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

Brian Rowe†
Bureau of Economics
Federal Trade Commission
browe@ftc.gov

September 2009

Abstract

In the United States, a majority of the drivers who receive a traffic ticket are male, and male drivers are more likely to receive a ticket after being stopped by the police. This paper develops and conducts an empirical test for the existence of police gender bias (taste-based discrimination) in traffic ticketing. The test is based on a model's prediction of how the gender composition of ticketed drivers should vary across groups of police officers who use unbiased, but potentially different ticketing standards. The test is useful for determining whether the gender disparity in traffic tickets results from gender bias or a higher tendency of male drivers to break traffic laws. In addition, the test offers an improvement over the "differences-in-differences" test for discrimination which has been applied in other contexts. When applied to data on traffic tickets issued by male and female police officers in Boston, the new test rejects the null hypothesis of unbiased ticketing.

JEL Codes: J16, J71, K42

*I am grateful to Martha Bailey, Alexia Brunet, Osborne Jackson, Zoe McLaren, Matt Rutledge, J.J. Prescott, Dan Silverman, Jeff Smith, 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 meeting for many helpful comments. I thank Kate Antonovics, Bill Dedman, and Nicola Persico for sharing data.

†The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Trade Commission.

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 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

[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).

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 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

[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.

[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.

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 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] This is done by including the interaction of subject and evaluator race or gender in the model for the outcome. The idea underlying the estimation of these "cross-gender" or "cross-race" effects (which are difference-in-differences estimates, as shown in the Appendix) 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.

1.1 Recent Related Literature

Makowsky and Stratmann (2009), 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

[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 (2009), Bagues and Esteve-Volart (2007), Price and Wolfers (2007), and Schanzenbach (2005).

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 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 (2009) 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 (2009). 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 (2009) 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 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.

2.1 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 $t(d_m, p_g) \neq t(d_f, p_g)$.*

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:

[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.

Proposition 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 (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 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:

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

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:

[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.

$$Odds(p_g) = \frac{P(T \mid d_f, p_g)}{P(T \mid d_m, p_g)} = \frac{1 - F_f\{\theta^*(d_f, p_g)\}}{1 - F_m\{\theta^*(d_m, p_g)\}} \tag{2}$$

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) shows how these ticketing odds are produced for each officer gender.

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) 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 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 (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. *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:

$$\frac{\partial Odds(p_g)}{\partial \theta^*} = \frac{\int_{\theta^*}^{\infty} f_m(\theta^*) f_f(\theta) - f_f(\theta^*) f_m(\theta) d\theta}{(1 - F_m\{\theta^*\})^2} < 0 \tag{3}$$

Proposition (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) 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 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 (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]

[12] Using Proposition (3), a test analogous to that proposed by Anwar and Fang (2006) can be derived. This is discussed in Section 4.

Violation severity θ is unobserved for individual tickets, but Proposition (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.

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 (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 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 $E[mph \mid T, \theta^*(p_m)] < E[mph \mid T, \theta^*(p_f)]$.*

[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.

See the Appendix for the derivation. Proposition (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 (4) shows that 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

[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.

the violation thresholds. Here the idea is that violations which are punished with higher fines are more dangerous.

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.

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 (2009).

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 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]

[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.

[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.

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 (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.

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:

$$\text{Gender disparity} = P(T \mid d_f) - P(T \mid d_m) \tag{4}$$

The quantities in equation (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 (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 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:

$$\frac{P(d_f \mid T)}{P(d_m \mid T)} = \frac{P(T \mid d_f)}{P(T \mid d_m)} * \frac{P(d_f)}{P(d_m)} \tag{5}$$

Equation (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:

$$EOdds(p_g) = \frac{P(d_f \mid T, p_g)}{P(d_m \mid T, p_g)} \tag{6}$$

Recall equation (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 (5) for both male and female officers and dividing gives:

$$\frac{EOdds(p_m)}{EOdds(p_f)} = \frac{Odds(p_m)}{Odds(p_f)} \times \frac{P(d_m \mid p_f)}{P(d_f \mid p_f)} \frac{P(d_f \mid p_m)}{P(d_m \mid p_m)} \tag{7}$$

The last term on the right hand side of (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:

$$\ln\left(\frac{P(d_f \mid T)}{1 - P(d_f \mid T)}\right) = \beta_0 + \beta_1(\text{Male Officer}) + \beta_2(\text{Controls}) \tag{8}$$

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 2009). 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) shows,

---

[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.

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

female police officers are distributed fairly evenly across the Boston police districts, and are never more than 20% of the force in any district. In 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 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 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 (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) 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

[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.

[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."

correspondence with the author, the Boston Police Department stated that: "Both male and female officers perform the same functions within the Department."

