Essays on the Economic Analysis of Discrimination, Law Enforcement, and Smoking

by

Brian C. Rowe

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2009

Doctoral Committee:

Assistant Professor Daniel S. Silverman, Chair
Professor James R. Hines Jr.
Professor Jeffrey A. Smith
Assistant Professor James J. Prescott

To my wife Christine and my family

ACKNOWLEDGEMENTS

I am grateful to everyone who helped me with this research and I sincerely thank them for their valuable comments, criticism, and suggestions. In particular I thank Martha Bailey, Alexia Brunet, Charlie Brown, Christopher Carpenter, Paul Courant, Lucas Davis, John Dinardo, Daniel Eisenberg, Jim Hines, Osborne Jackson, Jung-min Lee, Stephan Lindner, Yusufcan Masatlioglu, Zoe McLaren, Matt Rutledge, J.J. Prescott, Dan Silverman, Doug Smith, Jeff Smith, and Gary Solon. I also thank seminar participants at the University of Michigan, the Midwest Economic Association's 2008 meeting, the American Law and Economics Association's 2008 meeting, and the Conference on Empirical Legal Studies 2008. Special thanks to Kate Antonovics, Bill Dedman, and Nicola Persico for sharing data with me.

For comments on "The Effect of Smoking in Young Adulthood on Smoking Later in Life", my co-author Daniel Eisenberg and I thank Joshua Angrist, Tom Buchmueller, Richard Frank, Haiden Huskamp, Edward Norton, and seminar participants at the Ann Arbor VA, the American Society of Health Economists 2006 meeting, and the Triangle Health Economics Seminar. We also thank Joshua Angrist for sharing data and Clint Carter, Bob Krasowski, Maggie Levenstein, Peter Meyer, and Chris Rogers for assistance with accessing restricted data from the National Health Interview Survey. Funding is gratefully acknowledged from the University of Michigan Tobacco Research Network (UMTRN), which is supported by the American Legacy Foundation. The views expressed in our paper do not reflect the views of the Foundation, Foundation staff, or its Board of Directors. Our paper is forthcoming in *Forum for Health Economics and Policy.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table

CHAPTER I

Introduction

This dissertation examines issues related to the efficiency and effectiveness of government policies to provide public goods. In the first essay, I develop an empirical test for whether police officers discriminate on driver gender when enforcing traffic laws. The test is designed to only detect discrimination that is unrelated to providing safe conditions on the roads. The empirical method developed in the essay may be applicable in a number of other contexts where evaluators (such as police officers, judges, or mortgage lenders) may potentially discriminate when making decisions regarding subjects who belong to different demographic groups.

The second essay makes a theoretical argument showing that because police officers can detect many different crimes by making a traffic stop, the widespread practice of giving stopped traffic law violators a warning instead of a fine can be efficient. Warnings would at first seem to be inefficient because they lower the expected penalty from breaking the law, and thereby reduce deterrence for a given amount of public resources devoted to detecting and stopping violations. My argument therefore points out an efficiency rationale for providing individual government agents discretion in deciding which detected law breakers to penalize.

In the third and final essay, my co-author Daniel Eisenberg and I use the Vietnam draft lottery to test the commonly held presumption that smoking as a young person strongly predicts smoking in later adulthood. This presumption, well documented by many observational studies, underlies many anti-smoking policies in the United States. Yet some of the persistence of smoking over time might be attributable to individual factors, such as tolerance for health risks, which are difficult to account for in observational data. Using variation in smoking induced by the draft lottery, we do not find a strong relationship between smoking in early and late adulthood, suggesting that anti-smoking policies directed at young people may not be effective in achieving the policy goal of reducing adult smoking rates.

CHAPTER II

Gender Bias in the Enforcement of Traffic Laws: Evidence based on a new empirical test

2.1 Introduction

Traffic enforcement in the United States imposes a disparate impact on male drivers. In 2005, 63.4% of all traffic tickets in the U.S. were issued to males.[1] Furthermore, the gender disparity in tickets is in excess of the male share of the driving population. In 2005, 10.8 percent of all male drivers but only 6.8 percent of female drivers were stopped by police, and after being stopped males were more likely to be ticketed (Durose et al. 2007).

Traffic accidents are a significant public health problem in the United States.[2] Because of this, road safety would be viewed as a legitimate law enforcement objective by the courts. In the U.S., police practices which impose a disparate impact on a demographic group are often (but not always) upheld by the courts if the disparate impact is a byproduct of a legitimate law enforcement objective. On the other hand, police practices which seem based on prejudice or are unrelated to effective law enforcement are not permitted.[3]

In this way, the framework for determining the legality of police practices accords well with the distinction between statistical and taste-based discrimination (bias) in

[1] According to Durose et al. (2007) in a Bureau of Justice Statistics special report, 11 million male drivers and 6.9 million female drivers were stopped by police nationwide in 2005. 59.2% of the stopped male drivers were ticketed while 54.4% of the female drivers received a ticket. These figures imply that about 63.4% of all traffic tickets were issued to male drivers.

[2] Approximately 42,000 people were killed and 2.5 million people were injured in traffic accidents in 2006 (2006 Annual Assessment of Motor Vehicle Crashes, National Highway Traffic Safety Administration).

[3] See Knowles, Persico, and Todd (2001) for a more thorough discussion of the relevant legal background. The concept of an "unjustified disparate impact" is discussed in detail by Ayres (2002).

economics. Statistical discrimination can produce disparate impacts which are due to a legitimate objective and may be permissible. Similarly, omitted variables related to criminality and correlated with race or gender can produce disparate impacts even if the police are concerned only with effective law enforcement. This paper develops and conducts an empirical test for police gender bias in traffic enforcement.

It is difficult to determine empirically if the disparate impact of a police practice is at least partly due to bias. To solve this problem in the context of traffic enforcement, I develop a model of police preferences and driver behavior which provides a testable implication of gender biased ticketing. The testable implication is in terms of what I call the "officer gender effect": Conditional on breaking a traffic law, does the probability that a female driver receives a ticket depend on the gender of the officer who observes the violation? The model serves to clarify the conditions which are required to infer that a bias exists if this officer gender effect is found empirically.

In the model, the police receive a greater benefit from ticketing more dangerous traffic violations. In this way, the model provides an underlying motivation for traffic ticketing which is connected to the objective of safety on the roads. Officers incur a cost from ticketing a driver, and the cost of ticketing is allowed to vary with both the gender of the police officer and the gender of the driver. The ticketing costs reflect a taste for discrimination. If an officer's cost of ticketing male drivers is lower than his cost of ticketing females, then all else equal the officer will derive more utility from ticketing males. As the cost of ticketing male drivers increases, the officer increases his violation threshold for males, which is the least dangerous traffic violation for which he is willing to ticket a male driver.

The test for gender bias is based on the model's prediction for what the sign of the officer gender effect should be if male and female police use unbiased (equal for each driver gender) but different violation thresholds. In this case, the officer gender using the higher threshold will be relatively more likely to ticket male drivers who commit violations. This prediction depends on assuming that male drivers are more dangerous, in that they are more likely to commit a traffic violation of severity level above a given threshold. I show that this assumption is supported by several patterns in the Boston data, as well as by findings from other research.[4] Intuitively, relatively fewer female drivers commit violations which are dangerous enough to exceed a high threshold.

Estimating the officer gender effect is a difficult exercise because only drivers who

[4]For example, Levitt and Porter (2001) find that the two-car fatal crash risk for male drivers is 3 times higher than that of female drivers.

received tickets appear in the data, so it is impossible to condition on breaking the law. If male and female officers observed the same pool of drivers who broke the law, the officer gender effect is identified simply as the empirical effect of the police officer being male on the probability that a ticketed driver is female. In practice, male and female officers might monitor different areas of the city or tend to patrol at different times, and thereby observe different pools of drivers. To correct for this I use an extensive set of traffic stop level controls to account for any variation in the pool of drivers by observable characteristics such as time of day, day of week, and location in the city of Boston. If male and female officers observe the same pool of drivers after conditioning on this information, the empirical effect of the officer being male on the probability that a ticketed driver is female equals the officer gender effect. This follows from logic similar to that of Grogger and Ridgeway (2006). I examine the validity of this strategy by looking at a variety of evidence in the traffic ticket data and from some external sources.

To rank male and female officer's violation thresholds, I estimate how the miles-per-hour over the limit or dollar fine amount of ticketed violations depends on the gender of the police officer. I show that this method is valid if the rank order of average miles-per-hour preserves the rank order of average violation severity, and if on average male drivers commit violations which are at least as dangerous as those committed by females.[5]

When applied to data on traffic tickets issued in Boston, my test rejects the null hypothesis of no gender bias in favor of the alternative that at least one officer gender is biased. First, male officers were less likely than female officers to ticket female drivers. I find no evidence that this effect is due to differences in the pools of drivers observed, so I infer that male officers were less likely to ticket female drivers who broke the law. Second, male officers were "tougher" because they issued tickets for relatively less dangerous violations (lower miles-per-hour and fine amounts). According to the test, this pattern could not be observed if both officer genders were unbiased. If there was no bias, male police officers should have been *more* likely to ticket female drivers by virtue of being tougher (using a lower threshold).

Using the empirical results and some additional assumptions, I estimate the quantitative impact of the gender bias. In particular, I assume that driver behavior would not respond to the changes in violation thresholds which provide the thought experiment for a back of the envelope calculation. Supposing male police are biased while

[5]As will be explained in Section 2, the model actually suggests several ways of ranking violation thresholds. All of these produce the same ranking.

female police are not, my calculation implies that 1,902 tickets (1.3 percent of the total), would need to be re-allocated from male to female drivers to correct the gender bias. Alternatively, if female police are biased while male police are not, a similar calculation implies that only 136 tickets should be re-allocated to males from females.

After an article in the Boston Globe documented sizable racial and gender disparities in traffic tickets (Dedman and Latour 2003), the state of Massachusetts sponsored a follow-up study.[6] This study finds that males were ticketed in excess of a benchmark population, such as the share of males in the local driving population, throughout Massachusetts (Farrell et al. 2004). Perhaps in response to these findings, the Boston Police Department acted to limit police discretion in ticketing (Dedman 2004). My back of the envelope calculations suggest that at least with respect to gender, most of the disparity in tickets in Boston seems to result from gender differences in driving behavior.

Many studies of discrimination estimate how an outcome for subjects of a given racial or gender group depends on the racial or gender group of the evaluators who decide the outcome.[7] The idea underlying the estimation of these "cross-gender" or "cross-race" effects is that any dependence of the outcome on the subject-evaluator pairing of groups is difficult to reconcile as resulting from an important omitted variable or statistical discrimination. It remains difficult, however, to determine whether a cross effect *implies* bias. My analysis shows that cross effects can be generated when evaluators are unbiased but use different standards, and my test is one potential solution to the problem of drawing an inference about bias based on the estimation of a cross effect.

2.1.1 Recent Related Literature

Makowsky and Stratmann (2008), Blalock et al. (2007), and Rowe (2009) find that male drivers in Massachusetts are more likely to receive a ticket after being stopped by the police, even after accounting for many relevant controls. These results only confirm that in the benchmark population of stopped drivers, males are more likely to be ticketed.

Bagues and Esteve-Volart (2007) find that female candidates are more likely to

[6]The article gives the example of a 23 year old female college student who was pulled over four times in a three week period and never received a ticket. In the data used in this paper, containing records of all traffic citations in Boston from April 2001 to January 2003, male drivers received 71% of the citations.

[7]Recent examples include Antonovics and Knight (2007), Bagues and Esteve-Volart (2007), Price and Wolfers (2007), and Schanzenbach (2005).

pass the public examination for a position with the Corps of the Spanish Judiciary when the share of males on the evaluation committee is larger. They argue that this cross-gender effect suggests that committees are gender biased. Price and Wolfers (2007) find that black basketball players have more fouls called against them when the referees are white. They conclude that racial bias is the most plausible explanation for this cross-race effect after systematically ruling out several alternative explanations. My test offers an additional approach for interpreting the cross effects in these two studies, which I discuss in Section 4.

Broadly speaking, the literature on testing for racial bias in motor vehicle searches attempts to solve two critical problems which arise in the searches context.[8] First, omitted variables which are correlated with driver race could lead to incorrect findings of bias. Second, the researcher is unable to identify the least suspicious drivers who the police found worthy of searching (the marginal motorists). In the context of testing for bias in traffic ticketing, analogous problems appear, and my test is a potential solution. The model I develop allows for unobserved violation severity to affect officer's decisions, and the test does not require knowledge of the marginal violator.

The test I develop exploits a situation in which male and female police are unbiased yet have different costs of ticketing, and therefore use different thresholds. If police of different racial groups have different costs of search on average (i.e., one racial group of officers is more likely to search all racial groups of drivers), the test developed by Anwar and Fang (2006) has zero power to detect relative racial bias.[9] My test is able to detect relative gender bias (one group is more biased than the other) when officer's ticketing costs are different.

While my test exploits a difference in ticketing costs, the test developed by Antonovics and Knight (2007) requires the researcher to control for average differences in search costs by officer race. Also, their test requires conditioning on a qualified pool of drivers who are at risk of a search. In Section 4, I conduct tests for gender bias in ticketing which are analogous to those of Anwar and Fang (2006) and Antonovics and Knight (2007). These tests produce different results than my test, and I examine the reasons why.

[8]Knowles, Persico, and Todd (2001) developed the "hit rate" test for racial bias in searches of stopped motorists for drugs. Dharmapala and Ross (2004) analyze the hit rate test when some fraction of motorists always carry drugs. Anwar and Fang (2006) develop a test for relative racial prejudice based on ranking search rates and hit rates by officer race. Antonovics and Knight (2007) construct a test based on officer heterogeneity and the mismatch of officer and driver race.

[9]Anwar and Fang (2006) explain this in footnote 35 of their paper.

2.2 The Model

This model explains why the police choose to ticket some drivers who break the law but not others, which is a new application for a model of police behavior. After developing the model, I derive a testable implication of gender biased traffic ticketing.

2.2.1 Model set-up

The police patrol the roads and observe traffic violations committed by drivers, such as running a red light, driving faster than the speed limit, or changing lanes without signaling. The police have full knowledge of traffic laws and they know with certainty when a traffic law has been violated. A key aspect of traffic law enforcement is police discretion, because many observed violations are not ticketed. For example, according to the 2005 Police-Public Contact Survey, only 57.4% of all stopped drivers received a ticket (Durose et al. 2007). Also, from April to May of 2001, only 49 percent of Boston drivers who were stopped (and received written documentation) received a ticket as opposed to a warning. This model assumes that police discretion in ticketing operates by officers evaluating the severity, or the danger imposed on others, of the traffic violations they observe.[10]

The police officer observes the severity, $\theta \in (0, \infty)$, of each traffic violation, but θ is not observed by the researcher. The severity or danger level of a traffic violation depends on the speed of the motorist, the amount of traffic, the weather and road conditions, the presence of pedestrians, and other factors which may not be observed. All of this relevant information is summarized by θ. Police receive a benefit $b(\theta)$ from ticketing a violation of severity θ, with $\frac{\partial b(\theta)}{\partial \theta} > 0$ because officers are concerned about public safety. By ticketing a violator, officers incur a cost $t(d_g, p_g)$, which is allowed to depend on both the gender $g \in \{m, f\}$ of the driver d_g and the gender of the police officer p_g. Officers incur this cost because issuing a ticket requires labor effort in the form of stopping the driver, checking his license and registration, and dealing with any objections raised by the driver.

Definition of Bias. *A police officer of gender p_g is biased if $t(d_m, p_g) \neq t(d_f, p_g)$.*

[10] Rowe (2009) offers a rationalization for the existence of warnings (where the stopped driver receives no fine) in an efficient enforcement scheme, based on the idea that traffic stops act to detect other crimes. The model he uses does not explain how the police choose which stopped drivers to ticket. However, in that setting more dangerous offenses should be ticketed with a higher probability. This is consistent with the model developed here, which does explain officer's choices of warnings versus tickets.

This defines bias as taste-based discrimination, as originally described by Becker (1957). For instance, if an officer's cost of ticketing males is lower, for equally dangerous violations the officer will derive more utility from ticketing males. Since the utility from not giving a ticket is zero, officers use the following decision rule:

Ticketing Rule. *Officers ticket an observed violation if* $b(\theta) - t(d_g, p_g) \geq 0$.

The ticketing rule generates the following result:

Proposition 2.1. *Police officers ticket an observed violation only if* $\theta \geq \theta^*(d_g, p_g)$, *where the threshold violation* $\theta^*(d_g, p_g)$ *is determined by* $b(\theta^*) = t(d_g, p_g)$. *The threshold* $\theta^*(d_g, p_g)$ *increases monotonically as the ticketing cost* $t(d_g, p_g)$ *increases.*

Proposition (2.1) follows directly from the ticketing rule and the monotonicity of $b(\theta)$. The result says that if an officer is biased, he will find it optimal to use a different threshold θ^* for each gender of driver. For example, if it is more costly for a male police officer to ticket female drivers, he will set a higher threshold violation for ticketing females than males. For any severity $\tilde{\theta}$ where $\theta^*(d_m, p_m) < \tilde{\theta} < \theta^*(d_f, p_m)$, male police who are biased against males will ticket male drivers but not female drivers.

Define $F_g\{\theta\}$ as the distribution of violation severity θ among drivers of gender g, and $f_g(\theta)$ as the corresponding density function. Because $\theta \in (0, \infty)$ these functions are defined only for the population of drivers who violate traffic laws ($\theta > 0$). Think of violation severity as the external harm imposed by a violation, so under this formulation all violations impose positive harm. Therefore, $1 - F_g\{\tilde{\theta}\}$ represents the probability that a violation committed by a gender g driver is more harmful or dangerous than $\tilde{\theta}$.

2.2.2 Linking the model to the data

An important but unobserved quantity of interest is the probability that a driver receives a ticket, conditional on committing a traffic violation that is observed by a police officer. When a driver commits such a violation, I will say she is at risk of being ticketed.

Define the binary random variable $Ticket \in \{T, NT\}$ for whether a driver at risk is stopped and given a ticket (T) or not ticketed (NT). The probability of an at risk driver being ticketed by a police officer is then:

$$P(T \mid d_g, p_g) = 1 - F_g\{\theta^*(d_g, p_g)\} \tag{2.1}$$

The proportion of drivers from each gender group who are stopped and ticketed by p_g officers after committing a violation is simply the proportion whose violations were dangerous enough to exceed the officer's ticketing threshold, $\theta^*(d_g, p_g)$. The distribution of violation severity $F_g\{\theta\}$ is taken as exogenous. We can think of drivers having chosen how badly to violate traffic laws, taking as given the expected fine for committing various offenses.[11] Also implicit in this formulation is the idea that drivers don't know when they are being monitored by police, so they behave the same whether the police are observing them or not.

For officers of gender p_g, the odds of a female driver being ticketed conditional on committing a violation, referred to as the ticketing odds, is then:

$$Odds(p_g) = \frac{P(T \mid d_f, p_g)}{P(T \mid d_m, p_g)} = \frac{1 - F_f\{\theta^*(d_f, p_g)\}}{1 - F_m\{\theta^*(d_m, p_g)\}} \tag{2.2}$$

Notice that $Odds(p_g)$ may not be equal to 1 even if the police are unbiased. In the model, unbiased officers do not statistically discriminate by *ex ante* choosing ticketing probabilities based on driver gender. Rather, an unbiased officer will be more likely to ticket at risk male drivers if males tend to commit more dangerous violations. Such a systematic difference between male and female drivers could explain why males are ticketed in excess of feasible benchmarks such as their share of the local driving population.

The absolute ticketing odds $\frac{P(T|d_f)}{P(T|d_m)}$ cannot be identified in the data because the pool of drivers who commit violations is not observed. However, the empirical section shows that it is possible to determine if the ticketing odds are different for male and female police officers. The model is then linked to the data because equation (2.2) shows how these ticketing odds are produced for each officer gender.

2.2.3 The test for gender bias

Section 3 presents evidence that the ticketing odds for male officers, $Odds(p_m)$, is different from the ticketing odds for female officers, $Odds(p_f)$. Yet this finding says nothing about whether the police are biased. In particular, police officers might use different but unbiased ticketing thresholds, such as $\theta^*(d_m, p_m) = \theta^*(d_f, p_m) < \theta^*(d_m, p_f) = \theta^*(d_f, p_f)$. In this situation, equation (2.2) indicates that the ticketing odds may vary by officer gender even though there is no bias. To determine if an observed officer gender difference in the ticketing odds is consistent with unbiased

[11]The expected fine for an offense is determined by the statutory fine, the probability of being monitored by police, the thresholds used by male and female officers for each driver gender, and the probability of being monitored by a male or female officer.

policing, a prediction of how $Odds(p_g)$ should vary across unbiased male and female police is needed. To obtain this prediction I make the following assumption:

MLRP Assumption. *The density functions $f_m(\theta)$ and $f_f(\theta)$ satisfy the Monotone Likelihood Ratio Property, so that if $\theta_1 > \theta_0$ then $\frac{f_m(\theta_1)}{f_f(\theta_1)} > \frac{f_m(\theta_0)}{f_f(\theta_0)}$.*

The MLRP assumption is a way of formalizing the idea that men are more dangerous drivers than women. The MLRP implies that males are always more likely than females to commit a violation with severity or danger level above a given threshold, so that $1 - F_m(\tilde{\theta}) > 1 - F_f(\tilde{\theta})$ for all $\tilde{\theta}$. What confidence can we have in this assumption?

Figure (2.1) shows the empirical cumulative distribution functions of miles-per-hour over the speed limit for male and female ticketed drivers in the Boston data. To the extent that faster violations are more dangerous, the empirical distribution functions are consistent with the implication of the MLRP: $F_m(\text{MPH}) < F_f(\text{MPH})$.

Blackmon and Zeckhauser (1991) document adverse consequences in the automobile insurance market in Massachusetts after the state banned insurers from basing premiums on gender (and restricted the ways insurers could base premiums on age) in 1977. For instance, many insurers decided to no longer write policies. Levitt and Porter (2001), using national FARS data, find that the fatal two-car crash risk for men is three times larger than the same risk for women. Much of this effect is due to higher rates of drunk driving among males. Edlin and Karaca-Mandic (2006) find that various measures of automobile insurance costs and premiums increase as the percentage of young males in the population increases. All of this evidence supports the general idea that men are more dangerous drivers. Furthermore, Section 4 presents more specific evidence from the Boston data which supports the implication of the MLRP that more female drivers should be found at less serious violation levels. The MLRP makes it possible to predict how $Odds(p_g)$ will vary across unbiased police who use different thresholds:

Proposition 2.2. *If police officers are unbiased, but $\theta^*_{p_m} \neq \theta^*_{p_f}$, then $\theta^*_{p_m} > \theta^*_{p_f}$ implies $Odds(p_m) < Odds(p_f)$, and $\theta^*_{p_m} < \theta^*_{p_f}$ implies $Odds(p_m) > Odds(p_f)$.*

The result holds because it can be shown (see the Appendix) that:

$$\frac{\partial Odds(p_g)}{\partial \theta^*} = \frac{\int_{\theta^*}^{\infty} f_m(\theta^*) f_f(\theta) - f_f(\theta^*) f_m(\theta) d\theta}{(1 - F_m\{\theta^*\})^2} < 0 \tag{2.3}$$

Proposition (2.2) says that if police are unbiased but use different thresholds, the officer gender which sets a higher threshold will have a lower odds of ticketing female drivers. This follows from the MLRP, which implies that as an unbiased threshold θ^* increases, female drivers are relatively less likely to commit a violation above it.

Officer's violation thresholds are not observed, but the idea of Proposition (2.2) can be tested empirically with only a ranking of officer's violation thresholds. What is required is a reasonable way to rank officer's violation thresholds based on the available data. The first step is to notice that violation thresholds are linked to the average severity of ticketed violations in the following way:

Proposition 2.3. *The average severity of violations $\bar{\theta}(d_g, p_g)$ among drivers of gender d_g ticketed by officers of gender p_g increases monotonically as the violation threshold $\theta^*(d_g, p_g)$ increases. Therefore if $\theta^*(d_g, p_m) > \theta^*(d_g, p_f)$, then $\bar{\theta}(d_g, p_m) > \bar{\theta}(d_g, p_f)$. Likewise, if $\theta^*(d_g, p_m) < \theta^*(d_g, p_f)$, then $\bar{\theta}(d_g, p_m) < \bar{\theta}(d_g, p_f)$.*

The derivation is shown in the Appendix. The result says that the officer gender which uses a higher threshold for drivers of gender d_g will write tickets to those drivers for more dangerous violations on average. Consider the case when officers are unbiased but use different thresholds. Proposition (2.3) then says that the average severity of violations for both male and female drivers ticketed by the high threshold officers will be higher than the corresponding averages for the low threshold officers.[12]

Violation severity θ is unobserved for individual tickets, but Proposition (2.3) is in terms of the average severity $\bar{\theta}$ of ticketed violations. When averaging over tickets issued by male and female officers, it is reasonable to infer that a difference in average miles-per-hour (or average fine amount) represents a difference in average violation severity. In terms of speeding tickets written by male versus female officers, the required assumption is:

Average Severity Assumption. *If $\overline{mph}(p_m) > \overline{mph}(p_f)$, then $\bar{\theta}(p_m) > \bar{\theta}(p_f)$. Likewise, if $\overline{mph}(p_m) < \overline{mph}(p_f)$, then $\bar{\theta}(p_m) < \bar{\theta}(p_f)$.*

This guarantees that the rank order of average miles-per-hour (or fine amount) preserves the rank order of average violation severity. In other words, higher average miles-per-hour over the limit implies a higher average violation severity. This is reasonable in light of the fact that dollar fine amounts (which increase with miles-per-hour) are chosen by policy-makers so that more dangerous violations are punished with higher fines.

[12]Using Proposition (2.3), a test analogous to that proposed by Anwar and Fang (2006) can be derived. This is discussed in Section 4.

The last step is to link empirical rank orders of average miles-per-hour or average fine amounts (which are assumed to preserve the rank order of average violation severity) for tickets written by male and female officers to a ranking of ticketing thresholds. This could be done by estimating these four sample averages: $\overline{mph}(d_m, p_m)$, $\overline{mph}(d_m, p_f)$, $\overline{mph}(d_f, p_m)$, and $\overline{mph}(d_f, p_f)$. Using these, we could refer to Proposition (2.3) to rank the officer's thresholds. A limitation of this approach is that the four averages cannot be computed in a parametric specification (such as OLS) for miles-per-hour over the limit when a constant term is included.[13] To adjust for differences by officer gender in the pools of drivers at risk, a re-sampling procedure similar to that of Anwar and Fang (2006) could be used when computing the four sample averages. However, their procedure only corrects for geographic differences in the pools of drivers at risk.

