“Whose help is on the way?”
The importance of individual police officers in law enforcement outcomes

Emily Weisburst*

December 2018

Link to Download Latest Version

Abstract

The public’s perception of police fairness is essential to the willingness of citizens to cooperate with the police and is fundamental to establishing police legitimacy. However, little is known about whether police officers are actually fair and impartial in their application of the law. In this paper, I show that the likelihood of an arrest is not only a function of incident timing, geography, offense type, and other contextual factors but also critically depends on the identity of the police officer who responds to a call for service. The analysis examines detailed data on more than 1,850 police officers responding to over 230,000 offenses reported through calls for service from the Dallas Police Department. I find that police officers are important determinants of arrest outcomes, with individual officer behavior accounting for 10-15% of the explainable variation in arrests. Officers vary widely in their arrest behavior, with a 1 standard deviation increase in an officer’s propensity to arrest resulting in a 32% increase in the likelihood of arrest. Additionally, I apply a test of taste-based racial bias and fail to find conclusive evidence that officer differences are driven by racial bias in this setting.

JEL Classification: K40, K42, J45

Keywords: Police Discretion and Decision-Making, Racial Bias in Policing, Economics of Crime

*Assistant Professor, Luskin School of Public Affairs, University of California, Los Angeles. Contact address: Department of Public Policy, UCLA Luskin School of Public Affairs, 337 Charles E. Young Dr E., Los Angeles, CA 90095. Email: weisburst@ucla.edu. I am grateful to Professors Sandra Black, Jason Abrevaya, Leigh Linden, Becky Pettit, Rich Murphy, Steve Trejo, and Brendan Kline for their helpful comments and feedback. I also thank members of the Dallas Police Department that shared information about their operations, protocols and data practices for this project. This project received funding from the University of Texas Continuing Graduate Fellowship. The views presented in this paper are my own and do not necessarily reflect the views of the Dallas Police Department. First Version: July 2017.

1 Introduction

The public's perception of police fairness is essential to the willingness of citizens to cooperate with the police and is fundamental to establishing police legitimacy (Tyler et al., 2015; Frydl and Skogan, 2004). In recent years, public trust in police has become increasingly strained; recent survey evidence finds that less than 30% of individuals in high-crime and low-income areas believe that police "make fair and impartial decisions in the cases they deal with" or that police "make decisions based on the law and not their personal opinions or beliefs" (La Vigne et al., 2017). Distrust in police is related to the perception that officers are racially biased; according to national polling, 69% of Black and 54% of Hispanic Americans think that "police tend to target minorities," compared to only 29% of White Americans (Schneider, 2015). In 2015, American confidence in police officers reached its lowest point in more than 20 years, driven by high profile police use-of-force incidents and shootings (Jones, 2015). In the ensuing debate, pundits have often made conflicting claims about the causes of these high profile police incidents, sometimes asserting that "any officer would have responded in the same way" and at other times claiming that the events are "isolated incidents attributable to 'bad actors' that do not reflect the rest of a department."[1]

However, little is known about whether police officers are actually fair and impartial in their application of the law. On a more basic level, there is scant evidence of whether officer decisions actually matter to the outcomes of police interactions after considering the context of an incident. Further, if police officer decisions do matter, little is known about how much officers differ from each other in their actions. If there are differences in police officer behavior, how large are these differences? Is officer arrest behavior characterized by racial bias, and if so, how important is race in explaining officer behavior?

Managers in police departments face a similar principal-agent problem as managers in firms; they are impacted by the behavior of individual officers but cannot fully control officers' actions given limited resources. Understanding the trade-offs between alternative

[1]This phrasing is not directly attributed to any single pundit or public figure. An example of the first argument can be found in opinion pieces through the organization *Blue Lives Matter*, which was established as a reaction to the *Black Lives Matter* movement (BlueLivesMatter, 2016). The second argument was recently invoked by Attorney General Sessions as a reason to cease the Department of Justice's enforcement of consent decree agreements with police departments, which were established to address civil rights concerns related to law enforcement actions (Department of Justice, 2017).

monitoring policies is an important area of study, particularly in the setting of policing, where there are large potential consequences for both public safety and the civilians that interact with the police. This paper measures the scope of individual police officers to impact law enforcement outcomes, a necessary first step to clarifying these trade-offs.

Police officers often operate in the field, alone or in small groups, and have substantial legal latitude in their conduct and response to different situations. At the same time, police departments are increasingly incorporating technology and data to standardize operations, potentially limiting the importance of individual officer decisions in police work.[2] The ability of police officers to invest effort differently across different types of offenses can be a productive form of police discretion when police resources are limited and there is a trade-off between exerting effort on serious and non-serious crimes. However, within particular offense types, behavioral differences across officers are more likely to result from differences in officer skills, experience, and preferences. This study is the first to estimate the degree and importance of police discretion across officers, conditional on offense context.

I analyze police officer behavior using a sample of over 230,000 offenses reported through calls for service (or 9-1-1 calls) and over 1,850 police officers from the Dallas Police Department in Texas. I estimate the contribution of individual officers to predicting arrests, controlling for detailed information on the characteristics of offense incidents, including call urgency and dispatch code, complainant characteristics, and time and geographic factors. I then isolate each police officers' propensity to make arrests and estimate the degree of dispersion in this propensity across officers. Throughout the analysis, I pay particular attention to patterns of officer responses to particular offenses, and conduct a number of robustness checks to verify that the observed dispersion in arrest behavior across officers is not due to selection.

I find that a 1 standard deviation increase in an officer's arrest propensity results

[2] Criminologists have long noted that police work is characterized by discretion, with researchers adopting a broad definition of discretion that encompasses variation in work-related decisions, interpretation and implementation of the law, and the use of extra-legal factors, such as suspect race, in decision-making (Nickels, 2007; Frydl and Skogan, 2004; Mastrofski, 2004). Technological advancements have increased police reliance on data for surveillance of suspects, tracking and monitoring of police employee activities, automation of reporting, and focusing patrol on areas with high offending rates, or crime "hot spots." Advancements in technology have the potential to exacerbate differences in police treatment of civilians, or could serve to reduce police discretion (Brayne, 2017; Joh, 2016).

in a 32% increase in the likelihood of arrest, suggesting there is substantial variability in police officer responses to similar offenses. Further, individual police officers account for approximately 10-15% of the explainable variation in arrest outcomes. Observable officer characteristics, including salary, experience, age, race and gender, account for only a small portion of the variation in officer arrest behavior.

In economics, research on police decision-making has largely focused on measuring racial bias in police traffic stops (e.g. Horrace and Rohlin, 2016; Anbarci and Lee, 2014; Antonovics and Knight, 2009; Gelman et al., 2007; Anwar and Fang, 2006; Grogger and Ridgeway, 2006; Knowles et al., 2001). New work by Fryer (2016) extends this literature to officer decisions to use violent force. Collectively, this literature has found mixed evidence that police officers exhibit taste-based preferences for racial discrimination, with results that vary by the study setting and the test used to detect racial bias. While the literature has frequently exploited aggregate officer demographic characteristics, nearly all of the work in this space does not incorporate officer identity in measuring racial bias. This paper complements new work by Goncalves and Mello (2017), that measures individual officer effects in a racial bias test applied to officer decisions to issue speeding tickets.

I fail to find conclusive evidence of taste-based racial bias among officers, despite the fact that I document large differences in total arrest behavior across officers. I adapt a test of racial bias proposed in Anwar and Fang (2006) to the regression framework which leverages relative differences in officer arrest rates across officer and arrestee race groups. I find that White officers are more likely to arrest Black individuals than Black or Hispanic officers; however, I cannot reject the hypothesis that White officers arrest all suspect races more than other race officers. Further, I find that most of the variation in officer arrest behavior occurs within rather than across officer race.

In addition to providing the first estimate of officer-level police discretion, this paper makes several other contributions. First, the analysis in this study uses offenses reported by civilians through 9-1-1 calls, a setting that researchers have not exploited to study police decision-making.[3] I am able to explicitly measure the importance of officers versus offense

[3]To the best of my knowledge, this is the first study that uses high frequency complainant offense reports to study police officer behavior and decisions.

context because the incident setting is given at the time of the offense response.

With the exception of West (2015), who studies racial bias of state troopers who are randomly dispatched to motor vehicle accidents, researchers in economics have typically restricted their attention to interactions that are initiated by police officers, such as general traffic stops, stop and frisk interviews, and speeding tickets. Importantly, these interactions are a choice variable of the police officers involved. A growing body of research finds that race can also be a factor in police decisions to make traffic (or pedestrian) stops, and that studies that focus only on the outcomes of traffic stops neglect to consider police discretion that contributes to sample selection (Horrace and Rohlin, 2016; Gelman et al., 2007; Grogger and Ridgeway, 2006). A major advantage of using call for service data to study police discretion is that each observed offense is originally initiated by a complainant and not by a police officer.