3.3 Results

Table (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 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 (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 (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 (predawn, 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 (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 (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 (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) 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 (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 (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.

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 (1) and referring to Proposition (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 (5), and two additional controls: The gender of the driver and an interaction of driver and officer gender. According to Proposition (4), driver gender can be omitted in order to rank officer's thresholds, and specifications omitting driver gender (two are shown in Table (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 (2009). 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 variables, 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 (2009) create their mismatch variable as the sum of two interaction terms: Black Officer $\times$ White Driver + White Officer $\times$ Black Driver. If created in this way, the gender mismatch coefficient would be equal to the difference-in-difference shown above divided by two. The derivation of these equivalences are shown 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. Analogously, Price and Wolfer's (2007) interaction term of interest (the interaction of player race and referee crew race) captures how the racial

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

disparity in player foul rates varies by referee crew race. However, the ticketing model developed in Section 2 indicates that such variation by officer gender (or referee race) may not be due to bias when one group of officers (or referees) is tougher than the other.

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. Figure (3) illustrates this example graphically. In the Figure, the tough officer tickets drivers at 60 miles-per-hour while the easy officer tickets at 70, and both officers are unbiased because they apply their thresholds equally to male and female drivers.

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 percentage points) so I ignore this issue here.

The results used for ranking officer's ticketing thresholds are shown in Table (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 (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.

[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.

The ordered logit and OLS results shown in Table (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 (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 (8). The measures of tickets written had consistent effects across all the specifications. When more tickets were issued, ticketed violations occurred at 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

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

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 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 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 6 and 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

[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.

[27] To assess the sensitivity of my standard errors, I estimated the baseline specifications in Tables 5 and 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).

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 female 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 (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:

$$\frac{N_f^m + S}{N_m^m - S} = \frac{1}{\beta}\frac{N_f^m}{N_m^m} \Leftrightarrow S = \frac{(1-\beta)N_m^m N_f^m}{\beta N_m^m + N_f^m} \tag{9}$$

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.

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

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 (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 (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 (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 (2009), 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 the driver is different from the race of the police officer. Their theoretical model indicates that this cross-race effect is due to

[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.

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 (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 × 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 (10), I verified that creating gender mismatch as Male Officer × Female Driver + Female Officer × Male Driver results in coefficients on mismatch equal to those reported divided by two.

In their model, Antonovics and Knight (2009) 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.

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 (the difference-in-differences test). 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.

6 Appendix

6.1 Proof of Proposition 2

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

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.

6.2 Proof of Proposition 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

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

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

6.3 Proof of Proposition 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) it can be shown that $\frac{\partial s(\theta^*)}{\partial\theta^*} < 0$ (available by request). From Proposition (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 $\bar{\theta}(d_m) \geq \bar{\theta}(d_f)$.

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 (2009), 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:

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

7 Tables and Figures

Table 1: Sample means of traffic ticket variables by police officer gender.

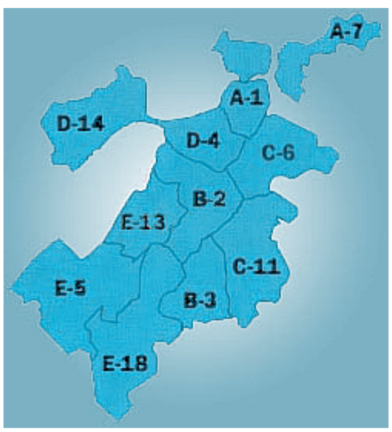
	Male Officers	Female Officers
Female Driver	28.4%	30.5%
Weekend	24.2%	21.2%
Commute	55.4%	50.8%
Daytime	66.8%	73.2%
Speeding	25.0%	17.6%
Failure to Stop	31.8%	32.1%
No Inspection Sticker	8.9%	4.6%
No Seat Belt	14.9%	5.8%
MPH Over Limit (speeding)	14.3	15.4
	(n=34,133)	(n=923)
Fine Amount (non-speeding)	$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: 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 1: Cumulative distribution of miles-per-hour over limit for ticketed drivers.