An advantage to parametrically estimating how the average miles-per-hour (or fine amount) varies by officer gender is that a large number of relevant variables can easily be included. Variables such as time of day, day of week, and the speed limit might all help to control for differences in the pools of drivers at risk which are observed by male and female officers. The following result clarifies how the coefficient on officer gender in an OLS specification for miles-per-hour over the limit allows us to determine which officer gender uses a lower threshold:

Proposition 2.4. *If the average violation committed by male drivers is at least as dangerous as the average violation committed by female drivers, then the average severity of ticketed violations for a given threshold θ^*, $E[\theta \mid T, \theta^*]$, increases monotonically as the violation threshold θ^* increases. Using the average severity assumption, we then know that: If $\theta^*(p_m) > \theta^*(p_f)$, then $E[mph \mid T, \theta^*(p_m)] > E[mph \mid T, \theta^*(p_f)]$. Likewise, if $\theta^*(p_m) < \theta^*(p_f)$, then $E[mph \mid T, \theta^*(p_m)] < E[mph \mid T, \theta^*(p_f)]$.*

See the Appendix for the derivation. Proposition (2.4) confirms the intuition that because faster violations are more dangerous, an officer who uses a relatively high violation threshold will end up ticketing relatively faster drivers. For this intuition to hold, on average male driver's violations must be at least as dangerous as those of females.[14] This condition is clearly supported by the evidence discussed earlier, which indicates that men are actually more dangerous drivers. Proposition (2.4) shows that

[13] When including a constant, one can only estimate three related quantities parametrically: How the average miles-per-hour depends on driver gender, officer gender, and their interaction.

[14] In fact, this is implied by the MLRP if we define θ on the interval $[0, \infty)$ and assume that $F_m\{0\} = F_f\{0\} = 0$. This setup also provides the intuitive condition that all violations impose positive harm.

we can rank officer's violation thresholds by the average miles-per-hour over the limit of the speeding violations they ticketed. To account for possible differences in the pools of drivers observed by male and female officers, we can condition on many observed characteristics (denoted by X) of the traffic citations. Using this way of ranking thresholds, my "composition test" for gender bias is:

Composition Test. *At least one police officer gender is biased if:*
$E[mph \mid X, T, p_m] < E[mph \mid X, T, p_f]$ *and* $Odds(p_m) < Odds(p_f)$*, or if:*
$E[mph \mid X, T, p_m] > E[mph \mid X, T, p_f]$ *and* $Odds(p_m) > Odds(p_f)$*.*

The test is easiest to interpret when the direction of both effects are significant in the statistical sense and we therefore conclude that bias exists. What can be said in other cases depends on the situation. For instance, if there is a small and statistically insignificant difference between $E[mph \mid X, T, p_m]$ and $E[mph \mid X, T, p_f]$, it would be logical to conclude that there is no difference in ticketing costs. In that case, Anwar and Fang's (2006) test would have positive power and could be used instead of the composition test. On the other hand, if there was a large difference in ticketing costs but no statistically significant difference in the gender mix of ticketed drivers, depending on the sign of the officer gender effect the composition test might still suggests that a bias exists. This is because if one officer gender is a great deal tougher, in the absence of bias it would be likely that the tougher officers would ticket many more females.

The test is based on the model's prediction that if there is no bias but officers use different thresholds, the officer gender which is more likely to ticket female drivers (conditional on the drivers being at risk) should also issue tickets for relatively less dangerous violations. Although this prediction was obtained with the help of some technical assumptions, the idea is intuitive. If females really are safer drivers, relatively few females should commit traffic violations dangerous enough to exceed a high threshold. Officers who use a high threshold must observe a relatively dangerous violation in order to issue a ticket, and therefore should ticket faster violations on average.

More generally, the model implies that the gender composition of ticketed drivers and the severity of ticketed violations are linked. Section 4 presents specific evidence from the Boston data which supports this theoretical link, and discusses how the existence of analogous links might be detected in other contexts. To apply the test to traffic violations for offenses other than speeding, the dollar amount of the ticket can be substituted for miles-per-hour over the limit in order to rank the violation thresholds. Here the idea is that violations which are punished with higher fines are more dangerous.

2.3 Estimation of the Officer Gender Effect

To implement the composition test, we must first estimate how the probability of a female driver receiving a ticket, conditional on committing a traffic violation, depends on the gender of the officer who observes the violation. This is the officer gender effect. The identification problem is to estimate the officer gender effect even though only the drivers who received tickets are observed in the data from Boston.

2.3.1 Data

The data I use comes from three sources.[15] The first source is a file containing information on characteristics of the driver and the traffic stop for traffic tickets issued in the city of Boston, Massachusetts, from April 2001 through January 2003. The second data source contains information on all stopped drivers in Massachusetts who received written documentation in the form of a ticket or a written warning, but only from April to May of 2001. I use records in this file from Boston to conduct a test analogous to Antonovics and Knight (2007).

The third data source is a file containing demographic information such as race, gender, and year of entry into the police force for the police officers in Boston. This officer-level data was merged in to the tickets data using the officer's identification number, which is present in both files. The merge successfully assigned officer-level data to 95 percent of the original 184,463 observations in the tickets file, leaving 175,021 observations in the merged file.[16] A similar merge was performed on the Boston tickets and warnings file. Finally, a comparison of the merged Boston tickets file to the merged Boston tickets and warnings file revealed duplicated observations in April and May of 2001 in the tickets file. By dropping observations with invalid fine

[15] I thank Kate Antonovics, Bill Dedman, and Nicola Persico for sharing these data sources with me.

[16] In some cases, the gender of the police officer was missing. When possible, I re-coded gender for these cases using the officer's first name, if the name was unambiguously a male or female name. Before re-coding, 149 officers who appear in the merged file (accounting for 12.1 percent of the citations) had missing officer gender, 143 officers were coded as female (accounting for 3.0 percent of citations), and 1,149 officers were male. After re-coding, 19 officers (accounting for 3.1 percent of citations) remain with missing officer gender, 179 officers are female (accounting for 3.9 percent of citations), and 1,243 officers are male. In addition, 2.8 percent (4,848 observations) of records in the merged file had missing information on driver gender, so these cases are not used in the regression analyses.

amount information for these two months in the tickets file, the number of citations issued in April and May of 2001 is consistent across the two files: 9,252 in the ticket and warning file compared to 9,396 in the tickets file.[17] There is no concern that warnings may also be present in the remaining months of the tickets file, as the information on warnings was not collected after May 2001 (Dedman 2003).

Table (2.1) shows sample means for some relevant variables calculated from the merged file, split up for tickets issued by male and female police officers. Overall, about 71% of ticketed drivers are males. Compared to male officers, female officers ticketed slightly more female drivers, issued more tickets during daylight hours, and issued fewer tickets for seat belt violations. In addition, female officers wrote fewer tickets (on days when they wrote at least one ticket), issued speeding tickets for higher miles-per-hour over the limit, and wrote non-speeding tickets for higher fine amounts. I will argue that the best explanation for these three facts is that female officers are not as "tough" (they use a higher threshold) in their enforcement of traffic laws.

2.3.2 Methodology

Let d_m and d_f denote the random variables that a male or female driver commits a traffic violation that is observed by a police officer. When a driver commits such a violation, I say she is at risk of being ticketed. A simple way to estimate the gender disparity in traffic tickets would be to compare the probability of a female driver at risk receiving a ticket (represented by T) to the probability of a male driver at risk receiving a ticket:

$$\text{Gender disparity} = P(T \mid d_f) - P(T \mid d_m) \tag{2.4}$$

The quantities in equation (2.4) cannot be calculated because the pool of drivers at risk of receiving a ticket is not known. Conditioning on the pool of drivers stopped by police will not enable calculation of (2.4) either, because the police do not stop each violator they observe, and do not give written documentation to all stopped drivers who are not ticketed. Indeed, the pool of drivers at risk can only be known to a researcher if data for all traffic law violators observed by a police officer were systematically recorded. However, the available traffic ticket data records the gender

[17] Before dropping these observations, April and May 2001 contained substantially more observations than the other months in the tickets file. All empirical results are similar if this issue is ignored. Results are also similar if the observations on tickets in the ticket/warning file are used in place of the observations for April and May 2001 in the tickets file, or if the April and May 2001 observations are dropped.

of the drivers who received tickets, so it is possible to calculate $P(d_m \mid T)$ and $P(d_f \mid T)$. Using Bayes' rule, we obtain that:

$$\frac{P(d_f \mid T)}{P(d_m \mid T)} = \frac{P(T \mid d_f)}{P(T \mid d_m)} * \frac{P(d_f)}{P(d_m)} \tag{2.5}$$

Equation (2.5) shows formally why it is not possible to tell if the gender disparity in tickets results because female drivers are less likely to receive a ticket after a violation (the first term on the right hand side) or because females are less likely to commit violations in the first place (the second term).

Again, p_f denotes the event that a female officer observes a traffic violation, while p_m represents the same for male officers. Define the empirical odds for female driver conditional on being ticketed by an officer of gender p_g as:

$$EOdds(p_g) = \frac{P(d_f \mid T, p_g)}{P(d_m \mid T, p_g)} \tag{2.6}$$

Recall equation (2.2), derived in the previous section, which shows the odds $Odds(p_g)$ of female drivers being ticketed by officers of gender p_g after committing a violation.[18] Refer to $Odds(p_g)$ as the ticketing odds. Forming equation (2.5) for both male and female officers and dividing gives:

$$\frac{EOdds(p_m)}{EOdds(p_f)} = \frac{Odds(p_m)}{Odds(p_f)} \times \frac{P(d_m \mid p_f)}{P(d_f \mid p_f)} \frac{P(d_f \mid p_m)}{P(d_m \mid p_m)} \tag{2.7}$$

The last term on the right hand side of (2.7) will be equal to 1 if the odds of a female driver committing a violation is independent of whether drivers are observed by male or female police officers. Therefore, comparing the empirical odds $EOdds$ across male and female police identifies how the ticketing odds $Odds$ depend on officer gender if the police officers observed the same pool of drivers.[19]

This discussion suggests that a natural way to estimate the officer gender effect is to estimate a logit model for the empirical odds that a ticketed driver is female, using an indicator for male officer as an explanatory variable. To account for possible differences in the pools of drivers observed by male and female police, I also include day of week, time of day, speed limit of road, and indicators for the geographic districts of the Boston Police Department as explanatory variables. The logit model I estimate is:

[18] $Odds(p_g) = \frac{P(T|d_f,p_g)}{P(T|d_m,p_g)} = \frac{1-F_f\{\theta^*(d_f,p_g)\}}{1-F_m\{\theta^*(d_m,p_g)\}}$.

[19] This reasoning is similar to that of Grogger and Ridgeway (2006), who estimate how the odds that a stopped driver belongs a racial minority group depends on whether the stop occurred in daylight.

$$\ln\left(\frac{P(d_f \mid T)}{1 - P(d_f \mid T)}\right) = \beta_0 + \beta_1(\text{Male Officer}) + \beta_2(\text{Controls}) \tag{2.8}$$

The coefficient β_1 on male police officer will show the effect of officer gender on *EOdds*.[20] If officers observed the same pool of drivers conditional on the controls, then β_1 captures the officer gender effect; how the odds of a female driver receiving a ticket conditional on being at risk depends on officer gender.

One concern about this identification strategy is that drivers might adjust their behavior in response to the gender composition of police officers in a given location. Such strategic driving may be plausible with respect to race. The areas of Boston which have greater proportions of minority residents also have greater proportions of minority police (Antonovics and Knight 2007). Thus when a minority driver travels into a predominantly white neighborhood, he can infer that he is more likely to be observed by a white police officer, and therefore might adjust his driving behavior.

However, strategic driving with respect to gender seems less plausible. As Table (2.2) shows, female police officers are distributed fairly evenly across the Boston police districts, and are never more than 20% of the force in any district. Throughout the city of Boston, the chance of being observed by a female officer is roughly uniform, which should make strategic driving simply not worthwhile.

In addition, since drivers cannot observe the gender of individual police officers *before* they decide whether to break traffic laws, they cannot respond directly to officer gender. A direct behavioral response to gender or race is more likely in other settings, such as Price and Wolfers (2007) and Bagues and Esteve-Volart (2007), in which gender (or race) is randomly assigned but is visible to all participants before behavioral choices are made.[21]

Another potential problem is that female and male police officers may have different job functions, which might somehow cause female police to observe a different pool of drivers. Female officers issued fewer traffic citations (see Table 2.1), which might indicate that job functions vary by officer gender. Yet if female officers simply spend less time monitoring the roads than male officers, this does not invalidate the empirical strategy. For example, female officers were less likely to work at night, but conditioning on the time of the traffic ticket will account for how nighttime drivers

[20] This is because $\frac{P(d_f|T)}{P(d_m|T)} = \frac{P(d_f|T)}{1-P(d_f|T)}$.

[21] In Price and Wolfers (2007), the NBA scheduling process guarantees that the racial makeup of the refereeing crew is unrelated to the racial makeup of the teams. For Bagues and Esteve-Volart (2007), the Spanish government assigns candidates to committees without regard to the gender makeup of the committee.

are different. In general, the inclusion of day of week and time of day controls, as well as day and time interaction terms, will account for systematic differences in driver behavior during the times that male and female police are engaged in monitoring traffic.

Table (2.3) shows means of several work-related variables from the 2000 Census for male and female police in the Boston metropolitan area. Clearly, male officers spend significantly more time on the job. Importantly, the higher hourly wage seen for male officers can be mostly explained by the higher pay rate received for overtime, along with a smaller contribution due to the higher rate of college degree attainment for male officers. In the Boston Police Department, a college degree guarantees a 20% bonus over the base pay rate, while police receive 1.5 times their base pay rate for overtime, which is hours worked in excess of 40 hours per week. Thus, the most prominent difference in the Census data between male and female police officers is the number of hours worked per year, a difference which the empirical strategy is able to account for.

Figure (2.2) displays the time pattern of total citations by officer gender. Although fewer citations are written by female police, the timing of the monthly fluctuations in citations matches up fairly well. This indicates that male and female officers are subjected to the same shifts in policing activity and driver behavior which account for the monthly changes in traffic tickets.

In their recruiting efforts, the Boston Police Department states that female officers are not pushed into systematically different or less desirable jobs than male officers.[22] Furthermore, in correspondence with the author, the Boston Police Department stated that: "Both male and female officers perform the same functions within the Department."

2.3.3 Results

Table (2.4) shows basic OLS and logit estimates for the effect of male officer on the probability that the ticketed driver is female, with the sample split into speeding tickets and tickets for other types of violations. In the logit specification for speeding tickets with no controls, the odds ratio coefficient on male police officer is not statistically different from 1. When miles-per-hour over the speed limit is included, the ticketed driver is less likely to be female if the officer is male because the coefficient on

[22] From the Boston P.D.'s Women in Policing web page: "Gone are the days of women serving solely in an administrative capacity or in positions deemed more suitable for women. Today we serve on the front lines. Women on the job serve in various capacities such as patrol officers, criminal investigators, motorcycle officers, and hostage negotiators."

male officer is less than 1 (0.808), an effect which is significant at the 1% level (s.e. = 0.057). The same pattern is observed in the OLS specifications for speeding tickets, which also provide a sense of the magnitude of the officer gender effect in terms of probabilities. If the police officer is male, the ticketed driver is about 5 percentage points less likely to be female, or equivalently 16 percent less likely to be female given that 30 percent of ticketed drivers are women.

The upward OLS bias towards zero results because miles-per-hour over the limit is negatively related to both female driver and to male police officer. Thus, the observed OLS bias when miles-per-hour is omitted suggests two key observations. First, male police officers give tickets at lower values of miles-per-hour, indicating that they may use a lower threshold than female officers. Second, relatively fewer female drivers are found at higher miles-per-hour over the limit, which is consistent with the MLRP assumption of the model.

For tickets issued for other traffic violations, such as failure to stop, no seat belt, or expired inspection sticker, the results in Table (2.4) show that even when no control variables are used, male police officers ticketed relatively fewer female drivers, an effect which is significant at the 1% level. The magnitude of the effect is a bit smaller than that observed for speeding tickets; the ticketed driver is 2.6 percentage points less likely to be female if the officer was male.

To link these results back to equation (2.7), note that because the odds ratio coefficient on male officer for speeding tickets is 0.808, this means $EOdds(p_m) = 0.808 * EOdds(p_f)$. If male and female police officers observed the same pool of drivers, this would reflect how the odds of being ticketed conditional on being at risk depend on officer gender, so the result would imply that $Odds(p_m) = 0.808 * Odds(p_f)$.

Linking the empirical odds directly to the ticketing odds requires assuming that male and female officers observed the same pool of drivers, conditional on the set of observable characteristics of each traffic ticket. For this reason, I include a rich set of control variables: The speed limit of the road, the driver's race and age, driver's age squared, whether the driver was from Boston (in-town), day of week dummies, weekend night and workday commute dummies, time of day dummies (pre-dawn, morning, afternoon, and evening), and dummies for the officer's geographic district of the Boston Police Department. In addition, specifications including interactions of all day of week and time of day dummies were estimated (these specifications do not include the workday commute and weekend night dummies). If the estimated male officer effects in Table (2.4) were due to male officers observing a pool of drivers which systematically differed by these observable characteristics, including such controls would tend to push the male officer effects towards zero.

The results in Table (2.5) show that for both categories of tickets, the negative effect of male officer on the odds of the ticketed driver being female becomes only slightly smaller as controls are included. In the specification for speeding tickets including the full set of controls, the odds that the driver is female falls by a factor of 0.845 (significant at the 5% level with s.e.=0.060) if the police officer is male. This estimate is within one standard error of the corresponding estimate (0.808) in Table (2.4) where the *only* control is miles-per-hour over the limit. The same pattern is seen in the results for other types of violations. This indicates that very little of the male officer effect is attributable to the effects of the control variables on the gender makeup of ticketed drivers.

To evaluate the robustness of these results, I estimated several alternative specifications. First, the log of the total number of traffic citations by month and Boston Police district was included as a control in both an OLS and a Logit specification. The number of tickets issued results from the interaction of driver behavior with enforcement intensity, so periods with high numbers of tickets issued must differ by at least one of these factors. Second, instead of the number of tickets by month and district I used the number of tickets by officer and day. The coefficient on this ticket variable will show how the gender mix of ticketed drivers depends on how many tickets the officer wrote that day. Third, I included unrestricted dummies for each month in the data as controls, as Figure (2.2) showed significant variation in total citations in Boston over time. These dummies net out the impact of monthly changes in the interaction of enforcement and driver behavior which drive the shifts in tickets, even though this is not necessarily desirable. For instance, if directed to write more traffic tickets, gender-biased officers might respond by lowering their ticketing threshold for only one gender of drivers. The robustness specifications are estimated with day and time dummies, Boston district dummies, and controls for driver demographics.

The results of these robustness checks are shown in Table (2.6). Consistently across the different specifications, the number of tickets issued has an impact on the gender mix of ticketed drivers. Adding up tickets issued by either month and district or officer and day, when more tickets (either for speeding or for other violations) were issued the ticketed driver was more likely to be female. The coefficients on male officer in these specifications are quite similar to those reported in Table (2.5). The only specification where the effect of male officer is attenuated is the specification for speeding tickets with unrestricted month dummies, and even here the effect (odds ratio of 0.882) is only slightly smaller than in the baseline results and is still statistically significant at the 10% level.

To summarize, the empirical results confirm that male officers ticketed relatively fewer female drivers than female officers. The available evidence related to the activities of male and female officers, together with the extensive controls to account for many factors which might plausibly affect the gender mix of drivers on the roads, suggest that this effect does not result because male and female police observed systematically different pools of drivers. I therefore conclude that female drivers are less likely to be ticketed, conditional on committing a traffic violation, if they are observed by a male police officer.

2.4 Application of the composition test

The next step in conducting the test is to rank the officer's violation thresholds by determining which officer gender must observe a more dangerous violation before deciding to issue a ticket. By examining the sample means in Table (2.1) and referring to Proposition (2.4), we could conclude that male officers use a lower ticketing threshold (they are "tough") because they issued tickets for lower fine amounts and lower miles-per-hour over the limit. This conclusion will be strengthened if it still holds when extensive controls are used to account for possible differences in the pools of drivers at risk.

For speeding tickets, I estimate OLS specifications for miles-per-hour over the limit as a function of officer gender, the control variables used to estimate the officer-gender effect in Table (2.5), and two additional controls: The gender of the driver and an interaction of driver and officer gender. According to Proposition (2.4), driver gender can be omitted in order to rank officer's thresholds, and specifications omitting driver gender (two are shown in Table (2.8), and others are available by request) produce the same rankings.[23] I include driver gender and the interaction of driver and officer gender (called "gender mismatch") for comparison to Antonovics and Knight (2007). In their model, the coefficient on mismatch of officer and driver race in a specification for the probability of being searched captures taste-based discrimination. Therefore a statistically significant coefficient on gender mismatch might suggest that officers discriminate via the miles-per-hour they charge ticketed drivers with.

In my miles-per-hour specifications, because a dummy for male officers, a dummy for female drivers, and their interaction (gender mismatch) are explanatory vari-

[23] Specifications using the percent over the limit as the dependent variable also produce the same rankings.

ables, the gender mismatch coefficient measures the following difference-in-difference: $[\overline{mph}(d_m, p_f) - \overline{mph}(d_f, p_f)] - [\overline{mph}(d_m, p_m) - \overline{mph}(d_f, p_m)]$. Antonovics and Knight (2007) create their mismatch variable as the sum of two interaction terms: Black Officer × White Driver + White Officer × Black Driver. If created in this way, the gender mismatch coefficient would be equal to the difference-in-difference shown above divided by two. I show the derivation of this equivalence in the Appendix. In either case, the mismatch coefficient shows whether the average male driver versus female driver disparity in miles-per-hour varies by officer gender. However, the ticketing model developed in Section 2 indicates that such variation by officer gender may not be due to bias when officers use different standards.

To see this, consider a simple example in which the probability of receiving a ticket conditional on breaking the law is known, so we can directly calculate $P(T \mid d_g, p_g)$. Suppose these 4 quantities were observed: $P(T \mid d_m, p_f) = 0.3$, $P(T \mid d_f, p_f) = 0.1$, $P(T \mid d_m, p_m) = 0.5$, and $P(T \mid d_f, p_m) = 0.4$. The male driver versus female driver disparity in this example is higher by 0.1 (which would be the coefficient on gender mismatch) when the officer is female. Notice that male officers were tougher because they were always more likely to ticket violations. The tougher officers ticketed relatively more female drivers, $Odds(p_m) = \frac{0.4}{0.5} > Odds(p_f) = \frac{0.1}{0.3}$, so according to my test the example is consistent with unbiased ticketing.

I rank officer's violation thresholds for other types of traffic violations separately from speeding tickets, as there is no continuous measure which reflects the severity of violations such as "failure to stop" or "expired inspection sticker". There is information on the dollar amount of the fine, which is mostly determined by the specific offense the driver was charged with. Tickets for non-speeding violations were by far most likely to impose a fine of either \$25, \$35, or \$50.[24] Because of the discrete nature of the fine variable, I estimate ordered logit models for 5 fine amount categories: less than \$26, from \$26 to \$35, from \$36 to \$50, from \$51 to \$100, and greater than \$100. It is difficult to interpret the effect of gender mismatch on the fine in the ordered logits, so I omit it from the reported ordered logit specifications. Instead, I report additional OLS specifications for the fine amount which include gender mismatch as an explanatory variable, and describe a calculation of the effect of gender mismatch based on an ordered logit in the text. Unfortunately the fine variable is missing for about 37% of the non-speeding tickets. The probability of the fine being missing is negatively related to male officer, but the effect is small in magnitude (about 1.6

[24]Out of 68,759 non-speeding tickets with valid fine amount and gender data, 18.4% were for \$25, 15.4% for \$35, 56.5% for \$50, and 8.0% for \$100.

percentage points) so I ignore this issue here.

The results used for ranking officer's ticketing thresholds are shown in Table (2.7). On average, conditional on all controls, male police issued speeding tickets for 1.47 fewer miles-per-hour over the limit (standard error of 0.23) than the female officers. According to Proposition (2.4), we can infer that male police use a lower ticketing threshold than the female police. The gender mismatch coefficient, which captures the difference-in-difference described above, is small (-0.47 miles-per-hour) and statistically insignificant. This might suggest that to the extent officers use discretion to assign miles-per-hour over the limit, they use this discretion similarly when faced with a driver of the opposite gender.

The ordered logit and OLS results shown in Table (2.7) for other types of traffic citations again suggest that male officers were tougher. The odds of the ticket being in a higher fine category (relative to all lower categories) falls by a factor of 0.66 (s.e.= 0.026) if the officer was male, which is significant at the 5% level. Therefore, male officers were more likely to issue tickets for violations which imposed smaller fines, consistent with male officers using a lower ticketing threshold. The OLS coefficient on gender mismatch when all controls are included is equal to 2.77 dollars with a standard error of 1.31. This could be reflecting a differential use of discretion to adjust charged fines, but the effect is only about half the size of the main effect of male officer (-5.84 dollars). I also estimated the effect of gender mismatch on the fine by adding it as an explanatory variable to the ordered logit model which includes controls in Table (2.7) and calculating the predicted fine amount for each observation. I assumed the fines associated with the fine categories were the following: \$25, \$35, \$50, \$100, and \$267.[25] The average of the marginal effects of gender mismatch on the fine amount, accounting for the implied changes in the gender dummies, equals 1.08 dollars with a bootstrapped (100 replications) standard error of 0.93.

The robustness of these results was tested by estimating a number of alternative specifications. First, I included the log of total citations issued by month and police district as an additional control. Second, I used the log of total citations by officer and day as a control, and also excluded driver gender and gender mismatch as explanatory variables. Third, I put driver gender and gender mismatch back in and included unrestricted dummies for each month in the data. The results for these specifications are shown in Table (2.8). The measures of tickets written had consistent effects across all the specifications. When more tickets were issued, ticketed violations occurred at

[25] I used \$267 for the highest category because it is the mean fine amount for non-speeding violations which received fines above \$100.

lower miles-per-hour and fine amounts. As we saw in the baseline specifications, for both speeding tickets and other violations male officers tended to issue citations for relatively less serious offenses. To the extent possible with the available data it does not appear that this result occurs because male officers observed a different pool of drivers. The results therefore indicate that male officers are tougher because they are willing to write tickets for less dangerous violations.