An additional contribution of this study is the variety of policing interactions that I examine. The analysis sample covers a diverse cross-section of police work, allowing examination of responses to different types of offenses. Detailed demographic information on officers, arrestees, and civilian complainants also provide rich controls in the model and enable tests of racial bias.

2 Institutional Background and Description of Data

2.1 *Dallas Police Department Data*

The setting for this study is the Dallas Police Department in Texas. Dallas is a large and diverse urban center, with over 1.2 million residents and a population that is 42% Hispanic, 24% Black and 29% White (Census, 2015).[4] Crime rates in the city of Dallas are similar to other cities of its size in the U.S., with 694 violent crimes and 3,440 property crimes per 100,000 residents in 2015 (FBI, 2015).

In recent years, the Dallas Police Department (DPD) has enacted several police reforms, including increasing officer training in de-escalation techniques and racial bias, em-

[4]I capitalize Black, Hispanic, and White for stylistic consistency throughout the paper.

ploying body cameras, and firing some of its most poorly performing officers (Haugh, 2016; Tsiaperas, 2016). In 2015, DPD joined the Obama Administration's Police Data Initiative and released a number of data sets on its operations. This project uses public data released by DPD covering responses to offenses reported by victims through 9-1-1 calls, records of persons involved in these offenses (arrestees and complainants), the names and badge numbers of responding officers and arrest records between June 2014 and mid-August 2018.[5]

The DPD offense data represents a subset of the universe of call for service incidents. First, these data exclude responses to calls for service where there is no determined offense. Second, for privacy reasons, the DPD data files also exclude records for sexually-oriented offenses, offenses involving juveniles, and offenses involving social service referrals. I use a liberal definition of arrests, coding an offense as having an arrest if any of the DPD data files obtained for the study indicate that an arrest was made. Additionally, I merge the DPD data sets with demographic information on police officers also obtained through an open records request to the city of Dallas. The officer data includes officer race, gender, salary, and job title (see Data Appendix A5 for more detail).

2.2 *Protocols for Dispatch and Offense Responses*

When a civilian calls DPD for police assistance, they are connected to a 9-1-1 call-taker. The call-taker creates an active call report that summarizes important facts related to the incident, including location and relevant descriptions of the events. Active call reports also include a dispatch code that categorizes the incident type. Given a set of open active calls, DPD dispatchers work with police officers to assign available officers to incidents. Calls are dispatched according to their priority, or their level of severity and urgency. When there is a long call queue, responses to low priority calls are postponed until more serious calls have been resolved. The pool of available officers when a call is received depends on patrol

[5]The offense data includes some interactions that were initiated by officers rather than complainants; I use a number of data fields to clean the data to exclude these offenses. I identify officer-initiated offenses through the listed complainant; excluding those that originate with a city or police, sheriff or fire department, and through the dispatch offense code; excluding codes associated with officer initiated offenses. Further, I supplement the offense data with data on all dispatched incidents obtained through an Open Records Request to the city of Dallas, which allows me to view officer dispatches to any incident, including non-offenses. I use this data to restrict the offense sample to calls with a match in the dispatch data.

responses to other incidents at the time. (Figure A1 depicts the response process in Dallas).

Patrol officers are the primary responders to calls for service. Officers are assigned to work in 1 of 7 police divisions in the city for 8-hour shifts, or watches, from 12am-8am, 8am-4pm, and 4pm-12am.[6] Regular patrol shift schedules are set once a year, based on the seniority of officers.[7] With minimal exceptions, calls are assigned to patrol officers who work within the geographic police division where the call incident occurred.

Officers typically conduct patrol in police cars, alone or in pairs. At the beginning of each shift, officers may choose to patrol with another officer, depending on the number of cars available for that shift. Each car is considered an "element" that can be dispatched to an incident. Throughout a shift, paired officers respond to all calls together.

If more than one patrol element is available to respond to an incident at the time of dispatch, dispatchers consider a number of factors in their assignment of available officers. More serious incidents may require or benefit from a response by multiple officers.[8] Additionally, officers who are geographically close to an incident are more likely to be dispatched to the incident, especially if the call is urgent. At the same time, depending on availability, officers may volunteer to take particular calls as they are posted, a potential source of selection.

When the assigned patrol element arrives at the scene of the incident, the responding officer(s) determines if an offense occurred, gather information, investigates the scene and assists the complainant or victim. The sample used in this study is restricted to the subset of calls where officers determine that an offense occurred, representing $\approx 10\%$ of dispatched responses.[9] The offense designation is driven by both the incident circumstances and the complainant/victim. Offense reports almost always require complainant interest in pursuing

[6]The three 8-hour shifts listed are approximate, in practice some officers work 10 hour shifts and other officers have start and end times that are slightly staggered across police shifts.

[7]Depending on the needs of the department, officers may choose to work overtime patrol shifts outside of their regular shift schedules, though these shifts are also set in advance, typically a month or a week prior.

[8]In serious incidents, additional officers may be called to the scene, through the request of the responding officer or dispatcher, or because another officer volunteers to participate in the call.

[9]The dispatch response sample may be inflated in the sense that it includes some officer-initiated incidents or officer "mark-outs" to dispatchers, responses to assist officers on existing responses and responses to multiple calls that are later determined to be associated with a single incident. I exclude records of offenses that cannot be matched to the dispatch data.

investigation, filing a formal report and/or pressing charges against a perpetrator.[10] Complainant compliance may be affected by the interaction with the responding officer(s), and in this way, the offense designation is a potential source of selection.

If officers choose to respond to offenses based on incident characteristics that are unobservable, the estimates in the study will impacted by selection bias. To address this concern, I conduct a series of tests to verify that selection does not affect the empirical estimates (see Section 4.3).

After the responding officer(s) files the offense report, it is submitted to a staff reviewer at DPD who examines the report for completeness. After this initial review, the offense may be assigned to a detective in an investigative unit based on the offense type. The assigned detective will then pursue additional investigation of the offense if warranted.

Over the course of an offense response, officers may identify a suspect and/or make an arrest. Alternatively, an arrest may be made at a later date after a detective assumes responsibility for a follow-up investigation of the offense. 8% of offenses reported through calls for service result in an arrest. Individual responding officers have the ability to influence arrests directly, by making the decision to apprehend an individual at the scene of the incident, or indirectly, by laying the groundwork for an investigation through gathering information for the offense report. In practice, most arrests do not involve a prolonged follow-up investigation and the responding officer is typically involved in the arrest.[11]

3 Empirical Model

I focus on two metrics to assess the importance of individual officers in arrest outcomes. First, I measure the collective importance of individual officers' arrest propensities in predicting arrest outcomes. Second, I estimate each officer's arrest propensity and measure the dispersion in this propensity across officers. The first measure summarizes how important

[10]In cases when there is substantial evidence at the scene, complainant or victim compliance is not necessary to designate an incident as an offense.

[11]Of the data with information on arrest dates (57% of arrests), 94% of arrests occur within a day of the incident offense. When there is information on the arresting officer (58% of arrests), 89% of arrests involve the original responding officer. The first rate is imputed from direct information on arrest dates when available, as well as the data upload date for an arrest record when not available.

individual officer decisions are relative to the context of an offense, while the second measure captures how different officers are from one another.

As a first step, I estimate the following linear probability model,

$$Arrest_{ikgt} = \theta_i + \theta_j + \pi X_{kt} + \delta_{gt} + \phi_g + \varepsilon_{ikgt}$$

where i indexes the responding officer, j indexes a co-responding officer (if present), k indexes the offense, g indexes geographic location, and t indexes time. The outcome $Arrest_{ikgt}$ is the primary focus of the analysis and denotes whether an arrest was made in association with the offense. X_{kt} are a set of incident specific characteristics, including 22 aggregated dispatch codes and 11 location type codes (e.g. street or residence) and indicators for the hour within a patrol shift. X_{kt} also contains complainant characteristics, including the number of complainants, whether there was a victim injury, and the race and gender of complainants.[12] Importantly, the model also controls for the *urgency or severity* of the call, defined as the number of minutes that pass between when a call occurs and the time of dispatch (entered as a linear and quadratic term).

ϕ_g are indicators for police beat locations and control for time-constant differences in arrest patterns across geography. There are 234 beats in Dallas and each is fully contained within 1 of the 7 police divisions in the city. δ_{gt} are shift indicators that capture time-varying location-specific arrest patterns that are associated with specific shift assignments. These variables are Police Division*Day-of-the-Week*8-hour Shift*Month*Year fixed effects. To increase power, I do not include a separate indicator for each individual shift, but rather aggregate them into month by year groups. For example, the four Tuesday evening shifts in the Central Division are grouped in January 2016.