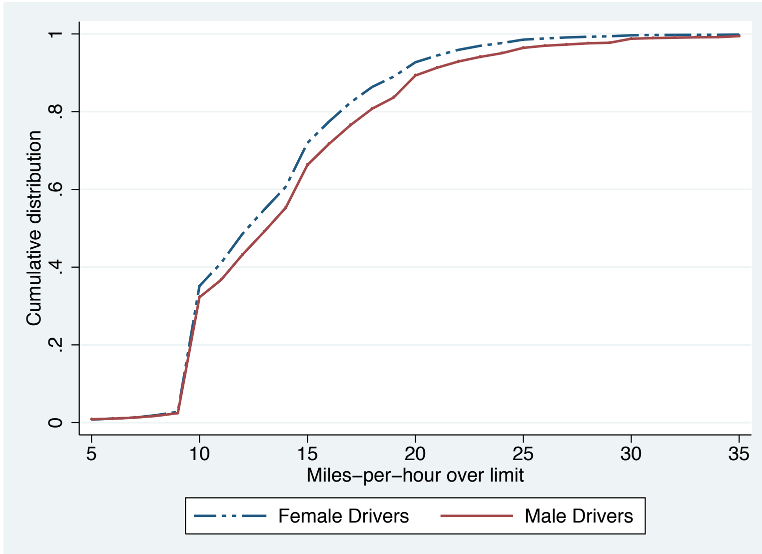


Figure 2: Total citations by month and officer gender.

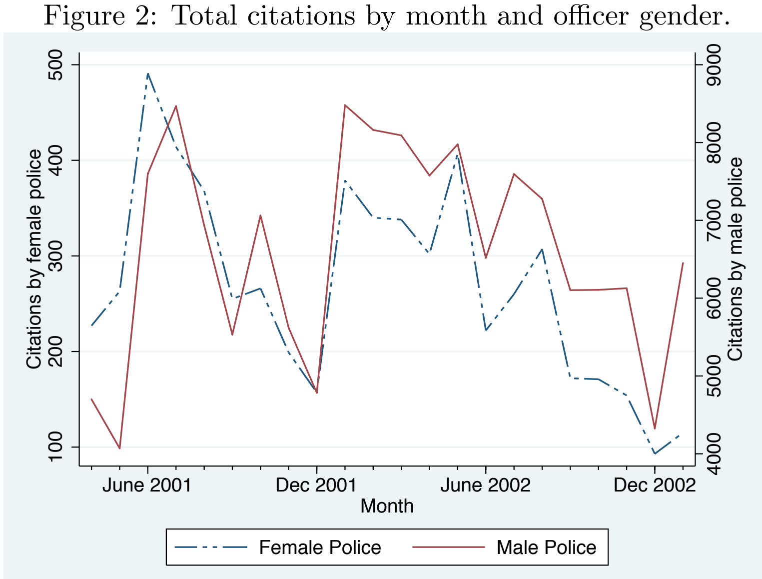


Table 3: Means of 2000 Census variables for Boston metropolitan area police officers

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

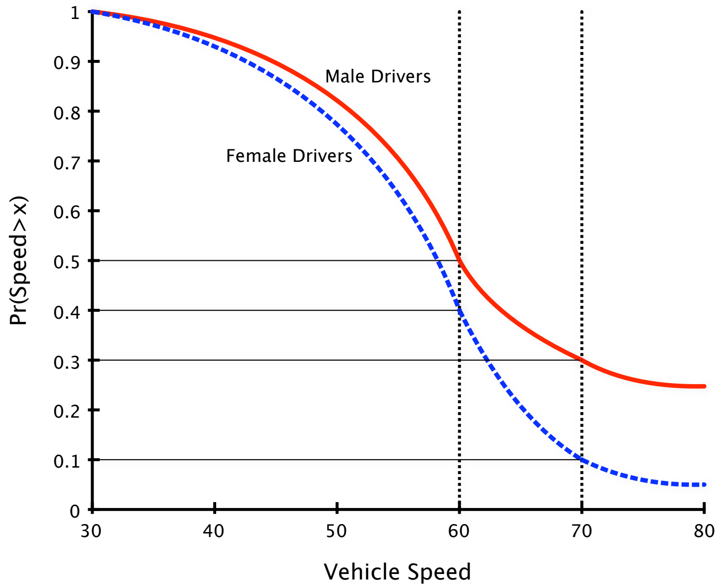
	Speeding tickets				Other violations			
Female Driver (Yes=1)	OLS	Logit	Logit	Logit	OLS	Logit	Logit	Logit
Male Officer	−0.032**	0.860**	0.828**	0.882*	−0.022**	0.900**	0.854**	0.901**
	(0.016)	(0.061)	(0.060)	(0.064)	(0.007)	(0.030)	(0.029)	(0.030)
MPH over limit	−0.006**	0.968**	0.969**	0.969**				
	(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

Figure 3: Example of Tough versus Easy ticketing.



Tough officers ticket drivers at 60, Easy officers ticket drivers at 70.

Both officers are unbiased, but Tough tickets relatively more female drivers.

Table 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 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 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 10: Test analogous to Antonovics and Knight (2009): 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

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”, *Review of Economics and Statistics*, February 2009.

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?”, *American Economic Review*, March 2009.

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.