Despite the consistent empirical pattern of male officers writing tickets for less serious violations, a relevant concern is that this may not reflect a difference in toughness but instead may result from officers adjusting miles-per-hour and fine amounts after deciding to write a ticket. Using the same Boston data as this paper, Anbarci and Lee (2008) observe that for speeding tickets, the histogram of miles-per-hour over the limit spikes at 10. They argue that this represents officer discretion in giving some motorists a "discount" on their ticket, and they find that male officers are more likely (by 33 percentage points) to write speeding tickets at exactly 10 miles-per-hour over the limit (when conditioning on the ticket being between 10 and 14 miles-per-hour over the limit).

Even accepting Anbarci and Lee's interpretation that male officers are more likely to discount miles-per-hour (this is not the main point of their paper), for several reasons I believe my results imply that male officers are tougher. First, assuming that officers randomly chose violations between 11 and 14 for discounting, Anbarci and Lee's result implies that discounting would reduce charged miles-per-hour for male officers by 0.85 miles-per-hour.[26] This cannot fully account for the male officer effect of -1.47 miles-per-hour in my baseline specifications. Second, there is no evidence of discounting for offenses other than speeding. Fine amounts for these offenses are clustered at the values of very common infractions, such as "Failure to Stop" (about 25,000 observations) which incurs a $50 fine in Massachusetts. Finally, discounting cannot explain why male officers wrote more tickets (as shown in Table 2.1). In contrast, because tough officers are willing to ticket drivers for less serious offenses, for a given mix of offenses observed a tough officer will see more that exceed his threshold and therefore issue more tickets. The three facts that male officers wrote tickets for lower miles-per-hour, lower fine amounts, and wrote more tickets (on days for which they issued at least one) are all consistent with male officers being tough.

We can now conduct the composition test by combining the conclusion that male

[26]In the Boston data, male officers wrote speeding tickets to 1,609 drivers at 11 m.p.h. over, 2,272 at 12, 2,016 at 13, and 2,162 at 14. From this, the average speed between 11 and 14 is 12.58. Male officers were more likely by a factor of 0.33 to mark the average 11 to 14 violation down to 10, so 0.33*(12.58-10)=0.8514 is the implied impact of the discounting.

officers are tougher with the empirical results of Sections 3. The results in Section 3 indicate that male police officers were relatively less likely to ticket a female driver who committed a violation. Both of these results are statistically significant at the 5% level.[27] According to the composition test, this pattern can only result if at least one gender group of officers is biased. The model implies that if there was no bias, by using a lower ticketing threshold male officers should have been *more* likely than female officers to ticket female drivers. Therefore, the null hypothesis that both officer genders are unbiased is rejected in favor of the alternative that at least one group of officers is gender biased.

Two critical assumption in the model are the MLRP and the average severity assumption. Besides suggesting the test for bias, the more general implication of these two assumptions is that there is a link between the gender composition of ticketed drivers and how fast (or expensive) ticketed violations are on average. If ticketed violations were slower on average, then relatively more ticketed drivers should be female (and vice versa). The observed effects of the changes in total tickets, added up by month and Boston police district or by officer and day, show a consistent pattern of empirical support for this link. When more tickets were issued, ticketed drivers were more likely to be female, and ticketed violations occurred at lower miles-per-hour and fine categories (see Tables 2.6 and 2.8). These effects are statistically significant in all of the relevant specifications.

There are two potential explanations for how changes in the number of tickets issued could produce this pattern. First, the police might be lowering (or raising) their ticketing thresholds in an unbiased fashion in order to write more (or fewer) tickets. Second, at certain times there are sometimes more drivers at risk for a ticket and relatively more of them are female. Whichever explanation is correct, the pattern provides confidence in the validity of the link between the gender composition of ticketed drivers and the speed of ticketed violations which is implied by the model.

To get a sense of the quantitative impact of the bias, I construct a back-of-the-envelope calculation of the number of "excess tickets" resulting from gender bias. First, we must assume there would be no behavioral response from drivers to the hypothetical policy change which drives the calculation.[28] Next, if we assume that fe-

[27]To assess the sensitivity of my standard errors, I estimated the baseline specifications in Tables 2.5 and 2.7 using OLS and clustered the standard errors by officer. When clustering, the male officer coefficients in the specifications for miles-per-hour and fine amount are still significant at the 5% level. In the specifications for the probability that the ticketed driver is female, the male officer coefficients are significant at the 10% level (p-values between 0.067 and 0.056).

[28]This would not be a good assumption if the policy change was large.

male police are unbiased and so use a single threshold $\theta^*(p_f)$, the pattern of violation thresholds consistent with the empirical results is $\theta^*(d_m, p_m) < \theta^*(d_f, p_m) < \theta^*(p_f)$, meaning that male police are biased against male drivers. Using the point estimate of the male officer effect for non-speeding tickets in Table (2.5), we obtain that $EOdds(p_m) = 0.9 * EOdds(p_f)$. Note that $EOdds(p_m) = \frac{N_f^m}{N_m^m}$, where N_f^m is the number of female drivers ticketed by male officers.

Think of correcting the bias by lowering male officer's threshold for females $\theta^*(d_f, p_m)$ and raising the threshold for males $\theta^*(d_m, p_m)$. The idea is that male police should have ticketed more female drivers and fewer male drivers. Holding the total number of tickets constant, let S represent the number of tickets to be shifted from males to females to equate the ticketing odds for male officers with that for female officers. We can do this by increasing the ticketing odds by a factor of $\frac{1}{\beta}$, where β is the odds ratio coefficient on male officer. The calculation for S is therefore:

$$\frac{N_f^m + S}{N_m^m - S} = \frac{1}{\beta}\frac{N_f^m}{N_m^m} \Leftrightarrow S = \frac{(1-\beta)N_m^m N_f^m}{\beta N_m^m + N_f^m} \tag{2.9}$$

The ticketing model implies that S is a lower bound. If $\theta^*(d_m, p_m)$ was increased and $\theta^*(d_f, p_m)$ was reduced until the two thresholds were equal, male officers using this new threshold $\theta^*(p_m)$ would have ticketed relatively more female drivers than the female police, because $\theta^*(p_m) < \theta^*(p_f)$. For non-speeding tickets with $\beta = 0.9$, $S = 620$. These 620 "shifted tickets" represent about 0.5% of the 110,556 non-speeding tickets issued during the 22 month sample period. The same calculation for speeding tickets, with $\beta = 0.85$, results in $S = 1,282$, which implies the shifted tickets are about 3.5% of the the 36,343 speeding tickets issued during the sample period.

Alternatively, when assuming that male police are unbiased while female police are biased, the empirical results would imply that female police are biased against female drivers. Making analogous calculations, for non-speeding violations the number of tickets S to be shifted from females to males is 100. For speeding tickets, $S = 36$. The quantitative impact of the gender bias is very small in this case because female officers issued relatively few traffic tickets.

2.4.1 Relating the composition test to the existing literature

First I compare the composition test to a test for gender bias in ticketing which is analogous to the test for racial bias in searches proposed in Anwar and Fang (2006). This test is based on Proposition (2.3), which suggests a test for gender bias based on comparing averages of miles-per-hour over the limit ($\overline{mph}$) for ticketed drivers in the following way pointed out by Anwar and Fang (2006):

Severity Test. *At least one police officer gender is biased if:*
$\overline{mph}(d_m, p_m) > \overline{mph}(d_m, p_f)$ *and* $\overline{mph}(d_f, p_m) < \overline{mph}(d_f, p_f)$*, or if:*
$\overline{mph}(d_m, p_m) < \overline{mph}(d_m, p_f)$ *and* $\overline{mph}(d_f, p_m) > \overline{mph}(d_f, p_f)$*.*

Critically, the severity test only compares average miles-per-hour for a gender group of ticketed drivers *across* the officer genders. For the test to reject the null, there must be a switching of the rank orders for male versus female drivers. Comparing average miles-per-hour for male and female drivers *within* officer gender is not informative about the relative positions of the ticketing thresholds, because the distributions of violation severity are different for male and female drivers.

Table (2.9) shows the results of conducting the severity test for miles-per-hour and fine amount. As both the male and female drivers ticketed by the female police were ticketed at greater miles-per-hour than the drivers ticketed by males, the severity test does not reject the null hypothesis of no gender bias in speeding tickets. The severity test also fails to reject the null for non-speeding violations, because male officers wrote less expensive tickets to both driver genders. In addition, according to Proposition (2.3) this empirical pattern of sample means corroborates the conclusion that male officers use a lower threshold on average (and therefore have a lower cost of ticketing on average). For this reason, the failure to reject the null hypothesis using Anwar and Fang's test is not surprising. Their test has zero power to detect bias when the groups of officers have different costs of ticketing on average, because there will never be a switching of rank orders even if one group of officers is biased.

Anwar and Fang conduct their test for bias by calculating rank orders of search rates and success rates by officer race for each racial group of drivers.[29] An analogous test for the ticketing outcome is to rank $P(T \mid d_m, p_m)$ versus $P(T \mid d_m, p_f)$, and $P(T \mid d_f, p_m)$ versus $P(T \mid d_f, p_f)$. If the ranking is different for male drivers than female drivers, then the null hypothesis of no gender bias is rejected. This test is not possible with the data at hand, but if the data were available for Boston we would expect this test to have zero power as well because male officers were tougher on average.

Antonovics and Knight (2007), using the same Boston Police Department data as I do, find that a search for contraband is more likely to be conducted when the race of

[29]For example, in the absence of bias, if white officers are more likely than black officers to search black motorists, then white officers should also be more likely than black officers to search white motorists.

the driver is different from the race of the police officer. Their theoretical model indicates that this cross-race effect is due to bias rather than statistical discrimination or omitted variables. An analogous test in the ticketing setting is to see if the probability of being ticketed, conditional on being stopped and receiving written documentation, depends on the interaction of officer and driver gender. As I showed in Section 4, according to my model, when officers use different ticketing standards on average it is possible for the coefficient on the interaction term (called gender mismatch) to be non-zero even in the absence of bias.

Table (2.10) shows the results of conducting this test for stops which occurred in Boston in April and May of 2001. The effect of the interaction term Male Officer $\times$ Female Driver on the probability of receiving a ticket is small and statistically insignificant in all six specifications. To compute this effect and its delta-method standard error for the probit specifications, I calculated the average of the partial effects of the interaction of male officer and female driver using the formulas described in Ai and Norton (2003). For the OLS specifications in Table (2.10), I verified that creating gender mismatch as Male Officer $\times$ Female Driver + Female Officer $\times$ Male Driver results in coefficients on mismatch equal to those reported divided by two.

In their model, Antonovics and Knight (2007) assume that if a bias exists, then both groups of officers are biased to the same degree against the drivers who do not belong to their group. If this does not hold, then the mismatch coefficient captures the average bias across the groups of officers. For example, if male police are slightly biased against male drivers while female police are unbiased, the average bias across the officers might be close to zero. This case is consistent with the data, and could be the reason why my test produces a different result than Antonovics and Knight's.

Price and Wolfers (2007) show that black basketball players in the NBA have more fouls called against them when the officiating crew is composed of white referees. The composition test can be applied to this cross-race effect as follows. In basketball, contact occurs on every play, so the referees must decide which instances of contact require a foul to be called. Suppose then that fouls vary by severity or by how obvious the infraction is, and assume that black and white referees use different, but unbiased severity (or obviousness) thresholds when calling fouls. Table 4 in Price and Wolfers shows that white referees tend to call fewer fouls, and that black players tend to commit fewer fouls. Assume then that black players commit less severe or obvious fouls while white referees use a higher threshold for calling a foul. Under these conditions, the NBA data is inconsistent with unbiased officiating. Unbiased white referees should call relatively more fouls against white players than black referees do because the white referees use a higher threshold, while Price and Wolfers find that the opposite pattern holds empirically.

Bagues and Esteve-Volart (2007) find that female candidates are more likely to pass the public examination for the Corps of the Spanish Judiciary when the share of males on the evaluation committee is larger. The idea of the composition test applies in this case, but not as cleanly because of a capacity constraint on the number of candidates each committee can pass. Table 11 in Bagues and Esteve-Volart shows that female candidates receive higher scores on average, and committees with more female members tend to assign higher scores. By using perhaps a lower objective standard, predominantly female committees might be expected to pass relatively more males. However, then the predominantly female committees should pass more candidates total, which is not possible because committees are only permitted to pass a fixed number of candidates.

2.5 Conclusion

If the police are gender biased in their traffic ticketing decisions, then the disparate impact of traffic ticketing on male drivers cannot be fully justified as resulting from the legitimate law enforcement objective of promoting safety on the roads. This paper developed a model of police and driver behavior which provides a testable implication of gender biased traffic ticketing. The test is based on the model's prediction that if the police are unbiased, the group of officers which is more reluctant to issue tickets should be relatively more likely to ticket male drivers who break the law. The test uses information on the miles-per-hour and fine amounts of ticketed violations to determine which group of officers is more reluctant to issue tickets. I reject the null hypothesis of unbiased ticketing in Boston because female police were more reluctant to ticket but were also relatively more likely to ticket female drivers. However, back of the envelope calculations based on the empirical results suggest the quantitative impact of the gender bias on traffic tickets received may be small. At least in Boston, this suggests that the sizable gender disparity in traffic tickets may be mostly due to differences in driving behavior by gender, rather than to biased policing.

Many empirical studies of discrimination estimate "cross-race" or "cross-gender" effects. These cross effects show how an outcome for subjects (drivers, players, candidates) depends on the racial or gender group of the evaluators (police, referees, committee members) who decide the outcome. The idea underlying this approach is that any dependence of the outcome on the pairing of subject and evaluator groups is difficult to explain as resulting from statistical discrimination or omitted variables. In my model, a cross-gender effect is generated when male and female police use unbiased but different threshold rules to decide which drivers to ticket. The test I developed in this paper offers a new method for determining if the direction of an observed cross effect is consistent with unbiased decision-making. The test can be applied if the demographic groups of subjects are systematically different on the outcome of interest, if a threshold decision rule is a reasonable approximation of the evaluator's decision process, and if there is a plausible way to rank the evaluator's thresholds.

2.6 Appendix

2.6.1 Proof of Proposition 2.2

$$\begin{aligned}
\frac{\partial Odds(p_g)}{\partial \theta^*} &= \frac{-f_f(\theta^*)[1-F_m\{\theta^*\}] + f_m(\theta^*)[1-F_f\{\theta^*\}]}{(1-F_m\{\theta^*\})^2} \\
&= \frac{-f_f(\theta^*)\int_{\theta^*}^{\infty} f_m(\theta)d\theta + f_m(\theta^*)\int_{\theta^*}^{\infty} f_f(\theta)d\theta}{(1-F_m\{\theta^*\})^2} \\
&= \frac{\int_{\theta^*}^{\infty} f_m(\theta^*)f_f(\theta) - f_f(\theta^*)f_m(\theta)d\theta}{\left(1-F_m\{\theta^*\}\right)^2} \\
&< 0
\end{aligned}$$

The sign of the derivative is negative because by the MLRP $f_f(\theta^*)f_m(\theta) > f_m(\theta^*)f_f(\theta)$, which makes the numerator negative.

2.6.2 Proof of Proposition 2.3

The average violation severity $\overline{\theta}(d_g, p_g)$ for a threshold $\theta^*(d_g, p_g)$ is

$$\overline{\theta} = E[\theta \mid \theta^*, d_g] = \frac{\int_{\theta^*}^{\infty} \theta f_g(\theta)d\theta}{1-F_g\{\theta^*\}}$$

The derivative of $\overline{\theta}$ with respect to θ^* is

$$\frac{\partial\overline{\theta}}{\partial\theta^*} = \frac{f_g(\theta^*)\int_{\theta^*}^{\infty} f_g(\theta)[\theta-\theta^*]d\theta}{\left(1-F_g\{\theta^*\}\right)^2} > 0$$

The derivative is positive because $\theta > \theta^*$.

2.6.3 Proof of Proposition 2.4

The unbiased violation threshold is θ^*, and $s(\theta^*) = \frac{N_f}{N_f+N_m}$ is the share of ticketed drivers that are female. The average violation severity for all drivers conditional on θ^* is:

$$E[\theta \mid \theta^*] = s(\theta^*)E[\theta \mid \theta^*, d_f] + (1-s(\theta^*))E[\theta \mid \theta^*, d_m]$$

From Proposition (2.2) it can be shown that $\frac{\partial s(\theta^*)}{\partial\theta^*} < 0$ (available by request). From Proposition (2.3) we know that $\frac{\partial E[\theta|\theta^*,d_g]}{\partial\theta^*} = \frac{\partial\bar{\theta}(d_g)}{\partial\theta^*} > 0$. Compute the derivative:

$$\frac{\partial E[\theta \mid \theta^*]}{\partial \theta^*} = \frac{\partial s(\theta^*)}{\partial \theta^*}[\bar{\theta}(d_f) - \bar{\theta}(d_m)] + s(\theta^*)\frac{\partial \bar{\theta}(d_f)}{\partial \theta^*} + [1 - s(\theta^*)]\frac{\partial \bar{\theta}(d_m)}{\partial \theta^*}$$

The last two terms are both positive, so $\frac{\partial E[\theta|\theta^*]}{\partial \theta^*}$ is guaranteed to be positive if the first term is greater than or equal to zero. This is satisfied if $\bar{\theta}(d_m) \geq \bar{\theta}(d_f)$.

2.6.4 Intepretation of gender mismatch coefficients

For the miles-per-hour (and fine amount) specifications, there are four key sample averages:

	Male Officer	Female Officer
Male Driver	m_1	m_2
Female Driver	m_3	m_4

I estimate OLS specifications of this form:

$$\text{MPH} = \beta_0+\beta_1\times\text{Male Officer}+\beta_2\times\text{Female Driver}+\beta_3\times(\text{Male Officer}\times\text{Female Driver})+\varepsilon$$

The equations for each sample average are therefore:

$$\begin{aligned} m_1 &= \beta_0+\beta_1 \\ m_2 &= \beta_0 \\ m_3 &= \beta_0+\beta_1+\beta_2+\beta_3 \\ m_4 &= \beta_0+\beta_2 \end{aligned}$$

Substitute the other equations into the equation for m_3:

$$\begin{aligned} \beta_3 &= m_3-m_2-[m_1-m_2]-[m_4-m_2] \\ \beta_3 &= [m_2-m_4]-[m_1-m_3] \\ \beta_3 &= [\overline{mph}(d_m,p_f)-\overline{mph}(d_f,p_f)]-[\overline{mph}(d_m,p_m)-\overline{mph}(d_f,p_m)] \end{aligned}$$

When the gender mismatch variable is created analogously to the racial mismatch variable in Antonovics and Knight (2007), the OLS equation and the equations for each m are:

$$\begin{aligned} \text{MPH} = \alpha_0 &+ \alpha_1\times\text{Male Officer}+\alpha_2\times\text{Female Driver} \\ &+ \alpha_3\times(\text{Male Officer}\times\text{Female Driver}+\text{Female Officer}\times\text{Male Driver})+\varepsilon \end{aligned}$$

$$\begin{aligned} m_1 &= \alpha_0+\alpha_1 \\ m_2 &= \alpha_0+\alpha_3 \\ m_3 &= \alpha_0+\alpha_1+\alpha_2+\alpha_3 \\ m_4 &= \alpha_0+\alpha_2 \end{aligned}$$

Start with the equation for m_3 and make substitutions:

$$\begin{aligned}
\alpha_3 &= m_3 - m_1 - m_4 - \alpha_0 \\
\alpha_3 &= m_3 - m_1 - m_4 + m_2 - \alpha_3 \\
\alpha_3 &= \frac{[m_2 - m_4] - [m_1 - m_3]}{2} = \frac{\beta_3}{2}
\end{aligned}$$

2.7 Tables and Figures

Table 2.1: Sample means of traffic ticket variables by police officer gender.

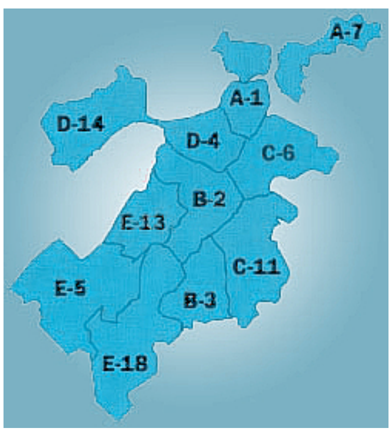
	Male Officers	Female Officers
Female Driver	28.4%	30.5%
Weekend	24.2%	21.2%
Commute	55.4%	50.8%
Daytime	66.8%	73.2%
Speeding	25.0%	17.6%
Failure to Stop	31.8%	32.1%
No Inspection Sticker	8.9%	4.6%
No Seat Belt	14.9%	5.8%
MPH Over Limit (speeding)	14.3	15.4
	(n=34,133)	(n=923)
Fine Amount (non-speeding)	$48.3	$55.8
	(n=68,090)	(n=2,847)
Number of Citations	141,224	5,675
Mean Citations per Officer-Day	9.5	4.9

Tickets file merged with officer data, excludes cases where officer gender was missing.
Officer-Day: A calendar day in which the officer wrote at least one traffic ticket.
12.6% of the police officers are female.

Table 2.2: The Boston Police Force by District

	Percent Female Officers	Number of Officers
A-1 Downtown/Beacon Hill/ Chinatown/Charlestown	12.0%	142
A-7 East Boston	12.5%	160
B-2 Roxbury/Mission Hill	10.1%	109
B-3 Mattapan/North Dorchester	11.7%	162
C-6 South Boston	11.8%	76
C-11 Dorchester	19.2%	78
D-4 Back Bay/South End/Fenway	8.0%	75
D-14 Allston/Brighton	16.4%	165
E-5 West Roxbury/Roslindale	15.2%	79
E-13 Jamaica Plain	15.4%	84
E-18 Hyde Park	11.8%	110
Special Operations	9.3%	182

Excludes cases where officer gender was missing.



Map of Boston police districts.

Figure 2.1: Cumulative distribution of miles-per-hour over limit for ticketed drivers.

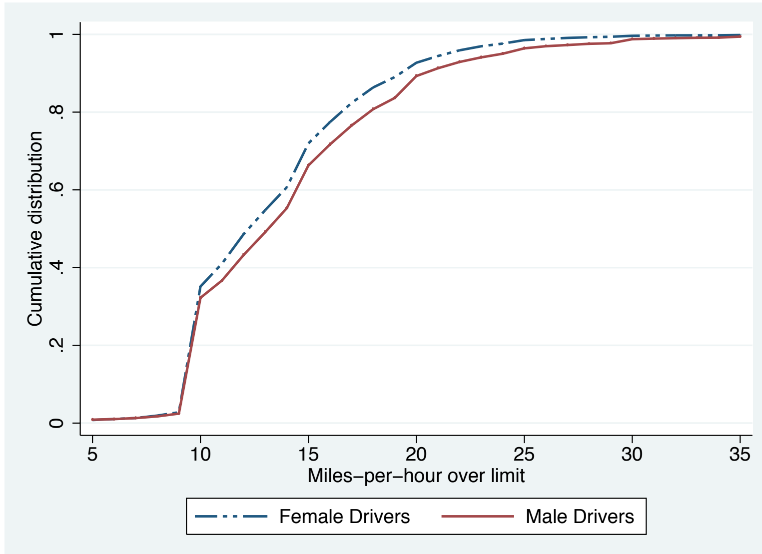


Figure 2.2: Total citations by month and officer gender.

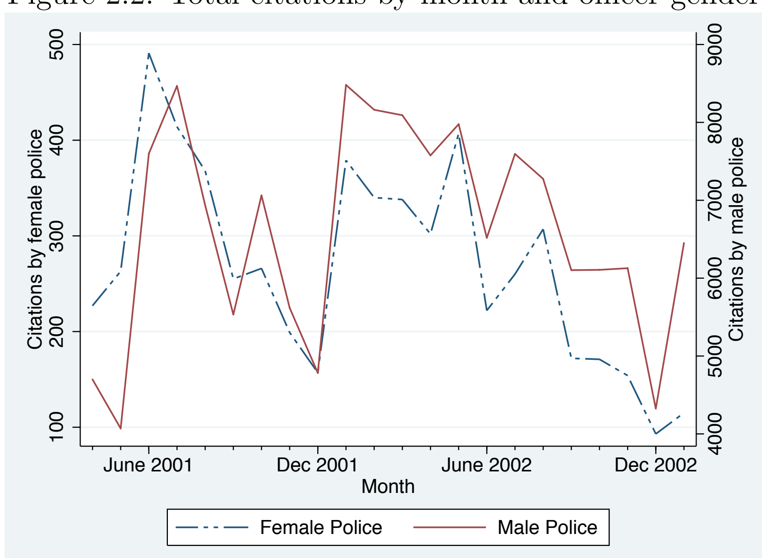


Table 2.3: Means of 2000 Census variables for Boston metropolitan area police officers

	Male Police	Female Police
Age	40.5	41.1
College Graduate	52.4%	40.3%
Weeks Worked	51.2	48.6
Hours Worked/Week	47.9	40.1
Annual Income	62,435	42,152
Implied Hourly Wage	25.5	21.6
Number of Officers	498	72

Author's calculations from IPUMS 5% sample of 2000 Census.

Table 2.4: Effect of male officer on gender of ticketed driver, no controls.

		Speeding tickets			Other violations	
Female Driver (Yes=1)	OLS	Logit	OLS	Logit	OLS	Logit
Male Officer	-0.020	0.915	-0.047^{**}	0.808^{**}	-0.026^{**}	0.882^{**}
	(0.015)	(0.061)	(0.016)	(0.057)	(0.007)	(0.029)
MPH over limit			-0.007^{**}	0.966^{**}		
			(0.0004)	(0.002)		
Observations	36,343	36,343	34,024	34,024	110,556	110,556

Dependent variable is Female Driver (Yes=1, No=0).
The only control variable is miles-per-hour over the speed limit, where indicated.
Coefficients from logit models are presented as odds ratios.
Heteroskedastic-robust OLS standard errors,**p<0.05, *p<0.10

Table 2.5: Effect of male officer on gender of ticketed driver.

