal amount of resources between the two groups. Both $\widetilde{C}_y$ and $\widetilde{C}_{\bar{y}}$ are strictly convex functions, which means that their sum, $\sum_{i \in \{y, \bar{y}\}} \pi_i \widetilde{C}_i(b_i)$, is also strictly convex. If the sum is strictly convex and the derivative is zero, then b is the minimum of the total loss, which means $b = b^{opt}$. This contradicts the suppostion that $b \neq b^{opt}$, so b is most robust for at most one value of L_c, given by eq. (D5).

∎

References

[1] Arrow, Kenneth J., 1973. The Theory of Discrimination. 3–33 in *Discrimination in Labor Markets*, edited by Orley Ashenfelter. Princeton NJ: Princeton University Press.

[2] Bar-Ilan, Avner, and Bruce Sacerdote, 2004. The Response of Criminals and Noncriminals to Fines. *Journal of Law and Economics* 47:1–17.

[3] Becker, Gary S., 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169–217.

[4] Ben-Haim, Yakov, 2006. *Info-Gap Theory: Decisions Under Severe Uncertainty*. 2nd ed. Academic Press.

[5] Benson, Bruce L., and Simon W. Bowmaker, 2005. Economics of crime. 101–36 in *Economics Uncut: A Complete Guide to Life, Death, and Misadventure*, edited by Simon W. Bowmaker. Northhampton: Edward Elgar Publishing.

[6] Blumkin, Tomer, and Yoram Margalioth, 2005. Economic Analysis of Racial Profiling Rules. Discussion Paper No. 05-06. Monaster Center for Economic Research, Ben-Gurion University of the Negev.

[7] Borooah, Vani K, 2001. Racial Bias in Police Stops and Searches: An Economic Analysis. *European Journal of Political Economy* 17:17–37.

[8] Burgess, Ernest W., 1928. Factors Determining Success or Failure on Parole. 205–249 in *The Working of the Indeterminate Sentence Law and the Parole System in Illinois*, edited by Andrew A. Bruce. Springfield, IL: Illinois Committee on Indeterminate-Sentence Law and Parole.

[9] Ehrlich, Isaac, and Zhiqiang Liu, 1999. Sensitivity Analyses of the Deterrence Hypothesis: Lets Keep the Econ in Econometrics. *Journal of Law and Economics* 42:455–87.

[10] Farmer, Amy, and Dek Terrell, 2001. Crime versus Justice: Is There a Trade-Off? *Journal of Law and Economics* 44:345–66.

[11] Harcourt, Bernard E., 2006. Muslim Profiles Post 9/11: Is Racial Profiling an Effective Counterterrorist Measure and Does It Violate the Right to be Free From Discrimination? Public Law and Legal Theory Working Paper No. 123 The University of Chicago.

[12] Harcourt, Bernard E., 2007. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. Chicago: The University of Chicago Press.

[13] Heaton, Paul, 2006. Understanding the Effects of Anti-Profiling Policies. Unpublished manuscript. University of Chicago.

[14] Hernández-Murillo, Rubén, and John Knowles, 2004. Racial Profiling or Racist Policing? Bounds Test in Aggregate Data. *International Economic Review* 45(3):959–989.

[15] Israel Police, 2007. Traffic Department.
http://www.police.gov.il/english/Traffic/equipment/xx_en_tr_equipment.asp (last updated August 28, 2007).

[16] Klick, Jonathan, and Alexander T. Tabarrok, 2005. Using Terror Alert Levels to Estimate the Effect of Police on Crime. *Journal of Law and Economics* 48:267–79.

[17] Knowles, John, Nicola Persico, and Petra Todd, 2001. Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy* 109:203–29.

[18] Lamberth, John, 1994. Revised Statistical Analysis of the Incidence of Police Stops and Arrests of Black Drivers/Travelers on the New Jersey Turnpike Between Exits or Interchanges 1 and 3 From the Years 1988 Through 1991. Unpublished manuscript. Temple University.

[19] Levitt, Steven D., 1997. Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *The American Economic Review* 87:270–90.

[20] Levitt, Steven D., 1998. Why Do Increased Arrest Rates Appear To Reduce Crime: Deterrence, Incapacitation, Or Measurement Error? *Economic Inquiry* 36:353–72.

[21] Malik, Arun S., 1990. Avoidance, Screening and Optimum Enforcement. *RAND Journal of Economics* 21:341–53.

[22] Persico, Nicola, 2002. Racial Profiling, Fairness, and Effectiveness of Policing. *The American Economic Review* 92:1472–97.

[23] Polinsky, A. Mitchell, and Steven Shavell, 2000. The Economic Theory of Public Enforcement of Law. *Journal of Economic Literature* 38:45–76.

[24] Pradiptyo, Rimawan, 2007. Does Punishment Matter? A Refinement of the Inspection Game. *Review of Law and Economics* 3.

[25] Tsebelis, George, 1990. Penalty Has No Impact on Crime? A Game Theoretical Analysis. *Rationality and Society* 2:255–86.