θ_i measures the time-invariant or permanent arrest propensity of officer i. Given the numerous controls in the empirical model, θ_i represents an officer specific effect that is measured within dispatch type, shift and geographic location. In cases when there are two responding officers, I include a control for the identity of the other responding officer, θ_j. Observations with two responders are duplicated, so that each officer gets a record of

[12]For calls with multiple complainants, I define each complainant variable using the maximum value for the complainant group, allowing complainants to have multiple races and genders.

participating in the offense response through the θ_i term. This procedure allows measurement of individual officer effects net of the effects of a co-responder, using a model similar to prior work on peer effects in production (e.g. Silver, 2016; Mas and Moretti, 2009). In this way, the specification addresses omitted variable bias related to police officer decisions to pair with another officer, as well as potential direct effects attributable to the co-responder.[13] I restrict the sample to observations where the officers responding have at least 50 offense records to improve precision in the estimation of θ_i.[14] [15]

Using this model, I calculate the two metrics of importance in this study. First, I measure the importance of the θ_i terms as predictors in the model, by calculating the proportion of explainable variation that is attributable to these parameters. I do this by estimating the R^2 (and Adjusted R^2) from the full model and the model without θ_i and θ_j terms included. I then calculate the proportion of the total R^2 (and Adjusted R^2) given by the including officer fixed effects: $(R^2_{total} - R^2_{w/oFE})/R^2_{total}$. I interpret this ratio as the relative importance of individual officers in explaining the model variation in arrest outcomes.

Second, I calculate the dispersion in officer-level permanent arrest propensity as the standard deviation of the distribution of θ_i across officers. In order to establish a conservative estimate of police officer dispersion, I adjust the estimates of θ_i terms using Empirical Bayes techniques.[16] Throughout this paper, I focus on results using these adjusted estimates, and

[13]The results in this paper are robust to two alternate formulations that exclude fixed effect terms for co-responders, θ_j. These alternate specifications are (1) observations that are duplicated when there are co-responders but exclude co-responder fixed effects and (2) observations that are not duplicated when there are co-responders but only consider effects from the first listed responder. The preferred model officer effects are highly correlated to the officer effects from alternate formulations with correlations of 0.73-0.82.

[14]This restriction excludes 6% of offenses in the sample, but allows estimation of fixed effects to be based on a reasonable number of observations per officer. 10% of co-responder offense responses include only records for only one responder given this restriction. Further limiting the sample to exclude these "one-sided" observations does not affect the results.

[15]I use an algorithm developed in Correia (2016) to estimate the large number of fixed effects indirectly through an iterative procedure that provides a point estimate value for each fixed effect. In the base model, there are 322,280 observations, 1,858 first officer categories, 2,658 second officer categories, 7,790 shift categories and 234 beats (after excluding singletons). Rather than estimating the model's fixed effects by including indicator variables as controls in the model, this algorithm effectively initializes each fixed effect within a group, and then iterates the estimation until both the sum of squared residuals is minimized and the coefficient on each set of fixed effect terms is 1. This procedure is programmed in the STATA command `reghdfe`, and is notable for its fast computation time. I estimate all sets of fixed effects in the model jointly in this way, or $\theta_i, \theta_j, \delta_{gt}$ and ϕ_g.

[16]I calculate the adjusted estimates of θ_i using the following steps. First, I construct a composite residual term, $\hat{r}_{ikgt} = \hat{\theta}_i + \hat{\varepsilon}_{ikgt}$, and an average officer residual, $\bar{r}_i = \frac{1}{N_i} \sum_{N_i} \hat{r}_{ikgt}$. This residual is estimated using a model that includes officer fixed effects in the first stage to allow for arbitrary correlations between

refer to these adjusted estimates as $\hat{\theta}_i$. The results are not an artifact of this adjustment and are qualitatively similar when unadjusted fixed effects from the first stage are used.

4 Results

4.1 *Summary Statistics*

Tables 1.A and 1.B summarize the data. The first column covers the total sample at the offense-level, while the second column covers the analysis sample, which restricts the sample to officers with 50 or more offense responses and duplicates observations with two responding officers. The analysis sample includes over 230,000 incidents and 320,000 observations for 1,858 officers.

Hispanic complainants are associated with 28% of offenses and are underrepresented relative to the Dallas population, which is over 40% Hispanic. In contrast, White officers and Black arrestees are over-represented relative to the population of Dallas. White patrol officers respond to 50% of offenses, while Black and Hispanic officers respond to 25% and 22% of offenses, respectively. Relative to the sample of offenses with demographic information for arrestees, 50% of offenses have a Black arrestee, 25% have a White arrestee, and 25% have a Hispanic arrestee.[17] The average officer arrest rate as a portion of his responses is 9%. Approximately 9% of offense responses involve a police officer in training, 1% involve a police sergeant, and 15% involve a female officer. Averaged across offense responses, DPD patrol officers earn approximately $58,000 per year. 38% of offense responses involve two responding officers.

On average, it takes 24-25 minutes for a patrol officer to be dispatched to an offense

---

responding officers and the other covariates in the model, in a manner similar to Chetty et al. (2014). Next, I calculate the adjusted officer arrest propensity using the following transformation: $\hat{\theta}_i^{EB} = \sigma_A^2/(\sigma_A^2 + \frac{\sigma_{\varepsilon,i}^2}{N_i}) \cdot \bar{r}_i$. The value of $\sigma_A^2 = \sigma_r^2 - \sigma_\varepsilon^2$, where σ_r^2 is computed using the sample analog of the average squared composite residual and σ_ε^2 is the average squared within officer composite residual, each calculated from the first stage model. The "shrinkage factor," $\sigma_A^2/(\sigma_A^2 + \frac{\sigma_{\varepsilon,i}^2}{N_i})$, adjusts officer arrest propensity toward zero when the number of observations per officer, N_i, is small, or the variation in the officer effect, $\sigma_{\varepsilon,i}^2$, is large. The $\hat{\theta}_i^{EB}$ values represent a "posterior" distribution of officer effects, correcting for sampling noise in the estimation (see Appendix A3 for more detail).

[17] Demographic information is not available for each arrest in the data, and covers 50% of arrests.

incident after a call is made, with a standard deviation of 28 minutes. The variation in this dispatch time lag highlights the fact that dispatchers prioritize calls based on their severity and that officers cannot immediately respond to all incidents. The most common dispatch codes are for major disturbances and burglaries of motor vehicles and residences. At the time of dispatch, only a small number of offense incidents are designated as known violent offenses; robberies, criminal assaults, armed encounters, and active shootings collectively comprise $\approx 10\%$ of offenses. A victim is injured in 10% of offenses.

Overall, the samples summarized in Tables 1.A and 1.B are very similar. The only material difference is mechanical; the analysis sample has a larger number of observations with two responding officers, because these responses are duplicated in this sample. This consistency suggests it is suitable to generalize results from the analysis sample.

4.2 *Baseline Results*

I find that the context of a call for service is relatively more important in predicting an arrest than the identity of the officer that responds to the call. However, individual officers are also a significant determinant of arrest outcomes. As discussed earlier, I measure the collective importance of individual officers to predicting arrest outcomes of offenses reported through 9-1-1 calls, by comparing the proportion of variation explained with and without officer fixed effects in the model of arrest outcomes.

Figure 1 shows the contribution of different controls to the total model R^2 and Adjusted R^2. I estimate these relative proportions sequentially from the bottom bar to the top, first adding the patrol shift fixed effects and indicators for hours passed since the beginning of a shift (δ_{gt} and hour indicators in X_{kt}), then police beat location fixed effects (ϕ_g), followed by location type variables, dispatch call types and call severity variables, complainant characteristics (each components of X_{kt}), and lastly officer fixed effects (θ_i and θ_j). Each percentage is calculated as an additional contribution of R^2 to the total or: $(R^2_{currentbar} - R^2_{priorbar})/R^2_{total}$. The explanatory power of the officer fixed effects is therefore the relative contribution of these controls after controlling for all other variables in the model.

I find that the officer fixed effects account for 15.6% of explainable variation measured using the R^2 measure and 12.5% of the explainable variation measured using Adjusted

R^2. Factors specific to an offense, including dispatch code, call severity, location type and complainant characteristics, account for 53 - 64% of the explainable variation in the model. Geography and time variables associated with an offense account for 24 - 31% of the model variation. Because the order of adding variables matters to calculating R^2, I add individual officer effects to the model last to ensure that the estimate of officer importance is as conservative as possible. The total R^2 and Adjusted R^2 of the model is 0.22 and 0.19, respectively, suggesting that officer effects explain 2.4 - 3.4% of the total variation in arrests. When officer fixed effects are the only controls in the model, they can explain 6.9 - 8.2% of the total variation in arrests. (See Appendix A2 for a discussion of the first stage.)