	Speeding tickets			Other violations		
Female Driver (Yes=1)	OLS	Logit	Logit	OLS	Logit	Logit
Male Officer	−0.034**	0.852**	0.845**	−0.021**	0.900**	0.912**
	(0.016)	(0.061)	(0.060)	(0.007)	(0.030)	(0.030)
MPH over limit	−0.006**	0.967**	0.967**			
	(0.0005)	(0.0026)	(0.0026)			
Speed limit	−0.007**	0.969**	0.968**			
	(0.0006)	(0.0025)	(0.0025)			
Black Driver	0.002	1.007	1.006	0.001	1.002	1.003
	(0.006)	(0.029)	(0.029)	(0.003)	(0.017)	(0.017)
Hispanic Driver	−0.072**	0.700**	0.698**	−0.048**	0.774**	0.774**
	(0.008)	(0.029)	(0.029)	(0.004)	(0.018)	(0.018)
Driver Age	0.010**	1.053**	1.053**	0.009**	1.049**	1.050**
	(0.001)	(0.006)	(0.006)	(0.0005)	(0.004)	(0.004)
In-town Driver	0.019**	1.091**	1.089**	0.029**	1.160**	1.166**
	(0.006)	(0.028)	(0.028)	(0.003)	(0.017)	(0.017)
Day and Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Day and Time Interactions	No	No	Yes	No	No	Yes
Boston District Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,941	33,941	33,941	110,531	110,531	110,531

Dependent variable is Female Driver (Yes=1, No=0).

Coefficients from logit models are presented as odds ratios.

Heteroskedastic-robust OLS standard errors, **p<0.05, *p<0.10

Table 2.6: Effect of male officer on gender of ticketed driver, robustness checks.

	Speeding tickets				Other violations			
Female Driver (Yes=1)	OLS	Logit	Logit	Logit	OLS	Logit	Logit	Logit
Male Officer	−0.032**	0.860**	0.828**	0.882*	−0.022**	0.900**	0.854**	0.901**
	(0.016)	(0.061)	(0.060)	(0.064)	(0.007)	(0.030)	(0.029)	(0.030)
MPH over limit	−0.006**	0.968**	0.969**	0.969**				
	(0.0005)	(0.003)	(0.003)	(0.003)				
Log(Total Citations)	0.026**	1.129**			0.015**	1.081**		
by month and district	(0.008)	(0.042)			(0.004)	(0.024)		
Log(Total Citations)			1.117**				1.106**	
by officer and day			(0.018)				(0.009)	
Sample Month Dummies	No	No	No	Yes	No	No	No	Yes
Day and Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boston District Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Driver Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,941	33,941	33,941	33,491	110,531	110,531	110,531	110,531

Dependent variable is Female Driver (Yes=1, No=0).

Coefficients from logit models are presented as odds ratios.

Heteroskedastic-robust OLS standard errors,**p<0.05, *p<0.10

Table 2.7: Effect of male officer on severity of ticketed violation.

	Speeding tickets			Other violations			
	Miles-per-hour over limit			Fine amount or category			
	OLS	OLS	OLS	OLS	OLS	OLogit	OLogit
Male Officer	-1.04^{**}	-1.47^{**}	-1.47^{**}	-7.76^{**}	-5.84^{**}	0.61^{**}	0.66^{**}
	(0.24)	(0.23)	(0.23)	(0.94)	(0.93)	(0.024)	(0.026)
Female Driver	-0.67^{*}	-0.23	-0.22	-3.26^{**}	-3.51^{**}	0.96^{**}	0.98
	(0.35)	(0.33)	(0.33)	(1.30)	(1.29)	(0.016)	(0.016)
Gender Mismatch	-0.18	-0.47	-0.48	2.17^{*}	2.77^{**}		
(Male Officer × Female Driver)	(0.36)	(0.34)	(0.34)	(1.32)	(1.31)		
Speed Limit	No	Yes	Yes	n.a.	n.a.	n.a.	n.a.
Driver Demographics	No	Yes	Yes	No	Yes	No	Yes
Day and Time Dummies	No	Yes	Yes	No	Yes	No	Yes
Day and Time Interactions	No	No	Yes	No	Yes	No	Yes
Boston District Dummies	No	Yes	Yes	No	Yes	No	Yes
Observations	34,024	33,941	33,941	68,759	68,744	68,759	68,744

Fine categories: less than \$26, \$26 to \$35, \$36 to \$50, \$51 to \$100, greater than \$100.

Coefficients from ordered logit models are presented as odds ratios.

Heteroskedastic-robust OLS standard errors,**$p<0.05$, *$p<0.10$

Table 2.8: Effect of male officer on severity of ticketed violation, robustness checks.

	Speeding tickets			Other violations				
	Miles-per-hour over limit			Fine amount or category				
	OLS	OLS	OLS	OLS	OLS	OLS	OLogit	OLogit
Male Officer	-1.51^{**}	-1.57^{**}	-1.69^{**}	-5.82^{**}	-3.91^{**}	-5.77^{**}	0.67^{**}	0.67^{**}
	(0.23)	(0.18)	(0.23)	(0.93)	(0.70)	(0.94)	(0.026)	(0.027)
Female Driver	-0.22		-0.23	-3.42^{**}		-3.43^{**}	0.99	0.99
	(0.33)		(0.33)	(1.29)		(1.29)	(0.017)	(0.017)
Gender Mismatch	-0.48		-0.44	2.68^{**}		2.70^{**}		
(Male Officer × Female Driver)	(0.34)		(0.33)	(1.31)		(1.32)		
Log(Total Citations)	-0.61^{**}			-0.64^{*}			0.88^{**}	
by month and district	(0.083)			(0.33)			(0.020)	
Log(Total Citations)		-0.36^{**}			-2.83^{**}			
by officer and day		(0.04)			(0.13)			
Speed Limit	Yes	Yes	Yes	n.a.	n.a.	n.a.	n.a.	n.a.
Sample Month Dummies	No	No	Yes	No	No	Yes	No	Yes
Driver Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day and Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boston District Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	33,941	34,970	33,941	68,744	70,921	68,744	68,744	68,744

Fine categories: less than $26, $26 to $35, $36 to $50, $51 to $100, greater than $100.

Coefficients from ordered logit models are presented as odds ratios.

Heteroskedastic-robust OLS standard errors,**p<0.05, *p<0.10

Table 2.9: Tests analogous to Anwar and Fang (2006).

		Male Officers	Female Officers	p-value
Miles-per-hour	Male Drivers	14.5	15.6	< 0.001
	Female Drivers	13.7	14.9	< 0.001
Fine amount	Male Drivers	48.6	56.3	< 0.001
	Female Drivers	47.5	53.1	< 0.001

p-values for null that mean differences are not different from zero.

Average miles-per-hour for speeding violations, fine amount for other violations.

Table 2.10: Test analogous to Antonovics and Knight (2007): Effect of gender mismatch on the probability of being ticketed conditional on being stopped. Gender mismatch is the interaction of male officer and female driver.

	Speeding			Other violations		
Ticketed (Yes=1, No=0)	OLS	OLS	Probit	OLS	OLS	Probit
Male Officer	-0.162^{**}	-0.217^{**}	-0.201^{**}	-0.015	-0.017	-0.028
	(0.049)	(0.048)	(0.036)	(0.023)	(0.023)	(0.018)
Female Driver	-0.031	-0.071	-0.042^{**}	-0.041	-0.041	-0.072^{**}
	(0.075)	(0.073)	(0.012)	(0.038)	(0.037)	(0.010)
Gender Mismatch	-0.026	0.029	0.030	-0.030	-0.033	-0.034
(Male Officer × Female Driver)	(0.076)	(0.074)	(0.073)	(0.039)	(0.039)	(0.039)
MPH over limit	No	Yes	Yes	n.a.	n.a.	n.a.
Driver Demographics	No	Yes	Yes	No	Yes	Yes
Day and Time Dummies	No	Yes	Yes	No	Yes	Yes
Observations	6,410	6,238	6,238	12,057	11,980	11,980

Data on warnings is only available for stops in April and May of 2001.

Probit estimates are the average of the marginal effects on the probability of being ticketed.

Delta-method standard errors for probit estimates.

Heteroskedastic-robust OLS standard errors,$^{**}p<0.05$, $^{*}p<0.10$

2.8 References

Ai, Chunrong and Norton, Edward. “Interaction Terms in Logit and Probit Models”, *Economics Letters* 2003.

Anbarci, Nejat and Lee, Jungmin. “Speed Discounting and Racial Disparities: Evidence from Speeding Tickets in Boston”, IZA Discussion paper 3903, December 2008.

Antonovics, Kate and Knight, Brian. “A New Look at Racial Profiling: Evidence from the Boston Police Department”, forthcoming, *Review of Economics and Statistics* 2007.

Anwar, Shamena and Fang, Hanming. “An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence”, *American Economic Review*, March 2006.

Ayres, Ian. “Outcome Tests of Racial Disparities in Police Practices”, *Justice Research and Policy*, Fall 2002.

Bagues, Manuel and Esteve-Volart, Berta. “Can Gender Parity Break the Glass Ceiling? Evidence from a repeated randomized experiment”, Working paper, June 2007.

Becker, Gary. “The Economics of Discrimination”, University of Chicago Press, 1957.

Blackmon, B. Glenn and Zeckhauser, Richard. “Mispriced Equity: Regulated Rates for Auto Insurance in Massachusetts”, *American Economic Review*, Papers and Proceedings, May 1991.

Blalock, Garrik; Jed DeVaro; Stephanie Leventhal; Daniel Simon. “Gender Bias in Power Relationships: Evidence from Police Traffic Stops”, Working paper, Feb. 2007.

Dedman, Bill and Latour, Francie. “Race, sex, and age drive ticketing”, *The Boston Globe*, July 20 2003.

Dedman, Bill. “Boston police to get tough on tickets”, *The Boston Globe*, Jan. 17 2004.

Dharmapala, Dhammika and Ross, Steven. “Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence”, *Contributions to Economic Analysis and Policy*, 2004.

Durose, Matthew; Erica Smith; Patrick Langan. “Contacts between Police and the Public, 2005”. U.S. Department of Justice, Bureau of Justice Statistics, Special Report. April 2007.

Farrell, Amy; Dean McDevitt; Lisa Bailey; Carsten Andresen; Erica Pierce. "Massachusetts Racial and Gender Profiling Final Report: Executive Summary", Northeastern University Institute on Race and Justice, May 2004.

Grogger, Jeffrey and Ridgeway, Greg. "Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness", *Journal of the American Statistical Association*, Sept. 2006.

Edlin, Aaron and Karaca-Mandic, Pinar. "The Accident Externality from Driving", *Journal of Political Economy*, 2006.

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

Levitt, Steven and Porter, Jack. "How Dangerous are Drinking Drivers?", *Journal of Political Economy*, 2001.

Makowsky, Michael and Stratmann, Thomas. "Political Economy at Any Speed: What Determines Traffic Citations?", forthcoming, *American Economic Review* 2008.

National Highway Traffic Safety Administration, Center for Statistics and Analysis. "2006 Annual Assessment of Motor Vehicle Crashes." Sept. 2007.

Price, Joseph and Wolfers, Justin. "Racial Discrimination Among NBA Referees", NBER working paper 13206, June 2007.

Rowe, Brian. "Discretion and Ulterior Motives in Traffic Stops: The Detection of Other Crimes and the Revenue from Tickets", Working paper, April 2009.

Schanzenbach, Max. "Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics", *Journal of Legal Studies*, Jan. 2005.

CHAPTER III

Discretion and Ulterior Motives in Traffic Stops: The Detection of Other Crimes and the Revenue from Tickets

3.1 Introduction

When a police officer in the United States stops a driver for a minor traffic violation, the officer often decides to let the driver go without issuing him a ticket. A recent analysis of U.S. Bureau of Justice Statistics data finds that 40.3% of stopped drivers were not required to pay a fine[1], an outcome which I refer to as a *warning*.

This use of discretion in traffic ticketing raises several important issues. First, does the high prevalence of warnings imply that traffic enforcement is grossly inefficient? Sizable public resources are expended to detect, stop, and prosecute violators of traffic laws.[2] Traffic laws are intended to promote public safety on the roads, and traffic accidents are a significant public health problem. Approximately 42,000 people were killed and 2.5 million people were injured in traffic accidents in 2006 (NHTSA 2007). Yet because many stopped drivers are only given warnings, the expected cost to drivers of committing a traffic violation is below what it would be if all stopped drivers were ticketed. As a result, less deterrence of traffic violations, and therefore more traffic accidents, may be realized for a given level of public expenditures on traffic enforcement.

Second, the impact of police discretion in traffic stops has generated a great deal of controversy with respect to racial profiling and racial disparities in searches of stopped

[1]These drivers received either a written warning, a verbal warning, or nothing at all (Durose et al. 2007).

[2]About 17.8 million drivers were stopped by the police in 2005 (Durose et al. 2007), and court cases related to traffic violations accounted for more than half of all cases in state courts in 2004 (National Center for State Courts, 2005).

motor vehicles for drugs.[3] The ability to give tickets or warnings provides another way in which the police can take different actions against drivers who committed identical traffic violations. Discretion in ticketing may therefore lead to unjustified differential treatment in ticketing based on characteristics such as race and gender.[4]

In light of these concerns, an important and policy relevant question is whether there is any good reason to allow the police to issue warnings. This paper provides an answer, based on the importance of the ulterior motive of detecting other crimes.

In the U.S., the police are legally permitted to stop drivers for minor traffic violations for the purpose of detecting other criminal activities. In *Whren et al. vs United States 1996*, the Supreme Court held that is legal for police officers to stop drivers for minor traffic violations if they suspect the driver is guilty of another crime. Police may do this even if a "reasonable" officer would not have stopped the driver for the minor violation alone.[5]

Detecting other criminal activities with traffic stops is an important aspect of traffic enforcement. In 2005 about 4.8% of all stopped drivers were searched for drugs or other contraband, and 2.4% of all stopped drivers were arrested.[6] Almost all of these searches and arrests occurred after the driver was stopped for a minor traffic infraction such as speeding or a vehicle defect. Since only 11.6% of searches uncovered evidence of criminal activity, many drivers were arrested for reasons unrelated to a vehicle search. Also, evidence from legal cases supports the idea that police commonly use minor traffic violations as a pretext for stopping motorists in order to look for

[3]See Harris (1997) for an influential argument that police discretion has an adverse impact on minority motorists. Ramirez et al. (2000) summarize many studies of racial profiling and document a dramatic increase in the collection of traffic stop data in response to concerns about profiling. According to Antonovics and Knight (2007), there have been over 200 court cases related to racial profiling in searches of stopped motorists.

[4]Dedman and Latour (2003) documented racial, gender, and age disparities in the probability of receiving a ticket after being stopped by police in Massachusetts. Their findings prompted the state of Massachusetts to fund a comprehensive follow-up study (Farrell et al. 2004), which reaches similar conclusions about these disparities in ticketing.

[5]Holding of *Whren*: "The temporary detention of a motorist upon probable cause to believe that he has violated the traffic laws does not violate the Fourth Amendment's prohibition against unreasonable seizures, even if a reasonable officer would not have stopped the motorist absent some additional law enforcement objective."

[6]I arrive at 4.8% based on Durose et al. (2007), who estimate that 854,990 drivers were searched out of a total of 17.8 million stopped drivers. The figure for the percent of stopped drivers arrested is from Table 7 of the same report.

other crimes.[7]

Therefore, in order to analyze the conditions under which warnings may be efficient I employ a theoretical framework in which traffic stops accomplish a parallel objective of detecting other crimes. This framework is based on Shavell (1991) and Mookerjee and Png (1992), who formulated the idea of "general enforcement", meaning that one detection probability applies to many different crimes.[8] Using a general enforcement model based on Shavell (1991), who does not consider warnings, I show that when police stop drivers for the ulterior motive of detecting other crimes it may be optimal to let many stopped drivers go with only a warning.

When the enforcement authority (in this paper, a local government) can detect other criminal activity with traffic stops, I show that more drivers are stopped for traffic violations than would be optimal if traffic offenses were the only crimes being monitored. Given this, the expected fine for minor traffic offenses may be higher than is socially desirable. The local government can then reduce the expected fine for traffic offenses by letting some (randomly chosen) stopped drivers go with only a warning. This is optimal because the dollar amount of the fines for various traffic offenses are fixed by state law.

This result provides a parsimonious justification for police discretion in traffic ticketing which is consistent with empirical evidence about how and why traffic laws are actually enforced. Alternatively, warnings could be rationalized by assuming they have a direct deterrence effect, perhaps by reminding drivers to be more careful. I consider some implications of this explanation for warnings, but I find little support for them.

Even accepting that warnings are theoretically possible in an efficient enforcement scheme, how can we evaluate whether they are being used efficiently in practice? Testing for efficient enforcement with respect to race or gender is beyond the scope of this paper.[9] Here I examine police treatment of out-of-town drivers, which offers an opportunity to test the general enforcement model against data on traffic stops by addressing another question: Do the police discriminate against stopped out-of-town

[7]For example, when summarizing racial profiling cases Ramirez et al. (2000) state that "By far the most common complaint by members of communities of color is that they are being stopped for petty traffic violations such as under-inflated tires, failure to signal properly before switching lanes, vehicle equipment failures..." and so on. Harris (1997) asserts that "In fact, searching cars for narcotics is perhaps *the* major motivation for making these stops."

[8]The main result of these two papers is to show that when the probabilities of detection for different crimes are linked, optimal fines should rise with the severity of the crime, and maximal fines should only be used for the most serious offenses.

[9]Rowe (2009) develops a test for gender bias in ticketing. He uses a theoretical framework which is able to address the question of how police decide which (otherwise equal) drivers to ticket.

drivers, by giving them warnings less frequently, because of the ulterior motive of ticket revenue?

Articles about traffic enforcement in the popular press often describe a sense that the authorities may be more concerned with generating revenue than providing safe road conditions.[10] Recently, public backlash about revenue raising with traffic fines forced the Virginia state legislature to back away from a plan to substantially raise fines for reckless driving offenses (Craig 2008). Facing a budget shortfall, Colorado is currently considering increasing traffic fines. In Florida there is a voter referendum to limit growth in government revenue, including the revenue from tickets.

My analysis leads to an example of how a concern for revenue can be an aspect of an efficient enforcement policy. I use the general enforcement model to show that giving warnings less frequently to stopped out-of-town drivers is part of an optimal policy for a local government. This is because the fine revenue from out-of-town drivers is a net transfer of wealth to the local community, while fine revenue from in-town drivers is a welfare-neutral transfer payment.

Empirically, I find that local police (in Massachusetts) are more likely to ticket out-of-town drivers stopped for speeding than similar in-town drivers, controlling for a rich set of observable characteristics. Yet this result does not speak to why this discrimination occurs. The police may be more likely to ticket out-of-town drivers because of the ulterior motive of ticket revenue, as predicted by the model. Alternatively, the police may treat out-of-town drivers more harshly for reasons unrelated to ticket revenue.

I use the model to derive a prediction for an empirical pattern which should appear if police are motivated by ticket revenue. If motivated by revenue, discrimination against out-of-town drivers should disappear for more serious violations, a pattern which I find support for in the data. At least with respect to the different treatment of out-of-town drivers, this suggests discretion in ticketing appears to be used efficiently. This result also provides additional confidence in the validity of the general enforcement framework for traffic stops.

3.1.1 Related Literature

Harrington (1988) considers a dynamic model based on the context of the enforcement of environmental regulations. The enforcement agency monitors firms repeatedly, fines and inspection probabilities can be conditioned on past compliance, and a maximum fine amount is exogenously fixed. Harrington (1988) finds that the optimal

[10] See for example Cardwell (2003), Howlett (2004) and Vitello (2007).

fine for "complying" firms is zero, as this provides the highest incentive for "violating" firms to comply with the law and thereby move into the "complying" group. My rationalization for warnings (a fine of zero for some violators) occurs in a static setting with no repeated interactions and no mechanism to condition enforcement on past compliance.

Dittman (2006) analyzes a model where the government maximizes a weighted sum of social welfare and the "residual budget", which is the revenue from fines net of enforcement expenditures. He shows that as the weight on the residual budget increases, the optimal probability of detection falls. Intuitively, the government induces more crime in order to catch and fine more criminals. Dittman excludes criminal utility from the welfare function because he focuses on serious crimes. In my analysis, criminal utility is included and the concern for fine revenue arises only from the maximization of social welfare.

Baicker and Jacobson (2007) find that police agencies increase the drug arrest rate in response to the net fiscal incentives of asset forfeiture laws, which allow agencies to keep assets seized in drug arrests. Garrett and Wagner (2008) find that in North Carolina, lagged negative county level revenue shocks lead to small increases in the number of traffic tickets. Makowsky and Stratmann (2008) find that Massachussetts police officers are more likely to issue a traffic ticket, and charge a higher fine amount, when their town's budget is tight. Their empirical specifications are probit models for whether stopped drivers were ticketed, and Heckman selection models of the fine amount charged to the driver.[11]

The normative consequences of these revenue effects are ambiguous. Regarding Baicker and Jacobson, we do not know if the police increased their total enforcement effort or shifted their efforts towards drugs and away from other crimes in response to the seizure incentives.[12] Regarding Garrett and Wagner, in times of fiscal distress it may be optimal for the county government to issue more traffic tickets if the alternative is to increase distortionary tax rates. The same argument applies to Makowsky and Stratmann's result that towns with tight budgets set a higher probability of giving a ticket. My use of the general enforcement framework provides a clear prediction of patterns we should expect in the data if the government is concerned with fine revenue as part of an optimal policy.

[11] To identify the Heckman model, an indicator for whether the driver had a commercial driver's license is included in the equation for whether a ticket was issued but excluded from the fine amount equation.

[12] Baicker and Jacobson provide evidence that police agencies shift their enforcement efforts towards heroin and away from marijuana in response to the budgetary incentives of seizure laws.

3.2 Warnings and the detection of other crimes

First, I demonstrate that warnings are inefficient in a basic model where traffic laws are enforced only to deter traffic violations. In the both the basic model and the general enforcement model, individuals are assumed to make decisions about crime by comparing their benefit from committing crime to the expected cost of breaking the law.

3.2.1 Traffic stops only detect traffic violations

Let b denote a driver's benefit from violating a traffic law, distributed according to the function $G(b)$. The probability of a violator being stopped is $p(x)$, where x is the level of public resources expended on monitoring the roads. Assume that $\frac{\partial p(x)}{\partial x} > 0$. The dollar value of the fine is F, which is exogenously fixed. The expected fine from breaking the law is $z = p(x)F$. Drivers are risk neutral and violate the law if $b > z$.

Normalize the number of drivers to 1 so that the number of violators is equal to $1 - G(z)$. Let c be the marginal external cost of violating the law, so that $c[1 - G(z)]$ is the total social cost incurred from traffic infractions. The government chooses enforcement spending x to maximize the following social welfare function:

$$\int_{p(x)F}^{\infty} b dG(b) - c[1 - G(p(x)F)] - x \tag{3.1}$$

According to equation (3.1), the government takes into account the private gains drivers receive from violations (time saved by speeding), the social costs violators impose on others (increased risk of injury), and the cost of monitoring the roads. At the optimum x^*, the following equation holds:[13]

$$p(x^*)F = c - \frac{1}{\frac{\partial G(z)}{\partial x}} \tag{3.2}$$

Recall that in the U.S. about 60% of stopped drivers are ticketed, while the rest are let go with only a warning. Let π denote the proportion of stopped drivers who are ticketed. For the same $p(x)$, compare the number of violators when $\pi < 1$ and when $\pi = 1$:

$$1 - G(\pi \times p(x)F) > 1 - G(p(x)F)$$

[13]The first order condition with respect to x is $\frac{\partial p}{\partial x} F g(z) c - \frac{\partial p}{\partial x} F^2 g(z) p(x) = 1$. The solution is below the Pigouvian solution of $z = c$ because increasing z requires increasing x when F is held fixed. See Polinsky and Shavell (1984) for a details.

Using the same amount of resources, fewer violators are deterred and the total social cost of traffic crime is higher when a fraction $\pi < 1$ of detected violators are ticketed. In this basic setting, enforcement is inefficient if *any* drivers receive warnings.

The same conclusion will hold in a setting in which marginal deterrence is important. For example, suppose that drivers decide by how much to exceed the speed limit, and that faster violations are more dangerous. Given exogenous fines, the best use of resources by the enforcement authority would be to set a lower probability of stop for slower violations. Setting a higher probability of a warning for slower violations is a waste of resources because it is costly to stop the slow speeders in the first place.

3.2.2 Traffic stops detect other crimes

Now I consider the case where the police use traffic stops to detect other crimes, using the model of Shavell (1991) to represent this idea. An individual's benefit from crime b is distributed by $G(b)$. There are many different crimes, each characterized by its social cost c, which is distributed by $S(c)$. Each individual is characterized by the pair (b_i, c_i). The corresponding probability density functions for $G(\cdot)$ and $S(\cdot)$ are written as $g(\cdot)$ and $s(\cdot)$ respectively. Assume the distribution of benefits $G(b)$ is the same for each crime c.

Shavell (1991) takes this setup to mean that individuals only commit one crime, where the assumption that $G(b)$ is the same for all c is only for simplicity. For individuals who break traffic laws and commit more serious crimes, the formulation is equivalent to having the same individual appear in the social welfare calculation twice. This individual is described by (b_i^1, c_i^1) and (b_i^2, c_i^2), where crime 1 might be speeding and crime 2 might be drug trafficking. Critically, the model implies that this individual's choices of traffic crimes and serious crimes occur independently of each other, a restriction which I justify shortly.

The schedule of fines as a function of the severity of the offense is $F(c)$. In Shavell (1991), fine amounts are choice variables. I assume that $F(c)$ is fixed exogenously, which makes sense in the context of law enforcement by a local government because the schedule of fines is fixed by the state.[14] Also, I assume that all fines $F(c)$ must be less than some maximal fine $\bar{F}$. The probability of detection is $p(x(c), y)$, where $x(c)$ is enforcement spending specific to detecting crime c and y is general enforcement

[14] Massachusetts state law Chapter 90, section 20 sets a schedule of fines for violating posted speed limits. Available at http://www.mass.gov/legis/laws/mgl/90-20.htm

expenditures. For example, drunk driving is detected by both the presence of police on the roads and by checkpoints set up specifically to detect drinking drivers.

Although the local government cannot choose $F(c)$, it can choose the probability of fining a detected criminal, denoted by $\pi(c)$. This implies that the probability of giving a warning to a detected criminal is $1-\pi(c)$. The expected fine for committing offense c is then $z(c) = \pi(c) \times p(x(c), y) \times F(c)$. The government's problem is to choose $\pi(c)$, $x(c)$, and y to maximize the social welfare function W:

$$W = \int_0^\infty \int_{\pi(c)\times p(x(c),y)\times F(c)}^\infty (b-c)dG(b)dS(c) - \int_0^\infty x(c)dc - y \tag{3.3}$$

The double integral in (3.3) is the net social benefit of crime. The two terms subtracted are the costs of enforcement. Shavell (1991) shows that the optimal y (call it y^*) will be positive if the average external cost $E[c]$ is sufficiently large or if the maximal fine $\bar{F}$ is large. I assume $y^* > 0$, so that all crimes are detected with at least probability $p(0, y^*)$.