Next, I show that individual police officers vary substantially in their arrest behavior. Figure 2 shows the estimated distribution of officer effects, $\hat{\theta}_i$, calculated using the procedure described in Section 3. For each officer, $\hat{\theta}_i$ represents his/her permanent or time-invariant arrest propensity, conditional on time and geography controls and offense, location, and complainant characteristics. This estimated distribution has a longer right tail, showing that a small number of officers have especially high arrest propensities.

Swapping an officer that has a low arrest propensity with one that has a high arrest propensity can critically change the outcome of a call response. Given that 9% of sample observations result in an arrest, with a standard deviation of 0.29, a 1 standard deviation in $\hat{\theta}_i$ corresponds to 0.1 standard deviations in the total arrest outcome. In percentage terms, a 1 standard deviation increase in an officer's arrest propensity results in a 31.5% increase in the likelihood that a given offense results in an arrest. Further, moving from the 10th to 90th percentile in the officer distribution translates to a 76% increase in arrest probability.

4.3 *Tests of Officer Sorting to Responses*

As discussed above, patrol officers responding to 9-1-1 calls determine whether an offense occurred and can choose to respond to certain to incidents. While I control for a detailed array of observable characteristics, the estimates of individual officer arrest propensity could be biased if officers systematically respond to offenses based on unobservable characteristics.

There are two potential patterns of selection. First, officer arrest propensity could be related to determinations about whether an offense occurred. In this case, the likely selection pattern is that officers who have a low arrest propensity may be less likely to designate marginal incidents as offenses. These officers may have a preference against filing reports and pursuing investigations and may try to persuade a complainant not to file an official offense report. Likewise, officers that have a high arrest propensity may be more likely to designate marginal incidents as offenses, because these officers may believe a lower evidence threshold is required to designate an incident as an offense. This pattern would imply that high arrest officers should respond to more offenses than low arrest officers. Further, this pattern could create a negative correlation between officer effects, $\hat{\theta}_i$, and the error terms, ε_{ikgt}. This negative selection bias would deflate the dispersion in officer fixed effects, and lead to an underestimate of this parameter.

To address this concern, I test whether officers with higher arrest propensities are more likely to designate incidents as offenses. The correlation between arrest propensity and the number of offense responses per officer is 0.029 (Table A2). This minimal relationship suggests that officer determinations about whether an offense occurred are not likely to bias the results. Moreover, the primary conclusion of this paper is that officers differ substantially in their arrest likelihood, so any potential negative selection related to officer designations of offenses will lead to a more conservative estimate of the key parameter.[18]

Selection related to officer choices to volunteer to particular 9-1-1 incidents is potentially more problematic. Here, officers who have a high arrest propensity could prefer to respond to incidents with a higher likelihood of arrest, and officers who have a low arrest propensity could prefer to volunteer for incidents with a lower likelihood of arrest. In either case, the estimates of $\hat{\theta}_i$ will be positively correlated with the error terms ε_{ikgt} and inflate the dispersion in officer fixed effects.

I conduct a number of tests of officer sorting to address this concern. First, I consider how important observable offense characteristics are in the estimation of $\hat{\theta}_i$. I calculate the correlation between the distribution of $\hat{\theta}_i$ from the full model to $\tilde{\hat{\theta}}_i$ estimated from a model that omits offense characteristics that could influence an officer's decision to respond to a call,

[18]In a separate analysis, I have used the full set of dispatch data to estimate the likelihood of each call being designated as an offense incident, conditional on officer identifiers and a limited set of covariates available in the dispatch data. The "offense designation" officer fixed effects are uncorrelated with the arrest propensities, θ_i, with a correlation coefficient of -0.05. I focus on the offense data as the base for the analysis because of the more detailed information available for these incidents.

X_{kt} and ϕ_g. If officer arrest propensity is orthogonal to observable offense characteristics, these distributions will be perfectly correlated. Figure 3(a) shows that the $\hat{\theta}'_i$ distribution is somewhat more disperse than the base $\hat{\theta}_i$ estimates that include offense characteristics; however, the estimates across these distributions have a correlation of 0.913. This high correlation suggests that observable offense characteristics are not very important controls in the estimation of the officer effects distribution.[19]

This test suggests that observable offense characteristics have a small but significant relationship to θ_i. These tests are informative if unobservable offense characteristics are correlated with observables, an assumption that is often applied in tests of endogeneity. However, It is important to note that the detailed controls in this paper capture close to the full set of information available when officers choose to volunteer for a dispatched incident. This feature is likely to reduce concerns about officers sorting to incidents based on unobservable characteristics.

Next, I focus on two settings where officer sorting is less likely to impact the results. The first setting consists of responses to offenses that are dispatched when few officers are available to respond to calls, so officers have less choice in their incident responses. I define this "Low Availability" sub-sample by calculating the proportion of officers on a shift that are not responding to other offense incidents at the time of each offense response. I then keep offense in the sample that have a lower percentage of available officers than the median value within a patrol shift*month cell, δ_{gt}, to account for variation in total staffing across shifts.[20]

In the second setting, I focus on a sample of incidents where an officer's observed response to an offense is *unlikely* to have occurred, and is therefore less susceptible to con-

[19] I have also measured the joint significance of offense and geographic characteristics that officers may use to choose responses, X_{kt} and ϕ_g, in predicting officer effects, θ_i. I do this by duplicating θ_i to the offense-level, regressing these effects on the full model specification, and taking a joint F-test of X_{kt} and ϕ_g. The F-statistic of this test is 1.97 and is significant (with standard errors clustered at the focal officer, i, and shift level, δ_{gt}, to account for duplicated offense records and officer effect outcomes). Despite this significant relationship, the total dispersion estimate of officer effects is not substantively affected by the inclusion of these covariates.

[20] I determine the total officers working on a shift and the rate of available officers using the full sample of dispatched calls, including those that do not result in an offense in the analysis sample. This larger data set permits observing officers responding to calls that result in offenses as well as those that do not. On average 33% of officers working are available to respond to calls in the unavailable sample, relative to an availability rate of 50% in the total sample.

cerns about officer sorting. I define this "Unlikely Response" sub-sample using techniques similar to propensity score matching. I construct the sample by obtaining a predicted response probability from a regression for each officer, i, across all observations. I then code observations as *unlikely* responses if the predicted response probability is below its median value among the actual offense responses for each officer. By construction, this sub-sample of *unlikely* responses consists of incident responses that are not characterized by predictable sorting of officers.

Figure 3(b) and 3(c) show the results of restricting the observations to the "Low Availability" and "Unlikely Response" sample. The graphs show a close match in the distributions across each of the samples and their corresponding baselines. The correlation in officer effects is approximately 0.8 between each of the robustness samples and their corresponding baseline.[21]

Table A2 shows that the estimated dispersion in officer effects is comparable across the analysis sample, the "Low Availability" sample and the "Unlikely Response" sample. First, the proportion of variation explained by the fixed effects is similar across the main sample and the "Low Availability" and "Unlikely Response" samples, accounting for 11 - 17% of the explainable variation across the samples. Second, a 1 standard deviation increase in officer fixed effects increases the probability of arrest by 31.5% in the primary sample, 38.2% in the "Low Availability" sample, and 39.3% in the "Unlikely Response" sample. If dispersion in officer behavior is increased by officer sorting, we would expect the estimates of dispersion to be lower in these robustness samples than the baseline. However, we observe the opposite pattern here. Across each of these settings, the officer effects are highly correlated and the estimates of officer dispersion are qualitatively similar. The evidence that officer dispersion is similar in samples where officer sorting is less likely reduces concerns that officer effects are systematically related to unobservable offense characteristics.

Overall, these tests suggest that the baseline estimates of officer dispersion are not likely to be substantively biased by officer sorting. This pattern is likely related to the

[21] I restrict each robustness sample to officers with at least 50 observations in the relevant sub-sample. I then compare estimates within these sub-samples to corresponding samples that include all observations for these officers. Tables A1.A and A1.B show that the characteristics of the robustness samples are similar to the primary analysis sample.

rich set of controls in the model, which include geography, time, patrol shift controls, and complainant characteristics, as well as a proxy for the urgency or severity of each offense.

4.4 *Robustness Tests of the Empirical Model*

How large is the dispersion in officer arrest behavior? One way to assert that the distribution of officer arrest propensities is meaningful is to benchmark the observed standard deviation in officer effects, $\hat{\theta}_i$, to the amount of variation that would be observed across officers if there were no "true" officer effects. Even in the absence of officer differences, there will be some measured variation in outcomes across officers, simply due to idiosyncratic variation in the error term or "noise."

I use a bootstrap simulation to confirm that confirm that the results in this study reflect actual variation in behavior across officers. This test estimates a distribution of officer dispersion and the explanatory power of officer effects under the assumption of a null hypothesis that the "true" prediction value of all officer effects is 0, and reference the actual model estimate to this distribution.[22] Figure A3 displays the results of these tests using 250 bootstrap replications. These graphs confirm that the estimated variation in officer effects and the contribution of the officer effects to explaining the model variation is not due to noise in the data. Each of the model estimates are well outside the 95% confidence interval given by the bootstrap tests.

I also test the robustness of the model by considering several alternate specifications in Table A3. In column (2), I substitute the 234 police beat categories with narrower geographic area controls of 1,147 police reporting areas in the city of Dallas. Column (3) alternatively

[22] In this test, I calculate the residuals, $\hat{r}$, and predicted outcome values, $\hat{Arrest}$, from a regression that does not include officer fixed effects, thereby assuming a null hypothesis that the true value of all θ_i and θ_j terms is zero. I apply the wild cluster bootstrap (Cameron et al., 2008) to allow errors to be correlated within shift clusters, δ_{gt}, accounting for common shocks within geography, time periods and officer groups, while also permitting error correlation across duplicated offense observations. I apply weights of $w \in \{-1, 1\}$ to residuals $\hat{r}$ that are constant within each shift cluster δ_{gt} and construct a new outcome variable as the predicted outcome plus the weighted residual, $\tilde{Arrest}_b = \hat{Arrest} + w_b\hat{r}$. I then regress this new outcome variable, $\tilde{Arrest}_b$, on the fully specified model that includes officer fixed effects and calculate the dispersion of the officer effects for each iteration. I also recover the R^2 and Adjusted R^2 contribution of officer fixed effects for each iteration of the bootstrap. I have also conducted this test imposing the restriction that $\tilde{Arrest}_b$ is binary in each iteration. To do this, I set the highest values of the outcome variable equal to one such that the mean of $\tilde{Arrest}_b$ equals the mean of $Arrest$ (approximately the top decile given an arrest mean of 10% in the sample. The results of this bootstrap test are similar and are available on request.

substitutes the police beat controls with geographic controls that vary by time, or 1,855 Police Sector*Month category variables. Column (4) includes 32,652 individual 8-hour shift level indicators (Date*8-hour shift*Division) instead of the monthly aggregated shift indicators, δ_{gt}, used in the primary specification. By conditioning on individual 8-hour shifts within police divisions, this specification absorbs variation in arrests at the date by geography level, accounting for factors such as changing weather conditions, holidays, and other day specific events in the city of Dallas. In column (8), I replace the 22 aggregated dispatch codes and 11 aggregated location codes used in the main specification with 119 specific dispatch code and 34 specific location type categories in the raw offense data.[23] Across specifications these additional controls do little to change the analysis. In fact, the correlation between the base distribution of officer fixed effects and these alternate specifications is above 0.95.

Columns (6) and (7) consider alternative procedures to adjust the estimates for precision instead of the Empirical Bayes' method used in the primary results. In column (6), I report the dispersion in unadjusted officer fixed effects from the first stage, where the sample is restricted to officers with more than 100 observations. In column (7), I weight the unadjusted officer fixed effects by the number of observations per officer and calculate a weighted standard deviation as the dispersion metric. Across both of these alternative precision methods, the dispersion estimates are similar to the base model, with a 1 standard deviation in officer effects corresponding to a 35-36% increase in arrest probability. Lastly, in column (8), I report dispersion in officer behavior using the unadjusted first stage officer fixed effects. With no adjustment for precision, this standard deviation estimate is moderately larger than the base model, corresponding to a 40% increase in arrest probability.

4.5 *Officer Demographics*

How does officer arrest propensity relate to officer demographics? A natural next step is to consider how the estimated officer fixed effects, $\hat{\theta}_i$, are associated with officer demographic characteristics. I regress $\hat{\theta}_i$ terms on officer race, gender, age, trainee or sergeant status,

[23] I collapse dispatch code categories to increase the estimation speed of the model and address small cell categories in the majority of the analysis.

experience and experience squared in Table 2.[24] These regressions offer information about whether officers with specific traits systematically differ in their arrest propensities.

Officers with more experience have higher arrest propensities. All else equal, the likelihood of arrest is 3% higher when a responding officer has 10 years of experience instead of 5 years of experience. Officer race does make a difference in the likelihood of arrest; Black and Hispanic officers are 4.6% and 5.3% less likely to make arrests relative to White officers, respectively. The regressions are very similar using officer effects derived from the "Low Availability" and "Unlikely Response" samples.

Overall, demographic factors do not explain a large share of the total variation in officer effects, with regression R^2 statistics of 0.02 - 0.03. Instead, this analysis shows that the substantial variation in arrest behavior observed across officers is due to unobservable characteristics of police officers, such as officer preferences or unobservable dimensions of productivity.

5 Racial Bias Among Officers

5.1 *Testing for Racial Bias*

Is there evidence of racial bias among police officers in Dallas? In this section, I test for the presence of racial bias by investigating relationships between officer race and arrestee race.

I adapt a test of racial bias used in Anwar and Fang (2006) examining officer bias in traffic stops. Anwar and Fang (2006) develop a test that finds evidence of racial bias when the relative ranking of officer arrest rates across officer race changes within different suspect race groups. For example, their test finds evidence of bias if White officers have higher arrest rates than Black officers for Black suspects *and* Black officers have higher arrest rates than White officers for White suspects. In this case, either Black officers or White officers are racially biased (or both). Their model allows officers of different races to behave differently from one another as long as these differences are independent of suspect race. For example,

[24]Salary is omitted from this regression because it is nearly perfectly correlated with experience, given the compensation formulas used by the department. Results are similar when salary is used instead of years of experience.

if White officers are more likely than Black officers to arrest a suspect of any race group, this does necessarily not imply that either group is racially biased. Critically, the test also does not find evidence of racial bias if *all officers* arrest individuals in some suspect race groups more than others. This feature allows officers to *statistically discriminate* against suspects, by using suspect race as a signal of offending characteristics that are correlated with race. Arrest rates can differ across suspect race groups in the test because differences in officer race arrest rates are always measured as a relative ranking within a suspect race group. (See Online Appendix A4 for a discussion of the economic framework in the call for service setting).

A key difference between the call for service setting and the traffic stop setting is that I do not directly observe suspect race for all offenses.[25] Instead, I consider unconditional arrestee race outcomes that take a value of 1 if an individual is arrested and is a particular race, either Black, Hispanic, or White, and 0 otherwise. As before, I estimate $\hat{\theta}_{i,r}$ terms for each unconditional arrestee race outcome, using the model outlined in Section 3. Next, I regress the $\hat{\theta}_{i,r}$ terms on the full set of officer demographic characteristics and examine the impact of officer race on officers' propensity to arrest individuals of different races. This regression framework permits a direct test of whether the rank of arrest rates across officer race changes for different arrestee race outcomes.

The test used in this paper offers three new advantages. First, I am able to test for racial bias among officers in a setting that is not affected by officers electing to initiate interactions. With the exception of West (2015), which studies racial bias of state troopers who are randomly dispatched to motor vehicle accidents, prior work studying racial bias in policing has typically examined interactions between officer and suspect race in officer-initiated incidents, such as traffic stops (e.g. Horrace and Rohlin, 2016; Anbarci and Lee, 2014; Antonovics and Knight, 2009; Anwar and Fang, 2006; Grogger and Ridgeway, 2006).These papers consider suspect race as a given characteristic of a traffic stop; however, in reality, suspect race is also a choice variable of the officer, who chooses to stop a particular individual. In this paper, I am able to test for the presence of racial bias in a setting where calls are initiated

[25] I have limited information on suspects in the data as not all offenses have a identified or recorded suspect. Further, suspect identification can be viewed as an outcome of a police interaction as it may be a function of officer effort.

by complainants and not officers.

Second, I use a regression framework to control for a large array of observable offense characteristics. Anwar and Fang (2006) address the fact that different officers may face different types of traffic stops by re-sampling their data to create comparable samples across officer race. In my setting, I use regression models to measure officer arrest propensity adjusted for observable differences in the composition of offenses across officers.

Lastly, I use officer identifiers to trace the distribution of officer effects by race for each of the arrestee race outcomes. Plotting this distribution reveals the relative importance of behavioral differences within officer race versus across officer race.

The racial bias test used in this paper cannot detect cases when all officers exhibit similar racial bias toward a particular group, a limitation of prior tests as well. As discussed above, Black arrestees are markedly overrepresented in the sample, making up 50% of total offense arrests and only 24% of the Dallas population, adjusted for the proportion of arrests with demographic information (Table 1.A). If arrest rates of Black individuals are higher than other race groups *across all officers*, this pattern could be consistent with statistical discrimination, institutional discrimination or uniform taste-based discrimination that is common across all officers. Institutional racial discrimination will occur when the organizational priorities of the department direct resources toward policing one race group relative to others, and all officers behave similarly given these priorities. At the same time, the higher representation of Black arrestees could also be consistent with uniform police officer attitudes of taste-based racial bias against Black suspects. Recent evidence on the importance of implicit racial bias in decision-making could be consistent with uniform taste-based discrimination against minority groups (e.g. Eberhardt et al., 2004).