The crucial restriction in this setup is that drivers who are also more serious criminals cannot alter their chances of detection by choosing how to violate traffic laws. This might seem unrealistic at first. However, the available evidence indicates that this restriction is actually *appropriate* in the context of traffic enforcement. Because traffic laws are numerous, complex, and often highly technical, it is almost impossible to avoid committing a violation that is grounds for a stop.

According to Harris (1997), a rule of thumb held by the police is that drivers cannot go more than three blocks without violating a traffic law. Ramirez et al. (2000) refer to a similar rule of thumb that drivers cannot go more than 1 or 2 minutes before a violation is observed. Harris (1997) also gives examples of laws that are sufficiently vague (such as that a driver may not "suddenly" slow down before turning) so that the police may invoke them as grounds for a stop at will. Knowles, Persico, and Todd (2001) similarly argue that because of *Whren* and the complexity of traffic codes, the legal constraint that police must have probable cause for a traffic stop is not binding in practice.

Criminals may also find it difficult to reduce their chance of being detected by refraining from breaking traffic laws because most other drivers violate traffic laws. In a study which conducted direct observations of drivers on the New Jersey Turnpike, Lamberth (1994) found that 98% of drivers were exceeding the posted speed limit of 55 miles-per-hour. It seems unlikely that an intelligent drug trafficker would choose to drive exactly 55 in this situation. The police would notice that he was driving differently than everyone else, and then use some other traffic regulation as the grounds for a stop.

I now demonstrate that it may be optimal for the government to release some detected violators without requiring the payment of a fine. Letting $*$ denote optimal values, I will show that $0 < \pi^*(c) < 1$ for some crimes c. The first step in the argument is:

Proposition 3.1. *Warnings and specific enforcement cannot co-exist in an optimal policy for crime c: If* $\pi^*(c) < 1$, *then* $x^*(c) = 0$.

See the Appendix for the formal proof, but the intuition is straightforward. If any warnings were given for crime c and specific enforcement $x(c)$ was positive, the same deterrence could be attained with fewer resources expended by increasing $\pi(c)$ and decreasing $x(c)$.

Definition of minor crimes. *If* $x^*(c) = 0$, *then crime c is a minor crime.*

Since warnings exist for most traffic crimes, we can say they are minor crimes. The detection probability for all minor crimes is $p(0, y^*)$. Having chosen $p(0, y^*)$, for minor crimes the local government can also choose $\pi(c)$.

Proposition 3.2. *For minor crimes, the optimal expected fine is equal to the external harm imposed:* $z^*(c) = c$, *where* $z^*(c) = \pi^*(c) \times p(0, y^*) \times F(c)$. *If* $p(0, y^*)F(c) > c$, *then* $\pi^*(c) < 1$.

See the Appendix for the proof. Intuitively, the general detection probability $p(0, y)$ takes into account the degrees of harm imposed on society by many different crimes simultaneously. When fine amounts are fixed by a higher level of government, if all detected criminals were punished it is possible that for minor crimes the expected cost of breaking the law may exceed the harm imposed. If this occurs, the optimal way to reduce the expected fine for crime c is to reduce the probability of imposing a fine on a detected criminal.

3.2.3 Discussion

When traffic stops detect other crimes, the enforcement authority stops more drivers for traffic violations than would be optimal if traffic infractions were the only crimes being detected. To see this, compare the expected fine for crime c under general enforcement with the expected fine when traffic stops do not detect other crimes (given in equation (3.2)):

$$
\begin{aligned}
\pi^*(c)p(0,y^*)F(c) &= c \\
p(x^*)F(c) &= c - \frac{1}{\frac{\partial G(p(x^*)F(c))}{\partial x}} \\
\Longrightarrow \pi^*(c)p(0,y^*) &> p(x^*)
\end{aligned}
$$

For the same c, the expected fine is higher under general enforcement. Therefore for a given fine $F(c)$, the probability of detection must be higher. Intuitively, because traffic stops detect other crimes the government stops more drivers than it would otherwise. Because of this and the fixed fine schedule, some stopped drivers should be released without a fine to avoid making the expected penalty for traffic violations too high.

Importantly, the probability of receiving a ticket $\pi(c)$ depends only on the degree of harm imposed. This implies that warnings can be given within otherwise identical groups of drivers. There need not be other factors, observed by police but not by the researcher, to explain why some drivers are warned and others ticketed. In the context of the model, the police exercise discretion by randomly choosing some drivers for warnings.

This contrasts with previous research on the use of discretion by government agents (such as prosecutors or police), where information that is private to those agents plays a central role. For instance, in Reinganum's (1988) model prosecutor's information about the strength of the case is the key variable in their decision of whether to dismiss the case, offer a plea-bargain, or take the case to trial. Shavell (2007) models the use of discretion as government agents basing their decisions on variables which they observe but which cannot be included in the law. A tradeoff results because given discretion, agents can base their decisions on factors which society views as relevant (such as extenuating circumstances or remorse), and also on factors which the agents care about but society has deemed irrelevant (such as race). Reinganum (1988) points out a different tradeoff by showing that restricting prosecutors to offer defendants charged with the same crime the same plea-bargain sentence is more likely to improve social welfare when the proportion of guilty defendants is higher.

Certainly there are potential costs of police discretion in ticketing that are not included in my model. For example, the implication that warnings exist within identical groups of drivers violates horizontal equity. People may have a preference for horizontal equity and therefore view this as undesirable. Using Shavell's (2007) framework, another cost of discretion in traffic ticketing is that some decisions may be made on the basis of characteristics such as race or gender which are not relevant under the law. Although I do not explicitly model these costs, my analysis suggests that discretion in ticketing can be desirable simply if the benefit from detecting other crimes outweighs the various costs of police discretion.

However, the model does not rule out a more active role for police discretion which takes into account information about the violation observed by the officer. Consider the traffic violation "Failure to Stop", which is punished in Massachusetts with a \$50 fine. Depending on the specific context, failing to stop can be relatively harmless or quite dangerous. If these offenses with different levels of harm are all minor crimes, they are all detected with the same probability $p(0, y^*)$. In this case, allowing the police to issue warnings more frequently for offenses which they think are less dangerous will avoid over-deterrence for less serious violations. Therefore, police discretion in giving warnings can be useful if the enforcement authority can choose a schedule of fines, but cannot perfectly condition the fines on the severity of the offense.

I have demonstrated that even if people only respond to the expected cost of breaking the law, when traffic stops detect other crimes it may be optimal to give warnings. A plausible alternative rationalization for warnings is simply that warnings have some deterrence value. For instance, receiving a warning might make a driver less likely to commit traffic violations in the future. If warnings have a deterrent value, we might expect warnings to be widespread in contexts where other crimes are not detected. Consider the case of automated traffic enforcement cameras, which take photographs of the license plates of vehicles which run red lights or exceed the speed limit. Automated cameras make for a clean comparison to traffic stops; both detect traffic violations, but only traffic stops can detect other crimes.

Speed cameras are widely used in the U.K., and while drivers may contest the fines in court, no warnings are issued.[15] A speed camera program in Belgium tickets all drivers who exceed a certain threshold (Eeckhout, Persico, and Todd 2005).

According to the Insurance Institute for Highway Safety, red light cameras are authorized by state law and in use in 16 U.S. states, and speed cameras are used in 9 states.[16] Colorado's speed camera program is the only one which issues any

[15] See the Department for Transport's website for a summary of the use of cameras at http://www.dft.gov.uk/pgr/roadsafety. An example of an effectiveness study is Christie et al. (2003).

[16] The states which use speed cameras are Arizona, California, Colorado, Illinois, Maryland, Mas-

warnings. In Colorado, first-time offenders caught exceeding the speed limit by less than 10 miles-per-hour are warned. This appears to be the sole exception. Clearly, warnings are much less prevalent when red light running and speed limits are enforced by cameras. This suggests that enforcement authorities see warnings as having little or no deterrence value.

3.3 Data on tickets and warnings

In Massachusetts, police officers have legal discretion over whether to ticket a stopped driver or let him go with a warning.[17] When a warning is given, the officer fills out the same information about the stop on a citation form, but checks a box at the bottom of the form. The data contain the information from the citation form for 166,368 traffic stops which occurred in Massachusetts in April and May of 2001. The sample is each traffic stop where a citation form was completed.

Some key variables in the data are the date and time of the traffic stop, the town where the stop was made, the driver's hometown, and the reason for the stop. For speeding stops, both the speed limit and the miles-per-hour over the limit the driver was charged are available. The driver's age, race, and gender are also included. Finally, the police agency which made the stop (Boston, state, or local police) is recorded.[18]

Table (3.1) shows that warnings are very prevalent in traffic stops in Massachusetts. The state police ticket about 73% of drivers they stop for speeding, and local police ticket about 40% of stopped speeders. Table (3.2) shows that warnings are also prevalent within specific categories of offenses other than speeding. For example, local police only ticketed about 37% of drivers who were stopped for the offense "Failure to Stop". This "catch and release" policy is consistent with the general enforcement model where fine amounts are fixed. If traffic stops did not detect other crimes, the high frequency of warnings would be difficult to reconcile with efficient enforcement

sachusetts, Ohio, Oregon, Tennessee, plus Washington D.C.. States which authorize and use red light cameras are the states above except for Massachusetts and Ohio, plus Delaware, Georgia, New Jersey, New York, North Carolina, Pennsylvania, Rhode Island, Texas, and Washington. The IIHS explains that cameras are sometimes used in municipalities in states where cameras have not been regulated by state statute. This information is from www.iihs.org/laws/automated_enforcement.aspx and www.iihs.org/research/topics/auto_enforce_cities.html.

[17] Massachusetts state law, Chapter 90C, section 3.

[18] There are 350 local police agencies, one for each of the 350 towns and cities in Massachusetts (excluding Boston). There are approximately 50 different units of the state police (including various troops, headquarters, and special units), and the troops are distributed geographically throughout the state.

unless warnings significantly influence driving behavior.

Table (3.3) shows that searches are very rare, and that both ticketed and warned drivers are searched. This suggests that the prevalence of warnings does not result because the police stop one group of drivers who are suspicious and another group who are not. Table (3.4) shows that both local and state police stop many drivers for traveling less than 10 miles-per-hour over the speed limit. The ulterior motive of detecting other crimes can explain why so many drivers are stopped for such minor violations. Importantly, warnings exist within groups of drivers who committed similar violations, as the general enforcement model predicts.

3.4 The Revenue from Tickets

Table (3.1) reveals that on average, out-of-town drivers stopped for speeding were more likely to receive a ticket. Table (3.4) shows that this out-of-town penalty exists within groups of drivers who violated the speed limit by similar amounts. However, there appears to be no such penalty for other types of violations. This section uses the general enforcement framework to analyze the treatment of out-of-town drivers. After incorporating a concern for ticket revenue into the model, the general enforcement framework can explain why out-of-town drivers receive *any* warnings, and why they are less likely to be warned. Towns in Massachusetts may consider ticket revenue when designing enforcement policy. By state law, one-half of the revenue from traffic tickets issued by local and state police goes to the treasury of the town or city where the violation occurred.[19]

3.4.1 Policy towards out-of-town drivers: Theory

After a traffic stop, the police can determine which drivers are from out-of-town by looking at the driver's license. This allows police to use their discretion differently for these drivers. Because the police don't acquire this information until after the driver has been stopped, the general enforcement probability of stop, $p(0, y^*)$, is the same for in-town and out-of-town drivers.[20]

Assume that out-of-town drivers have the same distribution of benefits $G(b)$ as in-town drivers.[21] Also, assume that in-town residents incur the same cost c from out-of-town traffic violators. Once $p(0, y^*)$ has been chosen, the enforcement authority's

[19] Chapter 280, section 2. Available at http://www.mass.gov/legis/laws/mgl/280-2.htm

[20] There is no numbering scheme on Massachusetts license plates which identifies the town where the car is registered.

[21] This assumption does not affect any qualitative conclusions.

problem is the same for each crime c.[22] Denote the ticketing probability for out-of-town drivers by $\pi^n(c)$. What is the local government's objective with respect to out-of-town drivers? One possibility is that the local government chooses $\pi^n(c)$ to minimize the external costs imposed on residents by out-of-town violators while ignoring out-of-town driver's utility:

$$\max_{\pi^n(c)} -c[1 - G(\pi^n(c) \times p(0, y^*) \times F(c))] \tag{3.4}$$

The solution to (3.4) is $\pi^n(c) = 1$ for all c. This deters the maximum number of out-of-town drivers from committing crime c for a given probability of stop and fine amount. Clearly, the data are not consistent with this prediction because out-of-town drivers are frequently given warnings. I now show that when fine revenue is a consideration, more realistic predictions for the treatment of out-of-town drivers are obtained.

In the general enforcement model (and the specific enforcement model), the revenue from fines was treated as a welfare neutral transfer of wealth. For this reason, fine revenue did not influence the optimal policy. However, when the local government tickets an out-of-town driver, this represents a net transfer of wealth to the community. Local social welfare must increase because the fine came from a non-resident. Therefore, ticket revenue from out-of-town drivers should add directly into the local social welfare function.

Let the expected fine for out-of-town drivers be $z^n(c) = \pi^n(c) \times p(0, y^*) \times F(c)$. When including ticket revenue but not out-of-town utility in the welfare function, the local government's enforcement problem (for each c) for out-of-town drivers becomes:

$$\max_{\pi^n(c)} z^n[1 - G(z^n)] - c[1 - G(z^n)] \tag{3.5}$$

The first term in (3.5) is out-of-town ticket revenue, and the second term is the total cost imposed on residents by out-of-town drivers who commit crime c. The solution to this problem is:

$$\pi^{n*}(c) \times p(0, y^*) \times F(c) = c + \frac{1 - G(z^n)}{g(z^n)} \tag{3.6}$$

The expected fine for out-of-town drivers is just the sum of the external cost c and the inverse hazard rate $\frac{1-G(z^n)}{g(z^n)}$. Both terms are positive. The analogous solution

[22] Recall that the mass of individuals who might commit crime c, written as $s(c)$, becomes a constant and so does not affect the optimal policy for each c.

for in-town drivers is given by Proposition 3.2: $\pi^*(c) \times p(0, y^*) \times F(c) = c$. Therefore the gap between the expected fines for out-of-towners and in-towners is:

$$z^n(c) - z(c) = \frac{1 - G(z^n)}{g(z^n)} \tag{3.7}$$

Because it cares about ticket revenue, the government should impose a higher expected fine for crime c on out-of-town drivers. Also, because the detection probability and statutory fine amounts are the same for in and out-of-town drivers, we see that:

$$p(0, y^*)F(c) \times [\pi^n(c) - \pi(c)] = \frac{1 - G(z^n)}{g(z^n)} \tag{3.8}$$

Equation (3.8) says the government should set a higher probability of ticketing stopped out-of-town drivers. In this way, discrimination against out-of-town drivers can be part of an efficient enforcement policy when fine revenue is taken into account. Makowsky and Stratmann (2007) show that the Massachusetts police exercise some discretion in choosing fine amounts as well. The departures from statutory fines are small but statistically significant. Assuming that police do manipulate fine amounts to some degree, according to equation (3.7) the police should assign higher fine amounts to out-of-town drivers.

However, any observed disparity in ticketing between in and out-of-town drivers may not necessarily be due to a concern for revenue. The police might treat out-of-town drivers more harshly for a number of reasons unrelated to ticket revenue. For instance, the local police might perceive a lower chance of being sanctioned for writing unjustified tickets to out-of-town drivers.

The model and the data together suggest a way to test for the revenue motive. For drivers ticketed by local police, Figure (3.1) shows the empirical probability that a speeding violation was faster than a given miles-per-hour. Call this probability $1 - H(mph)$, where $H(mph)$ is the cumulative distribution function for miles-per-hour over the limit. Notice that $1 - H(mph)$ is very similar for in-town and out-of-town drivers, and that above 15 miles-per-hour, $1 - H(mph)$ is concave. Assume that $H(mph)$ reflects the shape of the unknown distribution of driver's benefits from speeding $G(b)$, so that above some $\tilde{b}$, $1 - G(b)$ is concave. If $1 - G(b)$ is concave, the following property holds:[23]

Monotone Hazard Rate. *The hazard rate* $\frac{g(z)}{1-G(z)}$ *is increasing in the expected fine* z.

[23] See Fudenberg and Tirole 1991, page 267 for details.

This property holds for many distributions, including the normal, exponential, logistic, and uniform. When this holds, the inverse hazard rate $\frac{1-G(z)}{g(z)}$ is decreasing in z. Since the graph of $1 - H(mph)$ was only concave above a certain value of miles-per-hour, we suspect that the hazard rate condition on $G(b)$ only holds above some given benefit.

When the inverse hazard rate is decreasing, a marginal increase in the expected fine z will lead to relatively more criminals being deterred as z gets larger. Therefore, the inverse hazard term in equation (3.8) is predicted to be smaller for more harmful offenses which are punished with a higher expected fine. Thus, if discrimination against out-of-town drivers is motivated by revenue, the gap in the expected fines between in-town and out-of-town drivers is predicted to shrink for more serious violations. Intuitively, the transfer of revenue becomes less important. If the police discriminate against out-of-town drivers because of an animus or a lower chance of retribution for unjustified tickets, there is no particular reason why this pattern should hold.

3.4.2 Policy towards out-of-town drivers: Empirical evidence

In this section I estimate OLS specifications for the probability that a stopped driver receives a ticket, and Tobit models of the fine amount charged to the stopped driver. In the Tobit specifications, I set the fine amount to zero for drivers who received warnings. This is consistent with the theoretical formulation of warnings as a zero fine for breaking the law. The purpose of these specifications is to determine if there is an out-of-town penalty, in the form of a higher chance of being ticketed, a higher fine amount charged, or some combination of the two.

When estimating the out-of-town penalty, an important issue to consider is the possibility that out-of-town drivers might commit more dangerous violations. This could result in a positive estimate of the out-of-town penalty that is due to omitted variables, even when miles-per-hour over the limit is included as a control. I address this problem in two ways. First, because most of the 350 towns in Massachusetts are geographically small, it is unlikely that in each town there are two groups of drivers, one which only drives in-town and the other which ventures out-of-town. As such, the distinction between in-town and out-of-town drivers should result only from differences in where drivers happened to be stopped.

To capture this notion, I include fixed effects for each hometown in a linear probability model for receiving a ticket after being stopped. Suppose that out-of-town drivers tend to be from certain places, such as Worcester, and that drivers from

Worcester commit more dangerous offenses. The inclusion of hometown fixed effects will correct for this simple omitted variable problem by comparing how Worcester drivers are treated in Worcester versus everywhere else.

As another approach, I restrict the sample to traffic stops which occurred within 10 miles of the drivers hometown.[24] Drivers from far away might commit more dangerous offenses because they are unfamiliar with the local roads. Shrinking the sample in this way leaves out those drivers. Because the hometown fixed effects and the distance restriction do not make sense for out-of-state drivers, I exclude out-of-state drivers from the analysis.

In addition to an indicator equal to 1 if the driver is out-of-town, the explanatory variables in each specification are as follows: Miles-per-hour over the limit, the speed limit itself, dummies for day of week and time of day (morning, afternoon, evening, predawn), and driver's age, race (black versus white) and gender. Column 1 in Table (3.5) shows that conditional on the explanatory variables, local police were about 7 percentage points more likely to ticket out-of-town drivers stopped for speeding. When including hometown fixed effects and restricting the sample to drivers stopped within 10 miles of home (Column 4), the out-of-town penalty falls to 4 percentage points. This is an 11.4% penalty in the chance of being ticketed, because local police ticketed 35 percent of stopped in-town speeders. A high proportion (about 72%) of stops by local police satisfied the within 10 mile restriction. The estimated out-of-town penalty for drivers stopped by the state police is 1 percentage point, an effect which is not statistically different from zero.

Table (3.6) shows the results of estimating the out-of-town penalty with a Tobit specification for the fine amount charged to the stopped driver. I report the effect of being out-of-town on the expected fine amount charged (evaluated at the means of the explanatory variables). This marginal effect measures the out-of-town penalty as the combination of a higher probability of getting a ticket and a potentially higher fine amount, conditional on the fine being positive. In expectation, the local police charged stopped out-of-town drivers 8.4 dollars (standard error of 0.49) more than similar stopped in-town drivers. This is about 7.2% of the average speeding fine charged to in-town drivers (116 dollars). For the whole sample, the state police charged an additional 3.13 dollars to out-of-town drivers. This effect falls to 1.76 dollars and is not statistically significant when the sample is restricted to drivers stopped within 10 miles of home.

[24]Distance in miles from hometown to town where stopped was calculated using the longitude and latitude of the towns as published in the U.S. Census Gazetteer (www.census.gov/cgi-bin/gazetteer).

The above empirical evidence suggests that an out-of-town penalty exists for drivers stopped by local police, and that the penalty does not result because out-of-town drivers commit more dangerous violations. Yet this is not sufficient to conclude that the police are motivated by ticket revenue. Consider Figure (3.2), which shows the proportions of in-town and out-of-town drivers who received tickets at different values of miles-per-hour. Visually, the out-of-town penalty shrinks for more dangerous offenses. This is consistent with the model's prediction for how the out-of-town penalty should change if the penalty is motivated by a concern for ticket revenue. I now estimate some specifications to determine if this pattern holds up when other relevant factors are controlled for.

To determine if the out-of-town penalty shrinks for more serious offenses, a triple interaction term must be estimated. To do this, I include the following additional explanatory variables: First, an indicator for if the violation was above 20 miles-per-hour over the limit (called 1[mph >20]). I used 20 miles-per-hour based on the pattern shown in Figure (3.2). Second, the interaction of 1[mph >20] and the indicator for out-of-town. Third, the interaction of miles-per-hour over the limit and 1[mph >20], and fourth the interaction of miles-per-hour and out-of-town. Finally, the triple interaction term of interest is out-of-town $\times$ mph $\times$ 1[mph >20].

In this way, the triple interaction term captures how the effect of miles-per-hour on the outcome is different for out-of-town drivers once the miles-per-hour is above 20. If the pattern seen in Figure (3.2) holds up, the triple interaction term will have a negative effect on the outcome because the out-of-town penalty is initially positive.

Estimating the marginal effect of the triple interaction is a tedious exercise in a non-linear model such as a probit, because the effect is a combination of non-linear functions of all the coefficients.[25] For the reported probit specification, I computed the average of the marginal effects by taking discrete differences instead of derivatives, and I estimated the standard errors of the average marginal effects by bootstrapping with 100 replications.[26]

I estimated a probit for the probability of being ticketed, and OLS models for the probability of getting a ticket and the fine amount, conditional on the fine being positive. Table (3.7) shows that the results provide a consistent picture of a shrinking out-of-town penalty at high levels of miles-per-hour over the limit. The probit model for the probability of being ticketed shows that each mile-per-hour above 20 is associ-

[25]See Cornelissen and Sonderhof (2008) for an example.

[26]My attempts to compute the effect of the triple interaction term for a Tobit model were not successful. I am working on incorporating such an estimate into the paper.

ated with a 1 percentage point decline in the out-of-town penalty in the probability of receiving a ticket (standard error of 0.5 percentage point). The OLS specification for the fine amount shows that for drivers who received fines, each mile-per-hour above 20 is associated with a reduction of 1.69 dollars (s.e.=0.92) in the out-of-town penalty. This evidence suggests that discrimination against out-of-town drivers by the local police is motivated by ticket revenue. If the police simply disliked out-of-town drivers (or accorded in-town drivers preferential treatment), the out-of-town penalty would be the same across different degrees of violations.

3.5 Conclusion

This paper showed that under certain conditions, releasing criminals with only a warning instead of requiring them to pay a fine can be an aspect of an efficient enforcement policy. The result holds even if the behavior of potential criminals is unaffected by warnings. The required conditions are that the enforcement technology detects other crimes, and that fine amounts are either fixed or cannot be finely adjusted. Traffic enforcement by individual police officers satisfies these conditions, so we can infer that the widespread policy of giving warnings rather than fines to drivers who were stopped for breaking traffic laws is not likely to be grossly inefficient. Because police officers stop additional drivers based on the ulterior motive of detecting other crimes, it is beneficial to society to let some of those drivers go with only a warning.

Empirically, stopped out-of-town drivers in Massachusetts were more likely to receive a ticket than similar in-town drivers. I used a subtle prediction of my model to determine if this discrimination is indeed motivated by revenue. I conclude that the revenue motive exists because the penalty imposed on out-of-town drivers disappeared for more serious violations. Therefore by a much different method, I arrive at the same conclusion on this point as do Makowsky and Stratmann (2008). However, my paper also points out that this particular revenue motive does not conflict with effective law enforcement, if one thinks of the revenue from out-of-town traffic tickets as a net transfer of wealth to the local community.

3.6 Appendix

Assume that $y^* > 0$, so that all crimes are detected with at least probability $p(0, y^*)$.

3.6.1 Proof of Proposition 3.1

Suppose that $\pi^*(c) < 1$ and $x^*(c) > 0$. Then reduce $x^*(c)$ and increase $\pi^*(c)$ to hold the expected fine $z(c) = \pi^*(c) \times p(x^*(c), y^*) \times F(c)$ constant. Then the same number of individuals commit crime c but fewer resources are used, a contradiction.

3.6.2 Proof of Proposition 3.2

For minor crimes where $x^*(c) = 0$, the government takes as given $p(0, y^*)$ and $F(c)$. Therefore for a minor crime c, the government's problem is:

$$\max_{\pi(c)} \int_{\pi(c) \times p(0,y^*) \times F(c)}^{\infty} (b - c) dG(b) s(c)$$

The mass of individuals who might commit crime c is $s(c)$. Since $s(c)$ is just a constant it does not affect the solution and therefore can be dropped. Where $z = \pi(c) \times p(0, y^*) \times F(c)$, the first order condition is:

$$-[\pi^*(c)p(0, y^*)F(c) - c]g(z)d\pi(c)p(0, y^*)F(c) = 0$$

This simplifies to $\pi^*(c) \times p(0, y^*) \times F(c) = c$. Suppose $\pi^*(c) = 1$ and $p(0, y^*) \times F(c) > c$. Then at no cost, social welfare can be increased by reducing $\pi^*(c)$ so that an individual whose benefit b exceeds the social cost c will be induced to commit the crime.

3.7 Tables and Figures

Table 3.1: Proportion of traffic stops resulting in a ticket.

		Proportion Ticketed		Number of stops	
		Other offenses	Speeding	Other offenses	Speeding
State Police	Out-of-town	0.55	0.75	10,323	17,793
	In-town	0.54	0.68	3,291	2,177
Local Police	Out-of-town	0.46	0.43	26,355	36,162
	In-town	0.46	0.35	13,212	17,021
Boston Police	Out-of-town	0.49	0.49	6,093	2,879
	In-town	0.48	0.50	2,879	3,225

Excludes out-of-state drivers.