5.2 *Baseline Results of the Racial Bias Test*

I limit the sample used in the racial bias test to observations where responding officers have a single race (or respond alone) and each officer has more than 50 observations in this subsample. Tables A4.A and A4.B confirm that offense characteristics are comparable across officer race in this sample. The most meaningful difference is that each officer race is more likely to respond to a complainant that has the same race. This variation in complainant race also means that officers are more likely to respond to an offense with a suspect of their own race, given that social interactions are likely to be racially segregated. The empirical strategy used to conduct the test of racial bias controls directly for complainant race in the first stage to address this difference in exposure.

Table 3 shows the results of the test of racial bias. Each regression estimates the relationship between officer demographics and the measured officer effect, $\hat{\theta}_{i,r}$, calculated from a predictive model of whether the offense resulted in an arrest of an individual of race, r (corresponds to columns). White officers are the omitted officer race category in each regression. The F-tests show whether the officer race coefficients are statistically different from one another.

White officers are more likely to arrest Black individuals than Hispanic or Black officers, by 13.6% and 9.6% respectively. At the same time, White officer arrest patterns are not significantly different from other officer race groups for arrests of individuals who are not Black. This pattern could be viewed as suggestive evidence of bias in the sense that White officers are more likely to arrest Black individuals and "equivalently" likely to arrest individuals of other races.

However, the test does not provide conclusive evidence of racial bias among officers, in the following sense. The F-tests cannot reject the hypothesis that White officers are simply more likely to arrest individuals of all races. This is because the relative ranking of officer race coefficients in the Arrest Hispanic and Arrest White regressions does not reverse or strictly change relative to the pattern in the Arrest Black regression.

I leverage individual officer identifiers to better understand patterns of officer behavior across arrestee race, in Figure 4. These figures show that the distributions of officer fixed effects by officer race are quite close to one another for each arrestee race outcome. Importantly, these figures do not suggest that the rank order of officer effects by officer race changes across these outcomes. Moreover, the pictures suggest that there is more variation in arrest propensity *within* officer race than between officer races. In fact, if officer effects were constant across officer race, the dispersion in officer effects for each of these outcomes would decrease by less than 1%. In other words, race is not the most important determinant of an officer's arrest behavior.

The test cannot reject the null hypothesis that officers are not racially biased. This result could be consistent with reforms adopted by the Dallas Police Department. As discussed above, DPD enacted a number of reform initiatives prior to and during the sample period, including implicit racial bias training, de-escalation training, the use of body-worn cameras, and sharing data on its operations with the public.

5.3 *Robustness of the Racial Bias Test*

The racial bias test used in this paper has the advantage of being unlikely to yield a false positive claim of bias, an attractive feature given that a finding of racial bias can be a political flash point. However, a potential concern about the test used in this paper is that it has low power to detect racial bias. The first reason that the test may have low power is conceptual; the test cannot detect patterns of racial bias that are shared across all officers or patterns of racial bias that are consistent with certain officer race groups arresting all types of suspects more frequently than other officer race groups. In addition to this limitation, the test may have low power because it is a joint hypothesis test.

I empirically assess the power of the racial bias test using a bootstrap simulation. The simulation estimates the rate that the null hypothesis will be rejected when an alternative hypothesis of racial bias is imposed in the data. I do this by altering one officer race coefficient to become larger for different-race arrestee outcomes and smaller for the same-race arrestee outcome by a constant percent change, relative to other officer race groups. This change imposes the alternative hypothesis that the officer race arrest rate ranking differs across arrestee race outcomes.[26]

Figure A4 plots the power of the test for different values of the alternative hypothesis,

[26] For example, consider the simulation with respect to the Black officer coefficient, $\alpha_{Black,r}$. I first regress the second stage $\theta_{i,r}$ outcomes on officer demographics excluding $\alpha_{Black,r}$ and recover a predicted $\hat{\theta}'_{i,r}$ and residual $\hat{r}_{i,r}$. In each bootstrap iteration (of 250 iterations), I draw a wild bootstrap weight $w_b \in \{-1, 1\}$ with equal probability. I then impose an alternative hypothesis on the $\alpha_{Black,r}$ coefficient, and set $\tilde{\theta}^b_{i,r} = \hat{\theta}'_{i,r} + \Delta\alpha_{Black,r} + w_b\hat{r}_{i,r}$. I set Δ to be a constant percent increase for White and Hispanic ("other race") arrestee outcomes and an equivalent percent decrease in the Black ("own race") arrestee outcome. These percent changes are set relative to White or Hispanic officer averages for the total arrest outcome, using the max of the two officer groups (White or Hispanic) for percent increases and the min of the two for the percent decrease. I then regress these simulated values, $\tilde{\theta}^b_{i,r}$, on the full set of officer demographic variables and use F-tests to determine whether the ranking of officer race groups changes across arrestee race outcomes. For each imposed value of Δ, I conduct 250 bootstrap iterations (See Figure A4 for additional details).

or different percent deviations in arrests caused by the altered coefficient. I am able to accept an alternative hypothesis of a 9.5% deviation in arrest outcomes due to changes in the Black officer coefficient and accept a corresponding alternative hypothesis of a 21.5% deviation in the Hispanic coefficient, with 90% power. This power exercise shows that the test is unlikely to detect particularly small differences in racial bias but that it will succeed in rejecting the null hypothesis when there are moderately large patterns of bias.

In Table 3, I show that the racial bias conclusions are robust to considering a simplified version of the test, which does not include officer fixed effects in the first stage of the model but instead inserts officer characteristics as direct controls in the first stage. In this simplified specification, the test similarly cannot reject the null hypothesis that White officers have the highest arrest propensity for all arrestee race outcomes.[27]

6 Conclusion

Individual police officers are critical to the outcomes of police work. While the context of an offense, such as geographic location, time, and incident dispatch type, is the primary determinant of whether the response to the offense will result in an arrest, individual police officer decisions also matter. Analyzing high frequency data on offenses reported through calls for service in Dallas, Texas, I find that the individual officers account for 10-15% of the explainable variation in arrest outcomes.

Police work is characterized by discretion and police officers differ from one another. I find that a 1 standard deviation increase in officer arrest propensity increases the likelihood of an arrest by 32%. In general, observable officer demographic characteristics do not explain a large share of the variation in arrest behavior across officers.

Further, the variation in individual officer arrest propensity does not appear to be driven by racial bias. I apply a test of racial bias using relative ranking of officer arrest rates

[27] I have also benchmarked the findings in this paper by replicating tests used in the prior literature using aggregated arrest statistics for different demographic subgroups in the data, that are not adjusted for observable offense characteristics (e.g. Knowles et al., 2001; Anwar and Fang, 2006; Antonovics and Knight, 2009). In the majority of the paper, I treat suspect identification as an outcome of a police response, but these comparisons condition on suspects directly. These tests find mixed evidence of racial bias (available on request).

by officer race and fail to reject the null hypothesis that officers are not racially biased. I find that White officers are significantly more likely to arrest Black individuals than Black or Hispanic officers; however, I cannot reject the null hypothesis that White officers individuals of any race more than other race officers. Further, I find that most of the differences in officer behavior occur *within* rather than across officer race. The results may be related to a number of progressive police reforms that have been adopted by the Dallas Police Department, including implicit racial bias training, de-escalation training, and the use of body-worn cameras. However, more research is needed to understand how these reform initiatives may alter officer biases and actions.

Having established that individual officers are critical to the outcomes of police and civilian interactions, questions remain for future research. First, what are the marginal welfare costs (or benefits) of arrests that result from police discretion? An arrest may be a positive or negative welfare outcome depending on the incident context, culpability of the arrestee, severity of the offense and the subsequent burden for the arrestee and his/or family. A greater understanding of the quality of marginal arrests could improve response protocols of police departments.

Lastly, what are the best policy levers to reduce dispersion in arrest behavior? From a fairness perspective, investments in reducing dispersion in officer behavior could yield benefits in the form of increased trust in law enforcement and equal access to police protection services. Future work should assess the costs and benefits of different law enforcement practices that may be used to increase uniformity in officer behavior, including additional police training, monitoring procedures, mentorship programs, and targeted hiring and firing of officers.

References

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics 25*(1), 95–135.

Anbarci, N. and J. Lee (2014). Detecting racial bias in speed discounting: Evidence from speeding tickets in Boston. *International Review of Law and Economics 38*, 11–24.

Antonovics, K. and B. G. Knight (2009). A new look at racial profiling: Evidence from the Boston Police Department. *The Review of Economics and Statistics 91*(1), 163–177.

Anwar, S. and H. Fang (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review 96*(1), 127–151.