Table 3.2: Proportion of stops for selected other offenses resulting in a ticket.

		Proportion Ticketed			Number of stops		
		(1)	(2)	(3)	(1)	(2)	(3)
State Police	Out-of-town	0.56	0.61	0.84	844	1,061	1,176
	In-town	0.57	0.58	0.84	388	383	346
Local Police	Out-of-town	0.37	0.58	0.86	10,015	3,010	1,218
	In-town	0.37	0.56	0.86	5,404	1,393	433
Boston Police	Out-of-town	0.44	0.61	0.58	3,414	339	50
	In-town	0.42	0.61	0.55	3,115	276	49

(1) Failure to Stop (2) No Inspection Sticker (3) Seat Belt Violation
Excludes out-of-state drivers.

Table 3.3: Number of searches.

		Warned		Ticketed	
		Searches	Total Stops	Searches	Total Stops
Local Police	Other Offenses	80	20,545	181	17,357
	Speeding	44	30,279	76	21,335
State Police	Other Offenses	57	5,835	92	6,969
	Speeding	12	5,512	89	17,440

Table 3.4: Proportion ticketed by MPH over speed limit.

		Proportion Ticketed		Number of stops	
	MPH over limit	Out-of-town	In-town	Out-of-town	In-town
Local Police	1-10	0.40	0.31	7,328	3,372
	11-15	0.28	0.21	15,064	7,571
	16-20	0.55	0.48	9,945	4,326
	21-25	0.77	0.70	2,631	1,151
	26 +	0.91	0.89	856	389
State Police	1-10	0.54	0.49	5,905	795
	11-15	0.77	0.61	3,870	443
	16-20	0.92	0.87	3,978	423
	21-25	0.97	0.96	2,336	246
	26 +	0.98	0.97	1,229	197

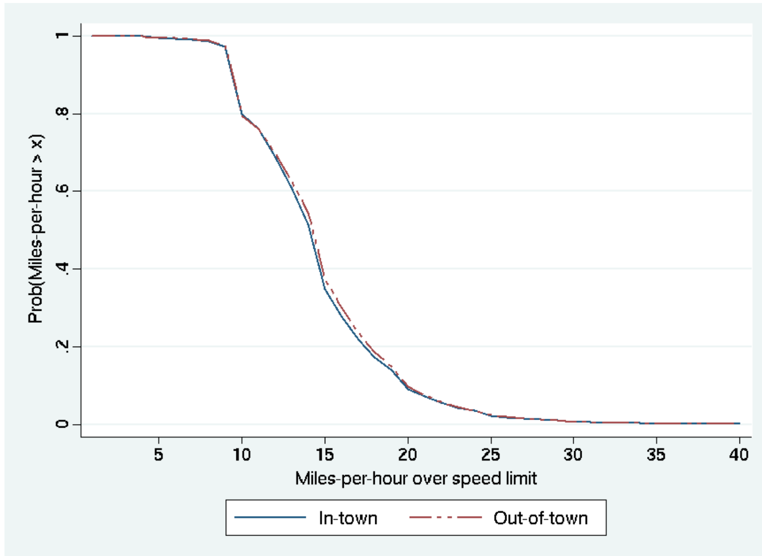


Figure 3.1: Probability that violation was faster than MPH (local police).

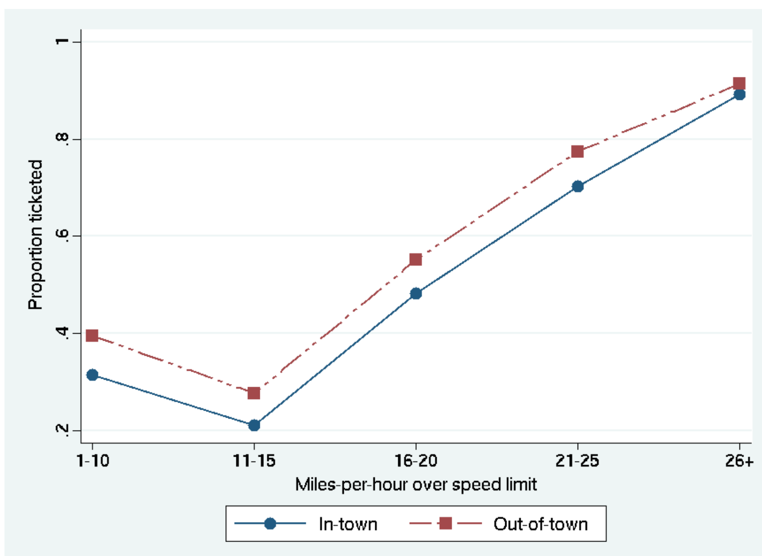


Figure 3.2: Proportion ticketed by MPH over speed limit (local police).

Table 3.5: Out-of-town penalty in the probability of being ticketed for speeding (OLS).

Ticketed (Yes=1)	Local Police 1	2	3	4	State Police 5
Out-of-town	0.07**	0.07**	0.04**	0.04**	0.01
	(0.004)	(0.005)	(0.005)	(0.005)	(0.009)
Linear mph	Yes	Yes	Yes	Yes	Yes
Hometown fixed effects	No	No	Yes	Yes	No
10 miles from home	No	Yes	No	Yes	No
Driver demographics	Yes	Yes	Yes	Yes	Yes
Day and time	Yes	Yes	Yes	Yes	Yes
Observations	52,398	37,394	52,398	37,394	19,314

Dependent variable: Ticketed after stop (Yes=1, No=0)

Heteroskedastic robust standard errors, **p<0.05

Table 3.6: Out-of-town penalty in fine amount charged (Tobit).

Fine Amount (warning=0)	Local Police 1	2	State Police 3	4
Out-of-town	8.40**	8.25**	3.13**	1.76
(effect on E[fine | X)	(0.499)	(0.553)	(1.56)	(1.81)
Linear mph	Yes	Yes	Yes	Yes
Hometown fixed effects	No	No	No	No
10 miles from home	No	Yes	No	Yes
Driver demographics	Yes	Yes	Yes	Yes
Day and time	Yes	Yes	Yes	Yes
Observations	46,875	33,889	16,736	5,921

Dependent variable: Fine amount charged (warning=0).

Marginal effect of out-of-town on fine amount, at means. **p<0.05

Table 3.7: Shrinking out-of-town penalty in speeding tickets.

Local police only	Prob(Ticketed=1) OLS	Probit	Fine amount OLS
Out-of-town×mph×1[mph > 20]	-0.005^{*}	-0.01^{*}	-1.69^{*}
	(0.0028)	(0.0056)	(0.916)
1[mph > 20]	0.378^{**}	0.157^{**}	-42.70^{**}
	(0.054)	(0.030)	(17.98)
Driver demographics	Yes	Yes	Yes
Day and time	Yes	Yes	Yes
Observations	52,398	52,398	15,642

Probit: Marginal effects on the probability of being ticketed.

Probit standard errors are bootstrapped with 100 replications.

3.8 References

Baicker, Katherine and Jacobson, Mireille. "Finders keepers: Forfeiture laws, policing incentives, and local budgets", *Journal of Public Economics* 2007.

Becker, Gary. "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, 1968.

Craig, Tim. "Legislators' Short View Killed Va. Driving Fees", *The Washington Post*, March 8 2008.

Cornelissen, Thomas and Sonderhof, Katja. "Marginal effects in the probit model with a triple dummy variable interaction term," Discussion Paper 386, Leibniz Universitat Hannover, January 2008.

Dedman, Bill and Latour, Francie. "Race, sex, and age drive ticketing", *The Boston Globe*, Jul 20 2003.

Dittman, Ingolf. "The Optimal Use of Fines and Imprisonment if Governments Do Not Maximize Welfare", *Journal of Public Economic Theory*, 2006.

Durose, Matthew; Erica Smith; Patrick Langan. "Contacts between Police and the Public, 2005". U.S. Department of Justice, Bureau of Justice Statistics, Special Report. April 2007.

Farrell, Amy; Dean McDevitt; Lisa Bailey; Carsten Andresen; Erica Pierce. "Massachusetts Racial and Gender Profiling Final Report: Executive Summary", Northeastern University Institute on Race and Justice, May 2004.

Garrett, Thomas and Wagner, Gary. "Red Ink in the Rearview Mirror: Local Fiscal Conditions and the Issuance of Traffic Tickets", Forthcoming, *Journal of Law and Economics*, 2008.

Harrington, Winston. "Enforcement Leverage when Penalties are Restricted", *Journal of Public Economics* 1988.

Harris, David. " 'Driving while Black' and All Other Traffic Offenses: The Supreme Court and Pretextual Traffic Stops", *Journal of Criminal Law and Criminology*, Winter 1997.

Howlett, Debbie. "Cities get lift from parking fines", *USA Today*, Apr. 13 2004.

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

Makowsky, Michael and Stratmann, Thomas. "Political Economy at Any Speed: What Determines Traffic Citations?", Forthcoming, *American Economic Review*, 2008.

Mookherjee, Dilip and Png, I. P. L. "Monitoring vis-a-vis Investigation in Enforcement of Law", *American Economic Review*, June 1992.

National Highway Traffic Safety Administration, Center for Statistics and Analysis. "2006 Annual Assessment of Motor Vehicle Crashes." Sept. 2007.

Polinsky, A. Mitchell and Shavell, Steven. "The Optimal Use of Fines and Imprisonment", *Journal of Public Economics*, 1984.

Ramirez, Deborah; McDevitt, Jack; Farrell, Amy. "A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned", Memo, Northeastern University, 2000.

Reinganum, Jennifer. "Plea Bargaining and Prosecutorial Discretion", *American Economic Review* 1988.

Rowe, Brian. "Gender Bias in the Enforcement of Traffic Laws: Evidence based on a new empirical test", working paper, 2009.

Shavell, Steven. "Specific versus General Enforcement of Law", *Journal of Political Economy*, 1991.

Shavell, Steven. "Optimal Discretion in the Application of Rules", *American Law and Economics Review*, 2007.

Vitello, Paul. "The Taxman Hits, in the Guise of a Traffic Cop", *The New York Times*, Jul 15 2007.

Whren et al. v. United States, 517 U.S. 806 (1996).

CHAPTER IV

The Effect of Smoking in Young Adulthood on Smoking Later in Life: Evidence Based on the Vietnam Era Draft Lottery

4.1 Introduction

One of the most well-established relationships in the smoking literature is that smoking during youth and young adulthood is a strong predictor of smoking later in life (e.g., Merline et al. 2004). A related well-known finding is that youth and young adults who manage to quit smoking are able to avoid or substantially mitigate lifetime health consequences (U.S. Department of Health and Human Services 1990). Together, these facts have motivated health policymakers to focus much of their prevention and cessation efforts on young populations (Glied 2003).

It is unclear, however, to what extent these well-known findings represent causal relationships, due to potentially omitted variables. Smoking during youth and young adulthood may be highly correlated with smoking later in life not only because of the addictive power of cigarettes but also because of imperfectly measured contextual factors (e.g., peers and family members, occupational and personal stresses, etc.) that influence smoking and are correlated over time. Analogously, young people who quit smoking may avoid long-term health consequences not only because of quitting but also because of imperfectly measured factors (e.g., motivation to become healthier) that influence both propensity to quit and lifetime health. In other words, smoking behavior during youth and young adulthood is far from exogenously determined, so outcomes in later adulthood should not necessarily be attributed to that earlier behavior.

In this context, a well-identified causal estimate requires exogenous variation in smoking during young adulthood. Randomized controlled studies of smoking cessation interventions offer one possible approach, but they generally do not follow subjects beyond a year or two. On the other hand, observational studies generally do not have truly exogenous sources of variation by which to estimate definitive causal relationships. Furthermore, observational studies that do have arguably exogenous variation still may not be able to separate short-term effects from long-term effects, if the source of variation remains present, or highly correlated, over time (e.g. a cigarette tax increase that remains in effect, as such increases typically do).

The Vietnam draft lottery, which randomly assigned eligibility for the draft based on birth dates, offers a unique opportunity to investigate this issue. In essence, it provides a population-level randomized experiment for which both a short-term and long-term follow-up are possible. By receiving a low draft number and thus becoming more likely to serve in the military, certain men became more likely to be exposed to a potentially large, positive shock to smoking behavior while they were 19-22 years of age. All men who served in the military during the Vietnam era had access to tax-free cigarettes at military bases and commissaries, and men in combat received free cigarettes in their rations (U.S. Public Health Services, 1989, p. 425). In addition, the effect of these reduced prices may have been enhanced by the close proximity of peers who smoked and the stress of potential combat.

Using the draft lottery and pooled cross-sectional data from the National Health Interview Survey (NHIS), we construct instrumental variable estimates of the effects of military service on smoking behavior at different time points later in life. We estimate two-stage least squares models in which the first stage predicts the probability of military service and the second stage predicts the probability of smoking. The key instrumental variable in the first stage is whether or not a man had a draft number (based on birth date) below or above the cut-off number that determined draft eligibility in that year. Using this analysis, we address two main questions. First, what was the effect of military service during the Vietnam era on smoking behavior as of several years after the war (1978-80), when the men in the relevant cohorts were 25-30 years old? Second, what was this effect approximately 20-25 years later (as of 1997-2005), when these men were 45-55 years old?

As of 1978-1980 we find that military service increased the probability of current smoking by 35 percentage points. This estimate is substantially larger than the corresponding ordinary least squares (OLS) estimate of 11 percentage points. The apparent downward bias in OLS suggests that men who served in the military had unobservable characteristics making them less likely to smoke than the general population, which may be related to the physical and mental standards for induction into the military.

By contrast, we find that this large short-term effect dissipated to being small and statistically insignificant as of 1997-2005, when the men were aged 45-55. This result is due to the simple fact that by this period, draft eligible men were no longer significantly more likely to smoke than non-eligible men. The IV estimate of the military service effect as of 1997-2005 is a relatively imprecise "zero", but point estimates of the effect of draft eligibility on smoking from two other later-period data sources (1987-1988 NHIS supplements and the 2002-2003 National Epidemiologic Survey on Alcohol and Related Conditions) are also close to zero and statistically insignificant.

Our findings suggest that the large exogenous increase in smoking early in adulthood was substantially attenuated by later adulthood. Before accepting this conclusion, however, we consider two alternative explanations for the disappearance of the smoking differential. First, there may have been differential attrition from the survey samples, due to mortality or other factors, between the two time periods. We find, however, that the proportion of men with draft eligible lottery numbers remained identical between the two samples, which is consistent with Angrist and Chen's (2007) findings with a much larger dataset. In addition, we show that even under extreme assumptions, differential mortality by draft-eligible smokers can account for at most a very small portion of the diminished smoking differential. Second, we consider the potential role of increased educational attainment among veterans. Based on the results of recent studies[1] we find that this factor can explain at most a small portion of the decline in the smoking differential.

The estimates produced by our empirical strategy represent the local average treatment effect (Angrist et al. 1996) of military service on smoking: the effect on smoking for those men induced to serve in the military as a result of being assigned to the draft-eligible group. Our results imply that for these "marginal smokers", cigarettes were not so addictive that at least 5 to 7 years of smoking during young adulthood was sufficient to induce smoking into middle and later adulthood. Furthermore, we find that as of 1997-2005 men with low draft numbers were not significantly worse on health measures, including self-reported health and lifetime prevalence of cancer.

In effect, this analysis represents a natural experiment in which certain men were randomized to be exposed to factors that made them much more likely to smoke for a number of years. The longer-term outcomes from this natural experiment are inconsistent with the widespread presumption that smoking in young adulthood strongly increases the likelihood of smoking later in adulthood, but consistent with the idea

[1] Angrist and Chen (2007), De Walque (2007), and Grimard and Parent (2007).

that young people who quit smoking can avoid serious long-term health consequences. As we discuss in Section 5, the extent to which these findings would generalize to other contexts (e.g., nonveterans, females, other cohorts, other countries) is unclear, but, as the first evidence based on effectively random assignment, the findings offer a new point of reference.

4.2 Background and related literature

In this section we review how the Vietnam draft lottery operated as a natural experiment, and discuss historical information related to cigarettes in the military. We then briefly review recent theory and evidence related to cigarette addiction, focusing on predictions about the links between current and future cigarette consumption.

4.2.1 The draft lottery, the war, and cigarettes

The Vietnam draft lotteries were held in each year from 1969 to 1972 to make induction into military service a fairer process. The lotteries randomly assigned a priority for induction based on date of birth by matching each birth date to a number from 1 to 365 (or 366 for leap years). Later in the year, after assessing its needs for manpower, the U.S. Department of Defense would choose a cut-off number for draft eligibility. Men with numbers below the cut-off were considered draft-eligible and were thereby more likely to end up serving, while men with numbers above the cut-off were not at risk for being drafted. Men from the cohorts we analyze were subjected to the draft lottery at age 19 or 20, and the typical term of service for these men was 2 years (Angrist 1990, Angrist and Chen 2007).

Previous studies have employed this natural experiment to look at the effects of military service on mortality, earnings, and other outcomes. In the first such study, Hearst, Newman, and Hulley (1986) used the draft lottery to estimate the effect of military service on mortality later in life. They found a small but statistically significant effect of service on total mortality, which was driven mainly by suicide and motor vehicle related mortality. As these results have bearing on potential attrition bias, we discuss them further when interpreting our results (Section 5). In another study using this natural experiment, Angrist (1990) estimated the effect of military service on lifetime earnings. He found a significant earnings penalty from military service for white veterans relative to similar white non-veterans.

Cigarette smoking may have been affected by military service during the Vietnam era for several reasons. First, men who served faced significantly reduced prices for cigarettes. Men in combat, or other places where cooking was not feasible, received C or K rations with free cigarettes. All servicemen could buy cigarettes at wholesale, tax-free prices at military bases and commissaries. We cannot determine the exact prices at these facilities, but we can infer that they would have been less than half of retail prices. In the U.S. in 1972, the average retail price of a pack of cigarettes was 40 cents and the average pre-tax wholesale price was 13 cents (USDA, 2007). Much of this difference is attributable to taxes: 8 cents per pack in federal taxes, and an average of 10 cents per pack in state taxes (Orzechowski and Walker 2006). Moreover, the price of cigarettes *relative to other goods* would have been especially low for servicemen who did not have access to many of the goods they would normally purchase as a civilian in the United States.

Second, peer effects may have augmented the effect of reduced prices, given the concentration of men at bases and other military facilities. Third, many servicemen in the Vietnam era were subjected to stressful, if not traumatic, experiences. For some men these experiences could have caused depression or anxiety, both of which are associated with an increased risk of smoking (Glassman et al. 1990, Winefield et al. 1989).

It is unclear how important this third factor would have been for the cohorts in our analysis. Major combat in Vietnam ended shortly after the first draft lottery was held in December 1969. Available evidence indicates that the proportion of men who faced combat, among those drafted in the 1969 to 1971 lotteries, was very small compared to earlier cohorts who participated in the war.[2] Nevertheless, the anxiety associated with the possibility of engaging in combat may have still played a role in these men's smoking behavior.

Multiple correlational studies have shown that veterans of the Vietnam era are significantly more likely to smoke than non-veteran males of the same ages (Stellman et al. 2000, Klevens et al. 1995). Furthermore, the lifetime risk of five types of smoking-related cancers is twice as high among the Veterans Administration population as among the general male population of the same age (Harris et al. 1989).

[2]The last major incursion by U.S. troops occurred in May-June 1970, and the last major confrontation for U.S. ground troops was the Battle of Fire Support Base Ripcord in July 1970. As draftees had a training period, these events would have occurred before men drafted in the lottery were sent to war. Starting in 1969, the U.S. steadily reduced the number of troops in Southeast Asia, and the Vietnam War officially ended in January 1973. Casualties similarly declined in this period, from 16,592 in 1968, to 11,616 in 1969, to 6,081 in 1970, to 2,357 in 1971, to 641 in 1972. (National Archives, http://www.archives.gov/research/vietnam-war/). Furthermore, of the 7,575,000 active duty Vietnam era servicemen, only 2,850,000 (38%) were sent to Southeast Asia (Baskir and Strauss 1978). Servicemen based outside of Southeast Asia would not have faced combat, but had the same opportunity to purchase tax-free cigarettes at military bases.

Our use of the draft lottery overcomes the problem of systematic selection into the military, which could in principle result in a correlation between military service and smoking even in the absence of any causal effect.

4.2.2 Theory and evidence related to cigarette addiction

The natural experiment of the Vietnam draft lottery offers an opportunity to evaluate how an exogenous change to smoking behavior in one time period persists over a long time span. Here we review what theoretical models of addiction and empirical studies have to say about this issue. While the natural experiment in our analysis does not enable us to distinguish between different economic theories of addiction, it provides new evidence related to the strength and persistence of addiction, whatever the underlying mechanism may be.

Under most economic models of addiction, a temporary price decrease would lead to at least some increase in contemporaneous smoking, simply because consumers respond to prices.[3] Also, in most economic models of addiction, the degree to which a short-term price decrease produces a long-term increase in smoking depends on the strength of the addiction, i.e., the degree to which present consumption increases the marginal utility of cigarettes consumed in the future.

Under a myopic model of addiction, a temporary fall in cigarette prices would result in both a short-term and a long-term increase in cigarette consumption. People simply increase their consumption of cigarettes in response to lower prices today, failing to anticipate that they will consume more cigarettes in the future as a result of their increased consumption today. However, there is some evidence against strictly myopic cigarette addiction, as several studies have found that consumption of cigarettes responds to expected future prices of cigarettes (Becker et al. 1994, Chaloupka 1991, Gruber and Koszegi 2001). Even if smokers do account for the future consequences of their present smoking decisions, it remains theoretically ambiguous whether a temporary price decrease would result in a long-term increase in cigarette consumption. In the rational addiction model (Becker and Murphy 1988), smokers take into account all the future consequences of their actions, and they discount the future in a time-consistent fashion. A key feature of the rational addiction model is the existence of unstable equilibrium levels of cigarette consumption. Thus under certain circumstances, a temporary price decrease could induce individuals

[3] In this discussion "price" can be thought of as encompassing monetary price as well as military service related factors which may affect the net benefit of current consumption (peer effects and stress).

to move from an unstable equilibrium where cigarette consumption is low to a new equilibrium where consumption is permanently higher (see Becker and Murphy 1988, p.692-693). However, in this model it is also possible that a temporary price decrease will leave long term consumption relatively unchanged.

A prominent competing theory is that individuals may be forward looking in their smoking behavior, yet discount the future in a time-inconsistent fashion (Gruber and Koszegi 2001). With time-inconsistent preferences, individuals may decide to smoke more today in response to a temporary price decrease, while they plan to quit tomorrow when prices go back up. Higher prices tomorrow may serve as a commitment device in that they will help smokers to bring their consumption levels back down to what they were before the price cut. In fact, if higher future prices are a strong enough commitment device, a temporary price cut may lead to an especially large short term increase in smoking, as forward looking individuals anticipate being able to quit when prices rise (see Gruber and Koszegi 2001, p.1284). Of course, a long term increase in cigarette consumption may still result if individuals are unable to follow through on their plans to reduce smoking in the future, as is typical in models where preferences are time-inconsistent.

Two recent empirical studies pertain to the issue of how changes to smoking during one time period persist over the long-term. Gruber and Zinman (2001) and Glied (2002) estimate how cigarette taxes during one's youth are associated with smoking later in life, controlling for current taxes in adulthood. Gruber and Zinman (2001) find that, for pregnant women older than 24, the overall elasticity of smoking is -0.22 and the participation elasticity is -0.08, with respect to cigarette taxes in the women's state of residence when they were 14-17 years old (which is proxied by state of birth). They note that these elasticities are about 25-50% of same-period tax elasticities typically estimated for youths. Glied (2002) finds, using a general sample of adults (in the National Longitudinal Survey of Youth), that as of age 44, the participation elasticity with respect to taxes at age 14-17 is not statistically different from zero for women and -0.2 for men. These elasticities are significantly smaller than Glied's estimates of contemporaneous tax elasticities. Thus, both studies suggest that the effects of cigarette taxes during youth dissipate substantially over the course of adulthood. A limitation of this approach is that state taxes may be correlated with unmeasured factors which affect smoking, such as anti-smoking sentiment. Also, it is difficult to determine which empirical specification properly captures the timing of the effects of taxes, given that typically cigarette tax increases are not one-time exposures but rather remain in place until the next increase. Our approach is different in that we trace out the impact of a single randomly induced exposure over time.

Continuing to improve understanding of how smoking behavior during youth and young adulthood affects longer-run smoking behavior and health is essential for informed policymaking. As Glied (2003) points out, the main justification for focusing on youths with tobacco control policy is not to reduce their current smoking so much as it is to reduce their long-term through adulthood, for which the health consequences are clearer. To project the long-term, population level effects of youth tobacco control policies, simulation models are typically used to calculate lifetime effects on smoking and health (Levy et al. 2000, Ahmad and Billimek 2007), but these models can only be as accurate as the empirical evidence on which their assumptions are based.

4.2.3 Smoking and military service

Two recent studies estimate the effects of military service on smoking and measures of health. First, Bedard and Deschênes (2006) estimate the long-run effects of military service during World War II or the Korean War on smoking, mortality and other measures of health, using instrumental variables approaches that exploit variation in the proportion of men who served across birth cohorts. They find that military service is associated with approximately a 30 percentage point increase in the probability of lifetime smoking, and this appears to account for significantly increased post-service mortality. Their results on smoking are only for the probability of having ever smoked, not current smoking, and thus do not address the evolution of smoking over time that is the focus of our paper.[4]

Second, Dobkin and Shabani (2007) use the draft lottery to instrument for Vietnam era military service in order to estimate effects of service on a large number of health-related measures in the NHIS. Their outcomes include smoking as of 1997-2004, and they do not find a significant effect of draft eligibility, consistent with our finding for 1997-2005. They do not examine smoking in earlier years, however. Also, they find no significant effects of military service on a large number of health measures during various periods (1974-1981, 1982-1996, and 1997-2004). As we discuss in Section 4.6, although Dobkin and Shabani caution that their estimates are too imprecise to rule out small but meaningful health effects, at the least their null findings

[4] Two other recent papers exploit variation across adjacent cohorts and male-female differentials in military service: de Walque (2007) and Grimard and Parent (2007). These papers estimate the effect of education on smoking by focusing on birth cohorts prior to 1950-52, who largely entered service prior to the draft lottery and its associated restrictions on educational deferments. As Card and Lemieux (2001) show, many men in these earlier cohorts were able to avoid the draft by attending or remaining in school.

bolster our conclusion that the large increase in smoking during young adulthood due to military service did not seriously affect health later in life.