BlueLivesMatter (2016, July). Opinion: Why do cops always support other officers? https://bluelivesmatter.blue/.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review (Forthcoming)*.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Census (2015). American community survey: Demographic and housing estimates. Technical report, Census Bureau, United States.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Correia, S. (2016). A feasible estimator for linear models with multi-way fixed effects. *Working Paper*.

Department of Justice (2017, March). Supporting Federal, State, Local and Tribal Law Enforcement. Memorandum, Office of the Attorney General, Department of Justice, United States.

Eberhardt, J., P. Goff, V. Purdie, and P. Davies (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology 87*(6), 876–893.

FBI (2015). Crime in the United States: Offenses Known to Law Enforcement. Technical report, Federal Bureau of Investigation, Department of Justice, United States.

Frydl, K. and W. Skogan (2004). *Fairness and Effectiveness in Policing: The Evidence*. National Academies Press.

Fryer, R. (2016). An empirical analysis of racial differences in police use of force. *National Bureau of Economic Research* (22399).

Gelman, A., J. Fagan, and A. Kiss (2007). An analysis of the New York City police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association 102*(479), 813–823.

Goncalves, F. and S. Mello (2017). Racial bias in policing: Evidence from speeding tickets. *Working Paper*.

Grogger, J. and G. Ridgeway (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association 101*(475), 878–887.

Guarino, C., M. Maxfield, M. Reckase, P. Thompson, and J. Wooldridge (2015). An evaluation of Empirical Bayes' estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics 40*(2), 190–222.

Haugh, C. (2016, July 9). How the Dallas Police Department Reformed Itself. *The Atlantic*.

Horrace, W. C. and S. M. Rohlin (2016). How dark is dark? bright lights, big city, racial profiling. *The Review of Economics and Statistics 98*(2), 226–232.

Joh, E. (2016). Surveillance discretion: Automated suspicion, big data, and policing. *Harvard Law and Policy Review 10*, 15–42.

Jones, J. (2015, June). In U.S., confidence in police lowest in 22 years. Technical report, Gallup.

Kane, T. and D. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *National Bureau of Economic Research* (14607).

Knowles, J., N. Persico, and P. Todd (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy 109*(1), 203–229.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015). Value-added modeling: A review. *Economics of Education Review 47*, 180–195.

La Vigne, N., J. Fontaine, and A. Dwivedi (2017, February). How do people in high-crime, low-income communities view the police? Technical report, Urban Institute.

Mas, A. and E. Moretti (2009). Peers at work. *The American Economic Review 99*(1), 112–145.

Mastrofski, S. (2004). Controlling street-level police discretion. *The Annals of the American Academy of Political and Social Science 593*, 100–118.

Morris, C. (1983). Parametric Empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association 78*(381), 47–55.

Nickels, E. (2007). A note on the status of discretion in police research. *Journal of Criminal Justice 35*, 570–578.

Schneider, B. (2015, January). Do Americans trust their cops to be fair and just? New poll contains surprises. *Reuters*.

Silver, D. (2016). Haste or waste? Peer pressure and the distribution of marginal returns to health care. *Working Paper*.

Tsiaperas, T. (2016, June). Top Dallas cops take classes to learn how to fight bias. *The Dallas Morning News*.

Tyler, T., P. Goff, and R. MacCoun (2015). The impact of psychological science on policing in the united states: Procedural justice, legitimacy, and effective law enforcement. *Psychological Science in the Public Interest 16*(3), 75–109.

West, J. (2015). Racial bias in police investigations. *Working Paper*.

Tables and Figures

Table 1: Summary Statistics

Table 1.A: Summary Statistics: Outcomes, Officers, and Complainants

	Total Sample		*Analysis Sample*	
	Mean	S.D.	Mean	S.D.
Total Observations	249,589		322,280	
Total Incidents	249,589		234,336	
Total Officers	3,158		1,858	
Outcomes				
Arrest	0.08	(0.27)	0.09	(0.29)
Arrestee Black	0.02	(0.16)	0.03	(0.17)
Arrestee Hispanic	0.01	(0.09)	0.01	(0.10)
Arrestee White	0.01	(0.10)	0.01	(0.11)
Arrest has Race Demographics	0.04	(0.20)	0.05	(0.22)
Officer Characteristics				
Officer Arrest Rate	0.09	(0.06)	0.09	(0.05)
Two Responders	0.38	(0.48)	0.54	(0.50)
Total Incidents	230.02	(149.06)	241.52	(142.28)
Trainee	0.09	(0.29)	0.08	(0.28)
Sergeant	0.01	(0.12)	0.01	(0.09)
Salary ($10,000s)	5.84	(1.14)	5.83	(1.12)
Years of Experience	12.06	(9.28)	11.91	(9.14)
Age	38.09	(10.25)	37.96	(10.17)
Female	0.17	(0.37)	0.16	(0.36)
Black	0.25	(0.43)	0.24	(0.43)
Hispanic	0.22	(0.41)	0.22	(0.41)
White	0.49	(0.50)	0.50	(0.50)
Officer has Race Demographics	0.98	(0.13)	0.98	(0.14)
Complainant Characteristics				
Victim with Injury	0.10	(0.30)	0.11	(0.31)
Number of Complainants	1.07	(0.33)	1.08	(0.33)
Female	0.46	(0.50)	0.47	(0.50)
Black	0.34	(0.47)	0.35	(0.48)
Hispanic	0.28	(0.45)	0.28	(0.45)
White	0.34	(0.47)	0.33	(0.47)
Complainant has Race Demographics	0.80	(0.40)	0.79	(0.41)

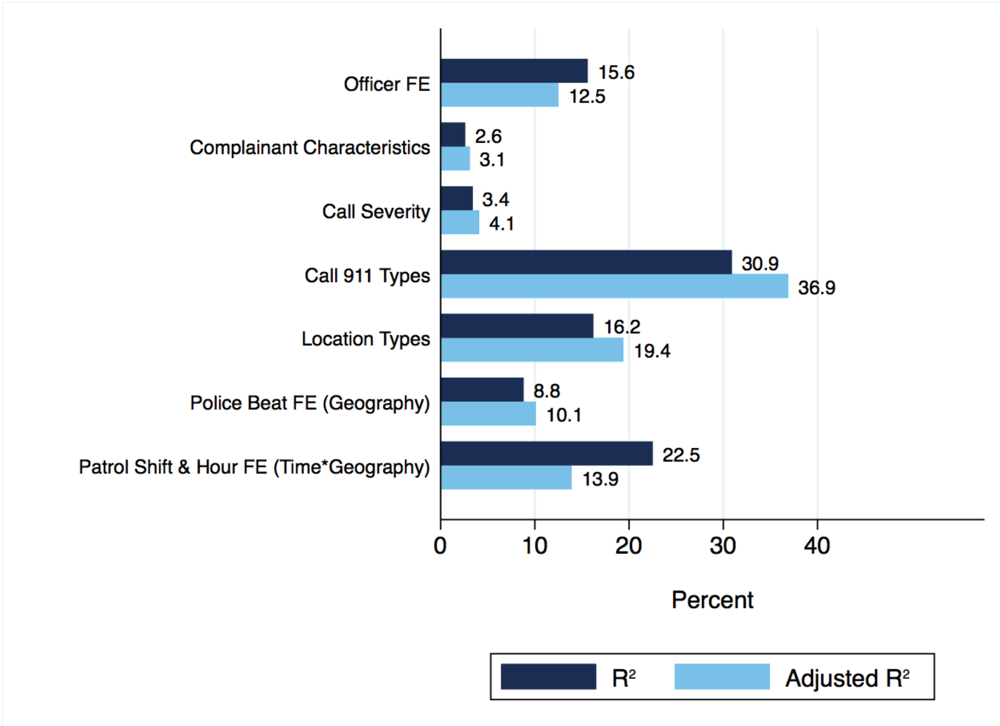
These tables display summary statistics for covariates used in analysis. The first column, "Total Sample", consists of all offenses reported through calls for service in the data, with each offense incident having only one record. The second column, "Analysis Sample", summarizes the primary analysis sample and excludes records for police officers that respond to fewer than 50 offenses. Additionally, this sample duplicates offense responses with two responding officers so that the records for each responding officer may contribute to the estimation of Officer Fixed Effects.