The novel aspects of our study are to: a) show that the effect of military service on smoking in early adulthood was large; b) show that the dissipation of this effect over time cannot be explained by factors such as mortality, education, or immigration. Therefore our use of the draft lottery natural experiment, in contrast to Dobkin and Shabani's, directly speaks to the persistence of cigarette smoking over time.

4.3 Data

Our primary data source is the National Health Interview Survey (NHIS), conducted by the Centers for Disease Control (CDC). The NHIS is an annual survey that asks various health-related questions to a nationally representative random sample. Although the NHIS goes back to 1969, questions related to smoking were not asked in every survey. The NHIS contains data on smoking related questions in the 1978, 1979, and 1980 surveys, and also in every survey year from 1997 through 2005.[5] By combining data from different NHIS survey years, we create two pooled cross-sectional data sets, one for 1978-1980 and one for 1997-2005.

Each year it was held, the Vietnam draft lottery applied to men from specific birth cohorts. In particular, the 1969, 1970, and 1971 lotteries applied in turn to the 1950, 1951, and 1952 birth cohorts, who were aged 19 to 20 when their respective lottery was conducted. We assigned respondents in the NHIS their lottery number, based on their year and day of birth, using the mapping published by the U.S. Selective Service (see http://www.ssa.gov). We obtained access to date of birth information via the CDC's restricted, in-house NHIS data files at a secure facility.

Table (4.1) shows the draft eligibility cutoff numbers and which cohorts they applied to. We created our instrument for military service as a dummy variable, Draft, equal to 1 for men whose lottery numbers were below the eligibility cutoff in the year they were subjected to the lottery, and 0 otherwise.

We leave out the 1944-1949 birth cohorts, who were subject to the first lottery in 1969, because most servicemen from these cohorts entered the military before the lotteries started (Angrist 1990). Therefore men from these cohorts who avoided service until 1970 are far from a representative sample. In addition, for these cohorts the draft lottery is a much weaker instrument than it is for the 1950-1952 cohorts.

[5]The NHIS also has supplements in 1987 (Cancer Epidemiology) and 1988 (Occupational Health) with smoking questions. We discuss results from these supplements in Section 4.5.

There was also a final lottery in 1972, applying to men from the 1953 birth cohort, but no men from this cohort ended up being drafted. Empirically, men with low lottery numbers in this cohort were slightly more likely to become veterans (1.5 percentage points in our sample), as some of them presumably anticipated being drafted and pre-emptively enlisted to receive more favorable terms of service. Due to the weakness of the lottery as an instrument for these men, we exclude the 1953 cohort.

Our sample is therefore men from the 1950 to 1952 birth cohorts. In our main analyses, we use 3 instruments: The Draft variable interacted with a dummy for each of the 1950, 1951, and 1952 birth cohorts.[6] This improves the predictive power of draft eligibility on veteran status, because the proportions of draft eligible and ineligible men who ended up serving in the military varies across the 1969-1971 lotteries.

In the 1978-1980 NHIS, respondents were asked whether they were veterans, and if so, in which conflict they served. This allows us to identify veterans of the Vietnam era military. The question related to veteran status is somewhat different in the 1997-2005 NHIS: "Have you ever been honorably discharged from active duty in the U.S. Army, Navy, Air Force, Marine Corps, or Coast Guard?" This formulation understates the true number of veterans by omitting those who were not honorably discharged. Information about this proportion does not appear to be available by year or cohort, but we do know that approximately 7% of all Vietnam era servicemen were given discharges in categories other than honorable (Baskir and Strauss 1978).[7] Drawing from multiple Department of Defense documents and reports, Baskir and Strauss find that instances of violations of military rules such as Absent Without Leave and prolonged absence ("administrative desertion") peaked during the last few years of the war. In addition, the military issued a peak number of "Chapter 10" discharges in 1972, 90% of which were classified as "Undesirable" discharges after administrative procedures. These facts suggest that the proportion of veterans from the 1950 to 1952 cohorts who were not honorably discharged is higher than 7%.

The measurement error in veteran status could bias our IV estimates for 1997-2005 in either direction, depending on the relationships between the likelihood of being other-than-honorably discharged, draft eligibility, and smoking propensity. Importantly, however, the measurement error in veteran status is not relevant to the re-

[6]The simple estimates discussed in Section 4.1 only use draft eligibility, and are quite similar to our main results using eligibility interacted with year of birth.

[7]Out of 7,575,000 active duty Vietnam era servicemen, 563,000 were given less-than-honorable discharges of various categories, including General, Undesirable, Bad Conduct, and Dishonorable. Approximately 34,000 of the less-than-honorable discharges were in the most serious categories of Bad Conduct or Dishonorable.

duced form relationship between draft eligibility (based on date of birth) and smoking. As we show later, this reduced form relationship is very small and not statistically different from zero in 1997-2005, and therefore our IV estimates would not be significant even if veteran status was measured perfectly.

In all years in which the NHIS asks about smoking, the questions about smoking begin with, "Have you smoked at least 100 cigarettes in your entire life?" The questions following this stem question differ slightly between the 1978-1980 and 1997-2005 survey periods. In 1978-1980, those who answer yes are then asked, "Do you smoke cigarettes now?" We classify those who answer yes as current smokers. In the more recent survey period, those who answer yes to the stem question are asked "Do you now smoke cigarettes every day, some days, or not at all?" We classify those who answer "every day" or "some days" as current smokers. We classify people as non-smokers if they answer no to either the stem questions about 100 lifetime cigarettes or the follow-up questions about current smoking.

4.4 Empirical methods and results

4.4.1 Reduced-form comparison of draft eligible to non-eligible men

The estimated effect of military service on smoking depends upon the reduced form relationship between draft eligibility and smoking behavior, and the first stage relationship between eligibility and veteran status. Table (4.2) shows that for men surveyed in 1978-1980, draft eligibility is positively related to both the probability of serving in the military and the probability of being a current smoker. Also, the differential diminishes by cohort, from 1950, to 1951, to 1952, suggesting that the differential might diminish as a function of time since service. Indeed, as of 1997-2005, draft eligibility is significantly related to veteran status but not to smoking. The difference in the proportion of current smokers between draft eligible and non-eligible men disappears completely for the 1950 and 1951 birth cohorts, and only a slight difference remains for the 1952 cohort.

The reduced-form patterns by draft eligibility for smoking in Table (4.2) drive our main results. The positive relationship between eligibility and smoking in 1978-1980 produces positive and significant estimates for the effect of Vietnam-era military service. For 1997-2005, the lack of a significant relationship between eligibility and smoking implies that the effect of Vietnam-era service is no longer significant in that time period.[8]

[8]These results and all other results reported in tables and figures are based on analyses with-

The columns labeled RF in Table (4.3) highlight this difference in the reduced form results between the two time periods. In 1978-1980, draft-eligible men have a 0.064 higher probability of smoking, whereas in 1997-2005 this differential is almost exactly zero (0.001). The relationship between eligibility and veteran status also changes somewhat, from 0.20 to 0.12. This is probably related to the difference in survey questions discussed earlier, as well as the fact that the 1997-2005 sample includes a small proportion of immigrants who were not subject to the draft during the 1970's.[9] As noted earlier, however, because the reduced form relationship between eligibility and smoking is close to zero and insignificant in 1997-2005, our estimates of the effect of military service would not be statistically significant regardless of the veteran status variable.

To see this, consider an extreme case using the "true" difference in the probability of service by draft eligibility according to the Defense Manpower Data Center (DMDC, see Figure 2). According to the DMDC, the true effect of eligibility on the probability of service is about 0.123 (for the 1951 and 1952 cohorts). Our 1997-2005 reduced-form effect of eligibility on current smoking is 0.001 with a standard error of 0.012. Treating the DMDC effect as a known constant, a simple two-sample IV estimate of the effect of military service on smoking is $\frac{0.001}{0.123} = 0.0081$, with a standard error of $\sqrt{\frac{0.012^2}{0.123^2}} = 0.0976$.

4.4.2 Wald estimates

The simplest IV estimate of the causal effect of military service on smoking is the Wald (1940) estimator. The Wald estimator may be computed as:

$$\hat{\beta}_W = \frac{\hat{S}_e - \hat{S}_{ne}}{\hat{V}_e - \hat{V}_{ne}} \tag{4.1}$$

Here $\hat{S}$ denotes the proportion of smokers, $\hat{V}$ the proportion of veterans, the e subscript denotes draft eligible, and ne not draft-eligible. The Wald estimator can be calculated using the information given in Table (4.2). A corresponding naive

out sample weights. All results remain nearly identical when re-estimated using the NHIS sample weights.

[9]The NHIS does not ask about country of origin, but we found in a different nationally representative data set from 2002-2003 - the National Epidemiological Survey of Alcohol and Related Conditions - that 13% of men born in 1950-1952 were foreign born. If we make the conservative assumption that none of these men were veterans, then the inclusion of these men in the NHIS estimates lowers our estimate of p(veteran) by about 3 percentage points. Also, the inclusion of these men in the NHIS estimates lowers our estimated differential in p(veteran) by draft eligibility by about 1 percentage point.

estimate for the effect of military service on smoking is simply to compare the sample proportion of veterans who smoke with the proportion of non-veterans who smoke. Letting v denote veterans and nv non veterans, the naive estimator is:

$$\hat{\alpha} = \hat{S}_v - \hat{S}_{nv} \tag{4.2}$$

The Wald estimator given in (4.1) is equivalent to estimating two-stage least squares (2SLS) with only a constant and veteran status as regressors, and with only draft eligibility as the instrument. We calculate the Wald estimator in this way in order to generate the usual 2SLS standard errors. The model to be estimated is the following:

$$V_{it} = \delta_0 + \delta_1 D_{it} + e_{it} \tag{4.3}$$

$$S_{it} = \beta_0 + \beta \hat{V}_{it} + u_{it} \tag{4.4}$$

Here V_{it} is the veteran status of person i in survey year t, D_{it} is a dummy for draft eligibility, and S_{it} is a smoking dummy variable, and $\hat{V}_{it}$ is predicted veterans status from equation (4.3). The IV estimate of the effect of military service is the coefficient β in equation (4.4). This IV estimate is a consistent estimator for the causal effect of military service on smoking if draft eligibility is related to military service and is uncorrelated with any other factors in the error term u_{it} that affect smoking behavior. To use draft eligibility as an instrument for military service, we must assume that the reduced form relationship between eligibility and smoking is entirely due to the relationship between eligibility and military service. This assumption is reasonable for draft eligibility because it was randomly assigned based on date of birth. In regression form, the naive estimate for the effect of military service is the following:

$$S_{it} = \alpha_0 + \alpha V_{it} + \varepsilon_{it} \tag{4.5}$$

Table (4.3) shows the Wald estimates (using 2SLS) and the naive estimates (using OLS) of the effect of military service for both sample periods. The naive estimates are quite similar across the two sample periods. Veterans were about 11 percentage points more likely to be smokers than non-veterans. The Wald estimates are quite different. The Wald coefficient on veteran status for 1978-1980 indicates that serving in the military increased the probability of being a smoker by 32 percentage points. The corresponding estimate of 1 percentage point for the 1997-2005 surveys is not statistically different from zero.

4.4.3 2SLS estimates with controls

While the Wald estimates are simple and show a marked pattern of results, there are potential problems which we address further now. First, it has been documented that eligibility assigned by the draft lottery is slightly correlated with month of birth for the first lottery held in December 1969, due to imperfect mixing of the balls drawn from a container.[10] Because of this association, we include month of birth dummies as controls. We calculate first stage F statistics for the significance of the excluded instruments to check that our instruments are not weak, and we conduct the Hansen J-test of the validity of over-identifying restrictions to provide evidence that our instruments are exogenous. The null hypothesis of the J-test is that the IV estimate using all available instruments differs only by sampling error from the IV estimate using a subset of instruments (one, in our case) that just identify the equation. Finally, we include race, age dummies, and cohort dummies as additional control variables.

For our main results we use a simple 2SLS specification, where the first stage is a linear probability model for veteran status, and the second stage is a linear probability model for current smoking.[11] Including controls, the model to be estimated becomes:

$$V_{it} = \delta_0 + \delta_1(D_{it} * B_{it}) + \delta_2 R_{it} + \delta_3 X_{it} + e_{it} \tag{4.6}$$

$$S_{it} = \beta_0 + \beta_1 \hat{V}_{it} + \beta_2 R_{it} + \beta_3 X_{it} + u_{it} \tag{4.7}$$

Equation (4.6) is the first stage and equation (4.7) is the second stage. $S_{it} = 1$ if individual i in survey year t is a current smoker, and $V_{it} = 1$ if the individual is a veteran. The variable R_{it} is a race dummy equal to 1 if the individual is black. X_{it} is a vector of birth cohort dummies, age dummies, and month of birth dummies. The instrumental variables are $D_{it} * B_{it}$, which are birth cohort dummies interacted with draft eligibility. The parameter of interest is β_1, which measures the causal effect of serving in Vietnam on the probability of being an active smoker. Current smoking is a good candidate for the linear probability model because the sample proportion of

[10] See Feinberg, S.E. (1971), "Randomization and Social Affairs: The 1970 Draft Lottery", Science, 171, 255-261.

[11] As compared to 2SLS, Limited Information Maximum Likelihood (LIML) has been shown to have some desirable finite sample properties (Anderson 1982), and two-step linear Generalized Method of Moments (GMM) is more efficient in the case of multiple instruments and a single endogenous regressor. We find, however, that point estimates and standard errors from LIML and two-step GMM (available upon request) are very similar to our 2SLS estimates. Thus for ease of interpretation we focus on estimates from 2SLS, using heteroskedastic-robust estimates of standard errors.

smokers is not too close to 0 or 1.

A natural question to ask is how to include education in the model, if at all. An argument for omitting education is that it is endogenous in the smoking model, as it may be correlated with factors in the error term that affect smoking, such as the subjective rate of time preference. A possible argument for including education is that educational attainment could be a mechanism, beyond military service per se, by which draft eligibility influences smoking. Card and Lemieux (2001) find that for men born in 1950 or later, the Vietnam draft lottery had little or no effect on educational attainment. They do find positive effects of the Vietnam draft on educational attainment for men born in the mid to late 1940's, i.e. men not in our sample. Using more recent data (the 2000 Census) and draft eligibility as the instrument for Vietnam-era military service, Angrist and Chen (2007) find small but statistically significant positive effects of service on educational attainment for white men born in 1948-1952.[12] Based on the phase-out of educational draft deferments during the draft lottery period and the time pattern of educational attainment (in the CPS) for Vietnam veterans born 1948-1952, Angrist and Chen argue that the effects on schooling are due to increased educational attainment (subsidized by the GI Bill) in the years after military service, and not to schooling obtained to avoid the draft.

Our results are very similar when we include education as a control variable in specification checks. Also, in the NHIS data, we do not find a statistically significant relationship between draft eligibility and educational attainment, although our sample size is too small to rule out the small effect estimated by Angrist and Chen (2007). Later, in Section 5, we discuss an upper bound estimate for how much of the decline over time in our estimated effects of service on smoking may be due to increased educational attainment among veterans.

Our main results, estimated by 2SLS using draft eligibility interacted with year of birth as instruments for military service, are presented in Table (4.4). For the 1978-1980 time period, the lottery-based 2SLS estimate is larger (35.3 percentage points for current smoking) than the OLS estimate of 11.3, indicating a downward bias in the OLS estimate. Both coefficients are significant according to the reported heteroskedastic-robust standard errors. The first stage F statistics are well above the usual rule of thumb of 10, and the J-test does not reject the overidentifying restrictions

[12] Angrist and Chen's draft-eligibility based estimates for whites indicate that military service increased years for schooling by about one-third of a year, and increased the chance of attaining at least a college degree by 0.05. The effects for non-whites are similar in magnitude but not statistically significant.

imposed by our 3 instruments.

While the 2SLS point estimates for 1978-1980 are substantially larger than the OLS point estimates, the standard errors for the 2SLS estimates are also fairly large. We use a regression based Hausman test for endogeneity to test whether the estimated 2SLS coefficients are significantly different from the OLS coefficients, and the test rejects the null hypothesis of no endogeneity at the 10% level (p=0.089).

The results for the 1978-1980 surveys in Table (4.4) are nearly identical to the Wald estimates presented in Table (4.3). This is consistent with the fact that because draft eligibility was randomly assigned, additional control variables are not essential to identify the parameter of interest. Table (4.4) also shows that we obtain very similar estimates when including indicators for college completion and marital status as additional control variables.

The fact that the lottery-based IV estimates for the effect of service on smoking are larger than OLS estimates in 1978-80 suggests that, on balance, unobservable factors made veterans *less* prone to smoking. This is because in order for the OLS estimate to be biased downwards, the correlation between the omitted variable (assuming there is only one) and smoking must have a different sign than the correlation between the omitted variable and military service. This may be related to the fact that health standards for enlistment into the military were relatively strict; as Angrist (1990) points out, over half of potential inductees in 1970 failed a physical or mental examination.[13] In economic terms, we can interpret our IV result for 1978-1980 as suggesting that, prior to military service, those who went on to become veterans had a lower reservation price for smoking initiation than those who did not.

Another potential explanation for the apparent downward bias of OLS may be that men induced to serve by the draft faced a different set of conditions than other men who served. Assuming that all of the reduced form effect of draft eligibility on smoking is attributable to increased military service, our lottery-based IV estimates represent the average effect of service on smoking behavior for those men induced to serve by the draft (i.e. a local average treatment effect, as shown by Angrist et al. 1996). It is possible that the smoking behavior of these "marginal" participants in the war was more affected by service than people who served in general. This could happen, for example, if men with low draft numbers who served were more likely to be assigned to places where access to cigarettes was higher than elsewhere in the military. We have not found any evidence, however, that conditional on serving, men with low

[13]Baskir and Strauss (1978) report that from 1967 to 1973, the failure rate for the pre-induction physical exam was 47%.

draft numbers were more likely to be sent to particular divisions of the military, nor any evidence that access to cigarettes varied significantly across divisions.

For the 1997-2005 survey period, Table (4.4) shows that the OLS estimate for the effect of military service on smoking (10.7 percentage points) is similar to the OLS estimates for 1978-1980. However, the 2SLS estimate is only 0.028 and is not statistically different from zero, which is very different from the result for 1978-1980. The J-test comfortably fails to reject the null that 2SLS estimates using a subset of instruments differ only by sampling error from estimates using all instruments for the current smoker specification. Using the regression based Hausman test, we cannot reject the null hypothesis of no endogeneity (p=0.448). In addition, we find that our 2SLS estimates remain similar if we control for college completion and marital status. Thus, our results suggest that military service caused a large increase in smoking as of young adulthood, but this effect did not persist into later adulthood.

4.4.4 Outcomes by draft lottery number intervals

The lottery-based 2SLS specification attributes all of the reduced form effect of draft eligibility on smoking to increased military service. Here we examine the validity of this assumption by looking at how smoking varies across more narrowly defined intervals of draft numbers. In particular, one concern is that having a very low draft number could have affected smoking via mechanisms other than military service, such as direct psychological distress or secondary effects resulting from a draft avoidance strategy. If this were the case, we might see distinct patterns in smoking by draft number within the draft-eligible group, such as highly elevated rates for men with very low numbers.

Figure (4.1) provides graphical evidence related to this possibility. Due to our modest sample sizes, we show means by groups of 50 consecutive lottery numbers (means by smaller groups are even noisier than those shown). Recall that the eligibility cut-offs were 195 for the 1950 cohort, 125 for the 1951 cohort, and 95 for the 1952 cohort. For the 1978-1980 sample, a sharp drop in smoking is observed to the right of the eligibility cutoff for all cohorts. These patterns are muted in 1997-2005 (consistent with our main results), and there is no clear pattern suggesting that very low lottery numbers had a distinctive effect. In addition, there is no evidence for a monotonic relationship between draft lottery number and smoking, which might have suggested some direct effect of lottery numbers. We also show in Figure (4.2), using a much larger data set from the Defense Manpower Data Center,[14] that the re-

[14]Joshua Angrist used these data in his 1990 paper, and we thank him for sharing the data with

lationship between draft number and veteran status approximates a relatively clean discontinuity at the cutoff numbers.

4.4.5 Other smoking variables

In order to gain a fuller understanding of our main pattern of results, we examine several additional smoking variables available in NHIS. First, we look at current smoking in supplements to the NHIS in 1987 and 1988. Second, we look at variables other than current smoking: in particular, lifetime smoking, age of initiation, and number of cigarettes per day as of 1978-80, 1987-88, and 1997-2005. For simplicity, in this section we focus on reduced form comparisons between men with and without draft-eligible numbers.

Current smoking in 1987-88 was ascertained by NHIS in the exact same way as described earlier for the 1997-2005 surveys. Comparing by draft-eligibility, we find that smoking prevalence is 0.37 for both groups (Table 4.5). This suggests that the large effect observed in 1978-80 had already disappeared as of 1987-88.[15]

Next we examine a dichotomous measure of lifetime smoking, which is based in all survey years on the answer to the stem question, "Have you smoked at least 100 cigarettes in your entire life?" As of 1978-80 we find that draft-eligible men have significantly higher lifetime smoking: 0.65 vs 0.59 (Table 4.5). By contrast, lifetime smoking is not significantly different by draft-eligibility as of 1987-88 (0.59 vs 0.61) or as of 1997-2005 (0.59 vs 0.59). It is interesting to note that the convergence in lifetime smoking comes from a decrease among draft-eligible men, rather than an increase among non-eligible men.

There are three possible, non-mutually exclusive explanations for this pattern over time: a change in the composition of the sample, misreporting of smoking history, or a true convergence in lifetime smoking probabilities. As we describe later in Section 5, a variety of evidence suggests that attrition is not a significant issue in this context, and we described earlier why immigration is not a significant issue either. By contrast, recall bias, or systematic errors made by individuals when they are asked to recall past events, may be important. Significant mispreporting has been documented for former smokers.[16] By contrast, self-reports of current smoking have been shown to be

us. The DMDC data are available for men born in 1951 and 1952, but not 1950.

[15]We do not include these data in our main analysis because the null 2SLS results we find for 1987-1988 are not definitive on when the smoking stopped, because standard confidence intervals for the estimates include sizable effect sizes.

[16]For instance, Kenkel, Lillard, and Mathios (2004) show that among the NLSY respondents who contemporaneously reported being smokers in 1984, 26 percent reported in 1998 that they were not

much more accurate. In the 1988-94 and 2001-2002 National Health and Nutrition Examination Surveys, only about 1 percent of people who said they were non-smokers had cotinine levels in their blood consistent with being an active smoker (Caraballo et al. 2001, West et al. 2007).

In our context, it seems plausible that misreporting by former smokers might operate differentially across draft eligibility status. Some veterans who were still smoking in 1978-1980 but who quit by the later sample period may have thought that smoking during their military days did not really "count". This seems particularly plausible given that misreporting has been shown to be highest among former light smokers (Kenkel et al. 2004, Stanton et al. 2007), and draft-eligible men were more likely to smoke 10 or fewer cigarettes per day than non-eligible men (Table 4.5). Another piece of evidence suggesting that former smokers are misreporting is that the overall reported lifetime prevalence *decreases* over time, despite our evidence in Section 5 that changes in sample composition is not a significant issue.[17] In addition, for *women* born 1950-1952, our calculations from public use NHIS data files show that the proportion reporting having ever smoked falls from 0.485 in 1978-1980 to 0.424 in 1997-2005. Given that mortality rates for women are even lower than those for men in these cohorts, this marked decline over time in the proportion of women reporting having ever smoked casts further doubt on the accuracy of reports by former smokers. The remaining explanation is that non-eligible men "caught up" to eligible men between 1980 and 1997. This may seem unlikely, but it is important to keep in mind that the differences as of 1978-1980 were due to "marginal" smokers; i.e. men who would not have smoked by that point if not for having had a low draft number. From that vantage point, it is perhaps not surprising that many "marginal" non-smokers in the non-eligible group went on to try smoking after 1980.

We next examine the age of smoking initiation. In each survey this is asked as,

smokers in 1984. Glied (2002) points out that among the NLSY respondents who reported that they had never smoked daily in 1992, eight percent reported smoking at least one cigarette daily in 1984.

[17] A factor relevant to the decline in reported lifetime smoking is the presence of foreign-born men in the 1997-2005 NHIS sample. As explained earlier, we estimate based on the NESARC that in the 1997-2005 period 13% of men from the 1950-1952 cohorts are foreign-born. We also find in the NESARC that foreign-born men are less likely to report having ever smoked: 39% as compared to 56% of U.S. born men in these cohorts. Given that foreign-born men constitute about 1/8 of the sample, their inclusion in our NHIS estimates likely deflates our estimated p(ever smoked) by about 2 percentage points. (Along with misreporting, this helps explain why p(ever smoked) fell over time). The presence of foreign-born men in the data, however, cannot explain why the smoking *differential by draft eligibility* diminishes to zero among the 1950-1952 cohorts. Given that the estimated differential for the overall sample (U.S. plus foreign-born men) is zero, and the differential for foreign-born men is logically zero (as the vast majority of them were not subject to the draft), then the differential for U.S. born men must also be zero.

"About how old were you when you first started smoking fairly regularly?" In each time period we find that the average age of initiation is not significantly different by draft-eligibility. The distribution of this variable, however, is probably more informative. In Figure (4.3) we show the probability that men initiated smoking by each age for the 1978-1980 sample. The first figure shows that the divergence in smoking behavior began around the age when men were subject to the draft (19-20).[18] In contrast, analogous figures for the later time periods (not shown) do not reveal any apparent differences in the distribution of initiation age by draft-eligibility; the graphs are nearly identical in each case. This supports the idea that misreporting by former smokers, rather than true convergence, explains the apparent convergence of lifetime smoking probability described earlier. If true convergence was the main explanation, then we would observe it in the distributions of initiation age.

Thus far we have focused on the extensive margin: the decision whether to smoke at all. We now look at the intensive margin, by examining responses to the question, "On the average, about how many cigarettes per day do you smoke?" As in the case of age of initiation, in each time period we do not observe any significant differences by draft-eligibility in the *mean* number of cigarettes (Table 4.5). As already noted, we do observe, however, a difference in the probability of being a "light smoker" (defined as 10 or fewer cigarettes per day) as of 1978-80. This foreshadows the apparent fact that draft-eligible men were more likely to quit after 1980, and, as noted above, is also consistent with the idea that in later periods, among former smokers, draft-eligible men were more likely to misreport (i.e., deny) having ever smoked.

4.4.6 Health effects

The evidence thus far suggests that draft-eligible men were much more likely to smoke through their mid to late twenties, but not much longer beyond that. We now examine whether this increase in smoking during young adulthood led to long-term health consequences. Although these men were still too young to experience significant mortality from smoking, as discussed in the next section, it is possible that they could have less severe health problems by the later sample period. In Table (4.6) we show 2SLS estimates, using our preferred specification (as in Table 4.4), for two health outcomes: self reported health and lifetime incidence of cancer. Respondents were asked to rate their health on a 5 point scale, where 1 is excellent and 5 is poor. The question about cancer asks the respondent if he has ever been

[18] The fact that the divergence started a year earlier than we would expect, at age 18, is presumably due to misremembering of the exact age.

diagnosed with cancer of any kind.

Although the 2SLS estimates are relatively imprecise, they indicate that, if anything, men with draft-eligible numbers were slightly better off for these measures of health. The NHIS, of course, also offers the opportunity to examine a large number of other health measures. As described earlier, Dobkin and Shabani (2007) use the draft-lottery natural experiment and find very few significant health effects in any of several time periods from 1974-2004. They caution that the estimates are not precise enough to rule out meaningful differences for many measures, but at the least their results, along with our own estimates for self-reported health and cancer, support the idea that the increased smoking in young adulthood did not have large, negative health effects.

Another caveat pertinent to this result is that military service may have had effects on health via mechanisms other than smoking. The direction of this bias, if any, is unclear a priori. As noted earlier, the cohorts in our analysis were not likely to be involved in combat, but the stress of potential combat may have affected health. Also, participation in the military may have affected physical fitness later in life (although Dobkin and Shabani do not find any significant differences in hypertension, for example).

4.5 Why does the smoking effect dissipate over time?

In this section we discuss two factors, post-service educational attainment and mortality, that could plausibly contribute to the disappearance over time of the smoking differential between men with and without draft-eligible lottery numbers. We find that these factors are unlikely to be significant in this context. We then discuss the implications of the fact that smoking behavior across the two groups appears to have converged after such a large differential in early adulthood.

Angrist and Chen (2007) use the draft-lottery IV approach to estimate that military service during the Vietnam-era led men to attain about one-third of a year of additional schooling in the years after they completed service. They find that this increase in education was likely related to educational subsidies offered to veterans. Given that there is a well-established inverse connection between education and smoking, this could account for a portion of the decline in smoking that we observe. Results from both De Walque (2007) and Grimard and Parent (2007) suggest that an upper bound on the size of the effect of educational attainment on smoking is about a 10 percentage point decline in the probability of being a current smoker for every year of education attained above high school. Combining these sets of results, it seems that increased education due to military service may account for at most a $\frac{0.10}{3} = 0.033$ decline in the differential probability of smoking. This is only a small portion (about one tenth) of the decline in our IV estimates between the early and later periods.

Attrition due to mortality is still a potential problem for our results for the 1997-2005 period. Given that we find no significant difference in smoking across draft eligibility status during that period, it is simplest to think about possible attrition bias for this reduced form relationship. The reduced form relationship between draft eligibility and smoking could be attenuated in the later survey period by either of two types of changes in sample composition. One possibility is that in the draft-eligible group, smokers were more likely to die than non-smokers. Another possibility is that in the draft-ineligible group, smokers were less likely to die than non-smokers. If smokers were more likely than non-smokers to die in both groups, which seems to be the most reasonable assumption, then there could still be attrition bias if mortality is higher in the draft eligible group.

In our data, however, we do not see evidence for differential mortality by draft eligibility status. If there were significant differences in mortality, we would see a reduction in the proportion of draft-eligible men between the earlier and later samples. As shown in Table (4.7), the proportion of draft-eligible men remained identical between the 1978-1980 and 1997-2005 NHIS samples: 38.0% in both cases. Furthermore, using a much larger sample (a subsample of the 2000 Census), Angrist and Chen (2007) also find no evidence of differential mortality; for men born in 1950-1952, the proportion who were draft eligible is almost identical to the proportion predicted based on a uniform distribution of dates-of-birth. The lack of evidence for differential mortality is perhaps not surprising, given that as of 1997-2005 these men were still in an age range (45-55) where smoking-related mortality is low (Sonnenschein and Brody, 2005).

We do not want to dismiss completely the possibility of differential mortality, however, given that Hearst et al. (1986) find for the 1950-1952 cohorts that the post-war mortality rate was slightly (1.04 times) higher among draft-eligible men than among non-eligible men, using records from California and Pennsylvania from 1974 to 1983. Using their results and other information, we now calculate an upper bound estimate of how attrition might attenuate the estimated differential in smoking by eligibility status in 1997-2005. We can approximate from national epidemiological data that 6.5% of men from the 1950-1952 birth cohorts died between the midpoints of the two sample periods (1979 and 2001, respectively).[19] National mortality rates by smoking status and age are not available, to our knowledge, so we make the conservative assumption that mortality was fully twice as likely among smokers during this period. If on average 40% of men were smokers during the time span of interest, then the 6.5% overall mortality rate would translate to about 8% for smokers and 4% for non-smokers. If we then assume that draft-eligible men were 1.04 times more likely to die, as estimated in Hearst et al. (1986), and make the additional conservative assumption that this "excess" mortality was entirely accounted for by *smokers* in the draft-eligible group (via a 1.08 relative risk), the mortality rates for the draft-eligible group would then be 8.64% for smokers and 4% among non-smokers, as compared to 8% and 4% for the ineligible group. Applying these mortality rates to the 1978-1980 smoking rates would cause the smoking rate for draft-eligible men to fall from 50.0% to 48.8% and that for draft-ineligible men to fall from 43.5% to 42.5%, meaning that the difference would fall from 6.5% only to 6.3%.[20] Therefore, even under conservative assumptions, it appears that mortality could explain very little of the dissipation in the smoking effect that we observe.

Another possible explanation for the dissipation of the smoking effect is that it is simply due to sampling error or some other anomalous feature of the NHIS data. As another check, we examine a different data source, the 2002-2003 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), which was conducted and sponsored by the National Institute on Alcohol Abuse and Alcoholism. This publicly available data set includes information on smoking and exact date of birth. It does not include information on veteran status so we cannot repeat our full IV analysis. Instead, we simply compare means for current smoking across draft-eligibility status. We find that draft-eligible men are only slightly more likely to be smokers (0.39 versus 0.37), and the difference is not significant (p-value=0.5), again supporting the idea that the differential dissipated over time.[21]

Given that this dissipation is evident in two different national data sets and cannot be explained by education or attrition, our results suggest that for the men in our analysis smoking behavior is highly malleable over time. More specifically, a large randomly induced increase in the likelihood of smoking in young adulthood

[19]We used annual mortality rates by age group for men, as reported by the National Vital Statistics in 2001. Reference: Table 1 in Centers for Disease Control (2001).

[20]The smoking rate among draft-eligible men falls from $\hat{S}^{e}_{1979}$=0.50 to $\hat{S}^{e}_{2001} = \frac{50-50*0.0864}{(50-50*0.0864)+(50-50*0.04)}$. The smoking rate among draft-ineligible men falls from $\hat{S}^{ne}_{1979}$=0.435 to $\hat{S}^{ne}_{2001} = \frac{43.5-43.5*0.08}{(43.5-43.5*0.08)+(56.5-56.5*0.04)}$.

[21]The sample sizes are 391 draft eligible men and 614 non-eligible.

(these men were ages 25-30 in 1978-80) diminished to being small and statistically insignificant in middle age (they were ages 45-55 in 1997-2005). For these men, the addictiveness of cigarettes does not appear to have been sufficient to compel them to smoke throughout adulthood. This finding runs counter to the conventional wisdom, based on epidemiological evidence cited earlier, that smoking early in life "dooms" people to smoke later in life.

When considering whether this basic result would generalize to other contexts, such as non-veterans, females, or more recent cohorts, it is important to consider a few key factors. First, serving in the military during the Vietnam era is obviously a different experience from civilian life on a number of dimensions. It is unclear whether these factors would, on balance, make people more or less likely to persist in their smoking through adulthood, given that they are smoking in young adulthood. On the one hand, to the extent that certain factors associated with military life – such as stress or access to subsidized prices (at military stores) or proximity to peer smokers – persist beyond the period of military service, continued smoking would only be *more* likely among veterans. The fact that we still do not observe this continued smoking would suggest that this result would be especially likely to hold in non-veteran populations. On the other hand, there was a shift in culture and policy related to smoking in the military during the 1980s and 1990s. For example, the Veterans Health Administration (VHA) increased efforts to identify and counsel smokers, to the point where it was referred to as a "model health care system for smoking cessation" by a leading tobacco control expert (Schroeder 2004). Assuming that health care providers in general have increased smoking cessation efforts in the past 10-15 years, the VHA of the 1990s may be roughly representative of health systems overall today.

The dissipation of the smoking differential over time may also be related to the fact that men induced to smoke by having a low draft number can be thought of as "marginal smokers"; they would not have been smoking as of 1978-1980 if not for having been draft eligible (and serving in the military as a result). People on the margin of smoking or not smoking are arguably the most relevant for policymakers considering whether to make incremental changes in one direction or another. It is important to note, however, that our main results do not necessarily apply to "always-smokers", people who would have been smoking as young adults regardless of their draft number.

4.6 Conclusion

Using the Vietnam era draft lottery as a natural experiment, we find that military service caused men born in 1950-1952 to be 35 percentage points more likely to smoke as of their mid to late 20s, but this effect was markedly attenuated later in adulthood. We also find that there were no measurable health consequences as a result of the increased smoking during young adulthood, which is consistent with Dobkin and Shabani's (2007) examination of a much larger set of health variables.

As described above, generalizing these findings to other contexts should be done with caution, particularly due to some of the unique features of military service during the Vietnam era. Nevertheless, with these caveats in mind, our results speak to the strength of cigarette addiction over the long-term for a large segment of the population at the time. In effect, this is the first large-scale study to examine the long-term consequences of a randomly assigned exposure that significantly affected smoking behavior. The results are consistent with evidence on the long-term effects of cigarette taxes during adolescence (Gruber and Zinman, 2001; Glied, 2002), in that the effects are substantially diminished over time. For health policymakers and consumers in general, our results can be viewed as an affirmation that smoking during young adulthood does not compel one into lifelong smoking, and by quitting in young adulthood one can substantially mitigate longer-term health consequences.

4.7 Tables and Figures

Table 4.1: Draft eligibility cutoff number by birth cohort and year.

Lottery Year	Cohort(s) Affected	Eligibility Cutoff
1969	1944-1950	195
1970	1951	125
1971	1952	95
1972	1953	N.A.

Data Source: U.S. Selective Service.

Table 4.2: Sample proportions of Vietnam era veterans and smokers.

Cohort	Surveys	Draft Eligible	Veteran	Smoker	obs.
1950	78-80	No	0.18	0.46	229
		Yes	0.32	0.48	258
	97-05	No	0.23	0.31	1,042
		Yes	0.32	0.30	1,163
1951	78-80	No	0.16	0.42	342
		Yes	0.36	0.49	192
	97-05	No	0.18	0.30	1,432
		Yes	0.28	0.30	752
1952	78-80	No	0.10	0.42	396
		Yes	0.34	0.52	143
	97-05	No	0.16	0.30	1,617
		Yes	0.31	0.33	590
All Cohorts	78-80	No	0.14	0.43	967
		Yes	0.34	0.49	593
	97-05	No	0.18	0.31	4,091
		Yes	0.31	0.31	2,505

Data Source: 1978-1980 NHIS Smoking Supplements, 1997-2005 NHIS. Males only.
78-80 Veteran: Yes to "Veteran of U.S. Armed Forces?"
97-05 Veteran: Yes to "Ever honorably discharged from U.S. Armed Forces?

Table 4.3: Reduced form estimates of the effect of draft eligibility, and Wald estimates of the effect of military service.

	1978-1980			1997-2005		
Dependent Variable	RF	OLS	2SLS	RF	OLS	2SLS
Smoker	**0.064**	**0.11**	**0.32**	0.001	**0.11**	0.01
	(0.027)	(0.03)	(0.14)	(0.012)	(0.01)	(0.09)
Veteran	**0.20**			**0.12**		
	(0.023)			(0.011)		
Observations		1,452			6,553	

RF: Regression of dependent variable on a constant and draft eligibility.
2SLS uses draft eligibility as the instrument for military service.
1950-1952 birth cohorts, males only, no control variables.
Heteroskedastic-robust standard errors.

Table 4.4: Lottery-based estimates of the effect of Vietnam-era military service. Basic controls are cohort dummies, age dummies, birth-month dummies, and race. Instruments are birth cohort interacted with draft eligibility.

Dependent Variable: Smoker	Surveys	OLS	2SLS	F-stat	J-test
Basic controls	78-80	**0.113**	**0.353**	20.71	0.92
		(0.032)	(0.145)		
	97-05	**0.107**	0.028	33.69	0.48
		(0.014)	(0.105)		
Basic controls plus college and marital status	78-80	**0.064**	**0.322**	20.63	0.85
		(0.033)	(0.148)		
	97-05	**0.080**	0.055	35.03	0.39
		(0.014)	(0.100)		

F-stat: F-statistic for the test that instruments are jointly significant in the first stage.
J-test: p-value for the Hansen J-test of the over-identifying restrictions.
Observations: 1,452 for 78-80 results, and 6,553 for 97-05.
Heteroskedastic-robust standard errors.

Figure 4.1: Sample proportion of current smokers by lottery number group and birth cohort.

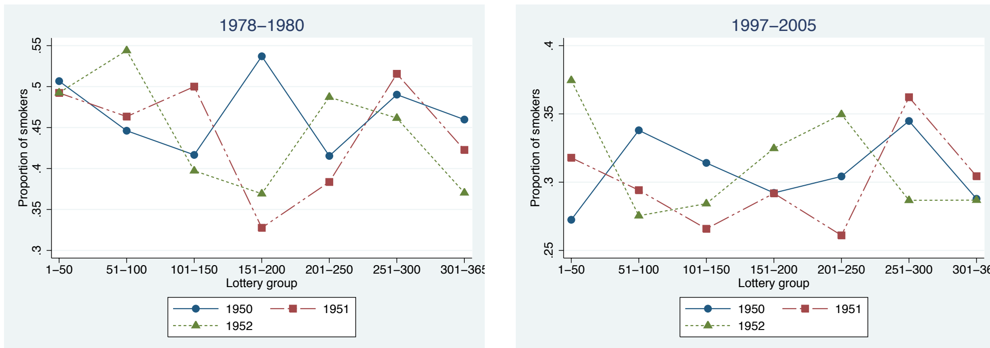


Figure 4.2: Proportion of men born in 1951 and 1952 who entered the military between July 1970 and December 1973, by draft lottery number. Whites only. Data source: Defense Manpower Data Center, provided to us by Joshua Angrist.

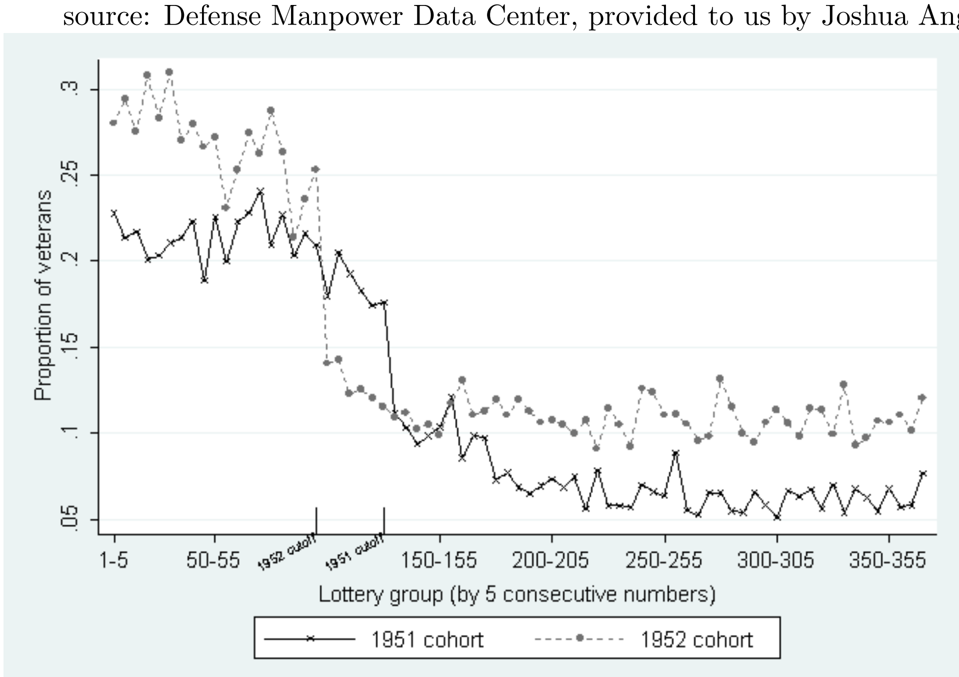


Table 4.5: Additional NHIS smoking variables: means by draft eligibility.

	1978-1980 Eligible Yes	No	p-val	1987-1988 Eligible Yes	No	p-val	1997-2005 Eligible Yes	No	p-val
Current smoker	0.49	0.43	(0.02)	0.37	0.37	(0.93)	0.31	0.31	(0.91)
Ever smoked	0.65	0.59	(0.02)	0.59	0.61	(0.42)	0.59	0.59	(0.81)
Age first smoked	17.0	16.9	(0.88)	17.5	17.6	(0.69)	19.5	19.6	(0.93)
Cigs/day	20.8	21.0	(0.85)	24.6	25.1	(0.66)	19.8	19.5	(0.57)
Light smoker	0.24	0.20	(0.26)	0.22	0.23	(0.62)	0.31	0.32	(0.78)

p-values: chi-squared tests for 0-1 variables, t-tests for age first smoked and cigs/day.

Figure 4.3: Probability of starting to smoke regularly by age.

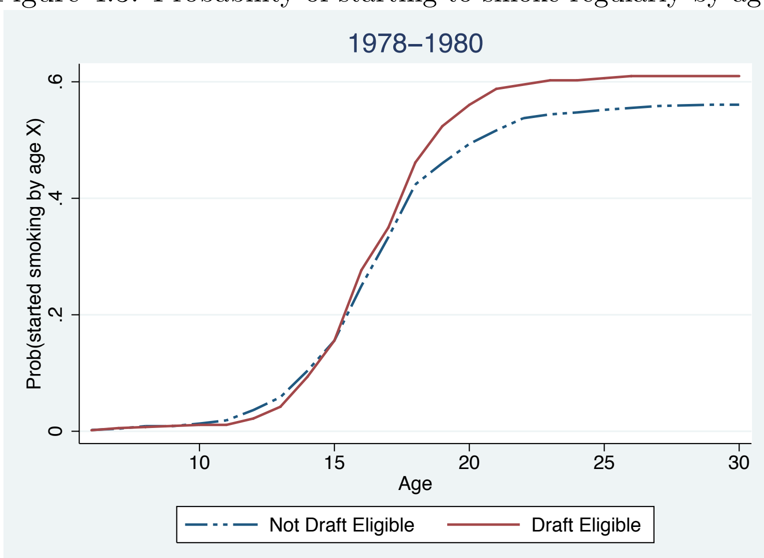


Table 4.6: Lottery-based estimates of the health effects of Vietnam-era military service. Controls are cohort dummies, age dummies, birth-month dummies, and race. Instruments are birth cohort interacted with draft eligibility.

Dependent Variable	Surveys	OLS	2SLS	F-stat	J-test
Reported Health	97-05	**0.084** (0.0321)	**-0.548** (0.252)	33.07	0.05
Cancer	97-05	0.009 (0.006)	-0.057 (0.042)	33.02	0.28

F-stat: test that instruments are jointly significant in the first stage.
J-test: p-value for the Hansen J-test of the over-identifying restrictions.
Observations: 6,592. Heteroskedastic-robust standard errors

Table 4.7: Sample proportions of veterans and draft eligible men.

	1978-1980 NHIS			1997-2005 NHIS		
Cohort	Draft Eligible	Veteran	obs.	Draft Eligible	Veteran	obs.
1950	0.53	0.25	487	0.53	0.28	2,208
1951	0.36	0.23	534	0.34	0.21	2,189
1952	0.27	0.17	539	0.27	0.20	2,212
All Cohorts	0.38	0.22	1,560	0.38	0.23	6,609

Data Source: 1978-1980 and 1997-2005 NHIS. Males only.

4.8 References

Ahmad, S, Billimek, J. "Limiting Youth Access to Tobacco: Comparing the long-term health impacts of increasing cigarette excise taxes and raising the legal smoking age to 21 in the United States." *Health Policy* 2007; 80(3):378-91.

Angrist, J. "Lifetime Earnings and the Vietnam Draft Lottery: Evidence from Social Security Administration Records." *American Economic Review* 1990; 80(3): 313-336.

Angrist, J, Imbens, G, Rubin, D. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 1996; 91(434): 444-455.

Angrist, J, Chen, S. "Long-term Consequences of Vietnam-Era Conscription: Schooling, Experience, and Earnings." NBER Working paper 2007; #13411.

Anderson, T, Kunitomo N, Sawa T "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator." *Econometrica* 1982; 50(4): 1009-1027.

Baskir, L, Strauss, W. "Chance and Circumstance: The Draft, the War, and the Vietnam Generation." New York: Knopf, 1978.

Becker, G, Murphy, K. "A Theory of Rational Addiction." *Journal of Political Economy* 1988; 96: 675-700.

Becker, G, Grossman, G, Murphy, K. "An Empirical Analysis of Cigarette Addiction." *American Economic Review* 1994; 84(3): 396-418.

Bedard, K, Deschênes, O. "The Long-Term Impact of Military Service on Health: Evidence from World War II and Korean War Veterans." *American Economic Review* 2006; 96(1): 176-194.

Caraballo, R, Giovino, G, Pechacek, T, Mowery P. Factors Associated with Discrepancies between Self-reports on Cigarette Smoking and Measured Serum Cotinine Levels among Persons Aged 17 Years or Older: Third National Health and Nutrition Examination Survey, 1988-1994. *American Journal ofEpidemiology* 2001; 153: 807-14.

Card, D, Lemieux, T. "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War." *American Economic Review* 2001; 91(2): 97-102.

Centers for Disease Control. *National Vital Statistics Report* 2001; 49(11).

Chaloupka, F. "Rational Addictive Behavior and Cigarette Smoking." *Journal of*

Political Economy 1991; 99: 722-742.

De Walque, D. "Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education." *Journal of Health Economics* 2007; 26(5): 877-895.

Dobkin, C, Shabani, R. "The Health Effects of Military Service: Evidence From the Vietnam Draft", *Economic Inquiry* 2007; 47(1): 69-80.

Glassman, A, Helzer, J, Covey, L, et al. "Smoking, Smoking Cessation and Major Depression." *JAMA* 1990; 264: 1546-1549.

Glied, S. "Is Smoking Delayed Smoking Averted?" *American Journal of Public Health* 2003; 93(3): 412-6.

Glied, S. "Youth Tobacco Control: Reconciling Theory and Empirical Evidence." *Journal of Health Economics* 2002; 21: 117-135.

Grimard, F, Parent, D. "Education and Smoking: Were Vietnam war Draft Avoiders Also More Likely to Avoid Smoking?" *Journal of Health Economics*, 2007; 26(5): 896-926.

Gruber, J, Koszegi, B. "Is Addiction 'Rational'? Theory and Evidence." *Quarterly Journal of Economics* 1991; 116: 1261-1303.

Gruber J, Zinman J. "Youth Smoking in the U.S.: Evidence and Implications." In: Gruber J, ed. Risky Behaviors among Youths. Univ of Chicago Press 2001: 69-120.

Harris, R, Hebert, J, Wynder, E. "Cancer Risk in Male Veterans Utilizing the Veterans Administration Medical System." *Cancer* 1989; 64: 1160-8.

Hearst, N, Newman, T Hulley, S. "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment." *New England Journal of Medicine*; 1986.

Kenkel D, Lillard, D, Mathios A. "Accounting for Misclassification Error in Retrospective Smoking Data." *Health Economics* 2004; 12: 1031-1044.

Klevens, R et al. "The Association between Veteran Status and Cigarette-smoking Behaviors." *American Journal of Preventive Medicine* 1995; 11(4): 245-250.

Levy D, Cummings, K, Hyland, A. "A simulation of the effects of youth initiation policies on overall cigarette use." *American Journal of Public Health* 2000; 90: 1311-4.

Merline, A, O'Malley, P, Schulenberg, J, Bachman, J, Johnston, L. Substance Use among Adults 35 Years of Age: Prevalence, Adulthood Predictors, and Impact of Adolescent Substance Use. *American Journal of Public Health* 2004; 94: 96-102.

Orzechowski, W, Walker, R. The Tax Burden on Tobacco 2006. V41. Arlington, VA.

Schroeder, S. "Tobacco Control in the Wake of the 1998 Master Settlement Agreement." *New England Journal of Medicine* 2004; 350: 293-301.

Sonnenschein, E, Brody, J. "Effect of Population Aging on Proportionate Mortality From Heart Disease and Cancer, U.S. 2000-2005." *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 2005; 60: S110-S112.

Stanton, C., et al. Consistency of Self-reported Smoking over a 6-year Interval from Adolescence to Young Adulthood. *Addiction* 2007; 102: 1831-1839.

Stellman, S, Stellman, J, Koenen, K. "Enduring Social and Behavioral Effects of Exposure to Military Combat in Vietnam." *Annals of Epidemiology* 2000; 10(7): 480.

U.S. Department of Agriculture Economic Research Service. Tobacco Briefing Room: www.ers.usda.gov/Briefing/tobacco, Tables 40, 131, and 132. Accessed June 28, 2007.

U.S. Department of Defense, Office of the Inspector General. "Economic Impact of the Use of Tobacco in DOD." Report 97-060, Dec. 1996.

U.S. Public Health Service, Surgeon General. "Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report from the Surgeon General." No. 89-8411, 1989.

U.S. Department of Health and Human Services. A Report of the Surgeon General: the Health Benefits of Smoking Cessation. Washington (DC): 1990.

Wald, A. "The Fitting of Straight Lines if Both Variables are Subject to Error." *Annals of Mathematical Statistics* 1940; 11(3): 284-300.

West, R., et al. Can We Trust National Smoking Prevalence Figures? Discrepancies Between Biochemically Assessed and Self-Reported Smoking Rates in Three Countries. *Cancer Epidemiology Biomarkers and Prevention* 2007; 16: 820-822.

Winefield HR, et al. "Psychological Concomitants of Tobacco and Alcohol Use in Young Australian Adults." *British Journal of Addiction* 1989; 84: 1067-1073.