Table 1.B: Summary Statistics: Incident Urgency, Location Type, and Dispatch Code

	Total Sample		***Analysis Sample***	
	Mean	S.D.	Mean	S.D.
Total Observations	249,589		322,280	
Total Incidents	249,589		234,336	
Total Officers	3,158		1,858	
Call Urgency				
Time to Dispatch (Minutes)	25.13	(28.29)	24.45	(27.94)
Location Type				
Apartment	0.14	(0.35)	0.14	(0.35)
Residence Other	0.16	(0.37)	0.17	(0.37)
Bar/Club/Entertainment	0.03	(0.18)	0.03	(0.18)
Retail	0.07	(0.26)	0.08	(0.27)
Business Other	0.05	(0.22)	0.05	(0.22)
Govt/Health/School/Religion	0.01	(0.10)	0.01	(0.10)
Motor Vehicle	0.02	(0.14)	0.02	(0.14)
Parking Lot	0.22	(0.42)	0.21	(0.41)
Street	0.16	(0.37)	0.16	(0.37)
Outdoor Other	0.05	(0.22)	0.05	(0.22)
Other Location	0.07	(0.26)	0.07	(0.26)
Dispatch Code Type				
Criminal Assault, High Priority	0.01	(0.12)	0.01	(0.12)
Armed Encounter/Active Shooter, High Priority	0.02	(0.13)	0.02	(0.14)
Robbery, High Priority	0.06	(0.24)	0.07	(0.25)
Burglary of Business, High Priority	0.01	(0.09)	0.01	(0.09)
Burglary of Business, Low Priority	0.05	(0.21)	0.05	(0.21)
Burglary of Residence, High Priority	0.03	(0.16)	0.03	(0.17)
Burglary of Residence, Low Priority	0.09	(0.29)	0.09	(0.28)
Burglary of Vehicle, High Priority	0.02	(0.15)	0.02	(0.15)
Burglary of Vehicle, Low Priority	0.16	(0.37)	0.14	(0.35)
Unauthorized Use of Vehicle, High Priority	0.01	(0.09)	0.01	(0.09)
Unauthorized Use of Vehicle, Low Priority	0.04	(0.20)	0.04	(0.19)
Theft, High Priority	0.02	(0.15)	0.02	(0.15)
Theft, Low Priority	0.06	(0.23)	0.05	(0.22)
Criminal Mischief, High Priority	0.01	(0.09)	0.01	(0.09)
Criminal Mischief, Low Priority	0.07	(0.25)	0.06	(0.24)
Major Disturbance, High Priority	0.12	(0.33)	0.14	(0.35)
Accident, High Priority	0.03	(0.18)	0.03	(0.18)
Accident, Low Priority	0.06	(0.25)	0.06	(0.24)
Injured Person, High Priority	0.00	(0.06)	0.00	(0.06)
Injured Person, Low Priority	0.01	(0.11)	0.01	(0.11)
Other, High Priority	0.05	(0.23)	0.06	(0.24)
Other, Low Priority	0.06	(0.24)	0.06	(0.24)

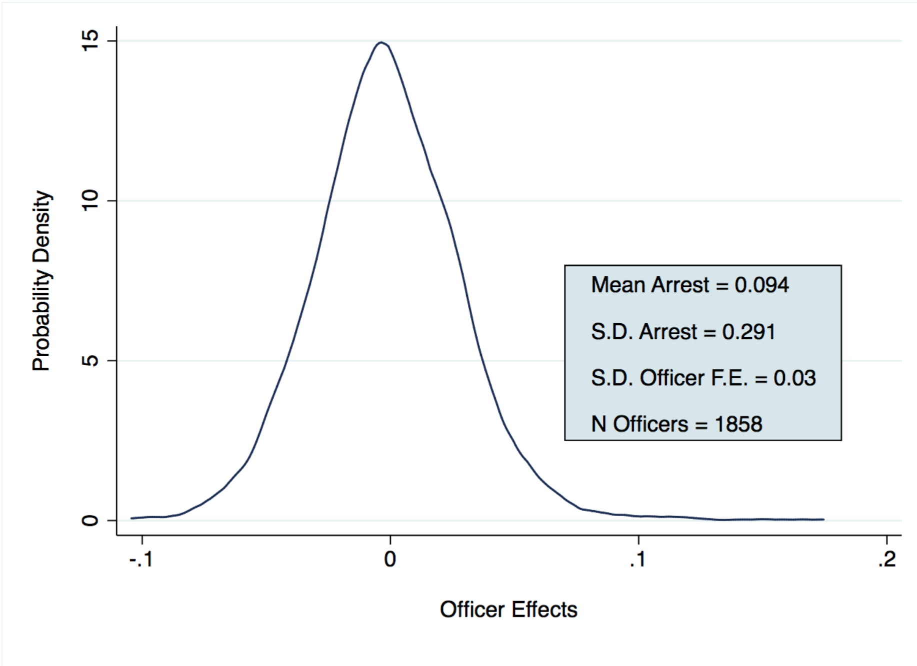
These tables display summary statistics for covariates used in analysis. The first column, "Total Sample", consists of all offenses reported through calls for service in the data, with each offense incident having only one record. The second column, "Analysis Sample", summarizes the primary analysis sample and excludes records for police officers that respond to fewer than 50 offenses. Additionally, this sample duplicates offense responses with two responding officers so that the records for each responding officer may contribute to the estimation of Officer Fixed Effects. Call Urgency or Time to Dispatch (Minutes) is a variable that captures the priority level or severity of a given call at the time of dispatch.

Figure 1: Relative Proportion of Total R^2, Components of Model



This figure shows the relative importance of explanatory variables in the estimation model. Each bar graph grouping represents the percent of total model R^2 (Adjusted R^2) accounted for by a grouping of variables. I estimate these percentages sequentially from the bottom bar to the top, with each percentage calculated as an additional contribution of R^2 to the total or: $(R^2_{currentbar} - R^2_{priorbar})/R^2_{total}$. Patrol Shift and Hour FE are monthly aggregated patrol shift indicators (day of the week*8 hour shift*month) and indicators for the hour within an 8 hour shift. Police Beat FE are fixed effects for geographic police beat locations. Location types are indicator variables for the type of location where the call occurred (e.g. apartment or street). Call 911 types are dispatch code categories associated with the call. Call Severity includes the number of minutes (and number of minutes squared) that pass between when a call is placed and when it is dispatched, as well as an indicator for whether the response included two officers. Complainant Characteristics include the number of complainants, whether there is demographic information for these complainants, complainant race and gender (max values across complainants), and whether there is a victim injury. Lastly, Officer FE include fixed effects for the first and second police officer responders, θ_i and θ_j. The explanatory power of Officer FE is therefore the relative contribution of these variables after controlling for all other variables in the model.

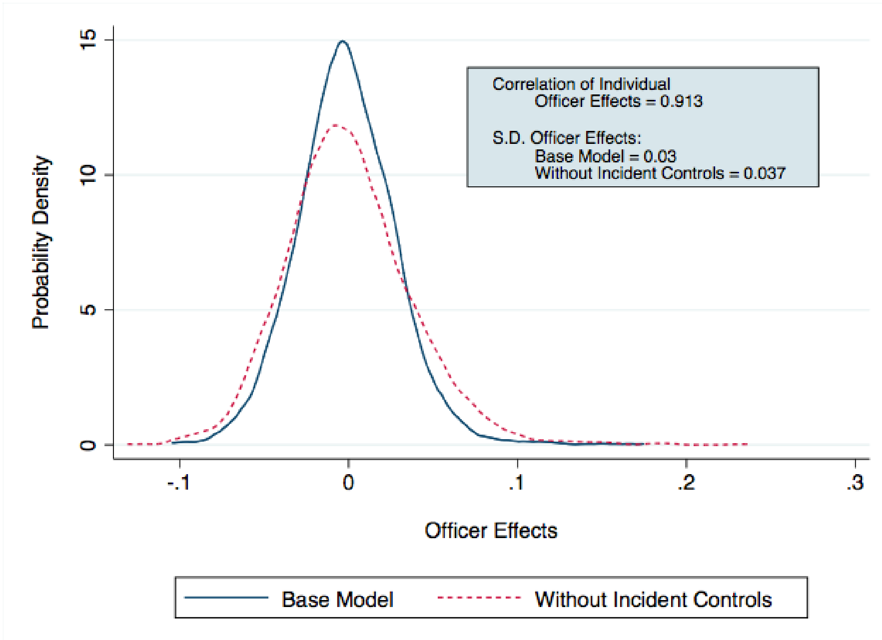
Figure 2: Distribution of Officer Effects, $\hat{\theta}_i$



This figure graphs the distribution of the Officer Effects, $\hat{\theta}_i$, measured in the primary arrest outcome model on the analysis sample. Each officer in the sample has at least 50 offense responses.

Figure 3: Tests of the Importance of Officer Sorting to Officer Effect Distribution

Figure 3.A: Officer Effects Measured with and without Offense Controls



This graph compares the base model officer effects, $\hat{\theta}_i$, to officer effects, $\tilde{\hat{\theta}}_i$, that are estimated from a model that does not include offense characteristics and police beat fixed effects, X_{kt} and ϕ_g. These offense characteristics are those that an officer may choose at the level of a call response, as shifts and partners are determined prior to a call event.

Figure 3.B: Officer Effects in Full Sample and Low Availability Sample

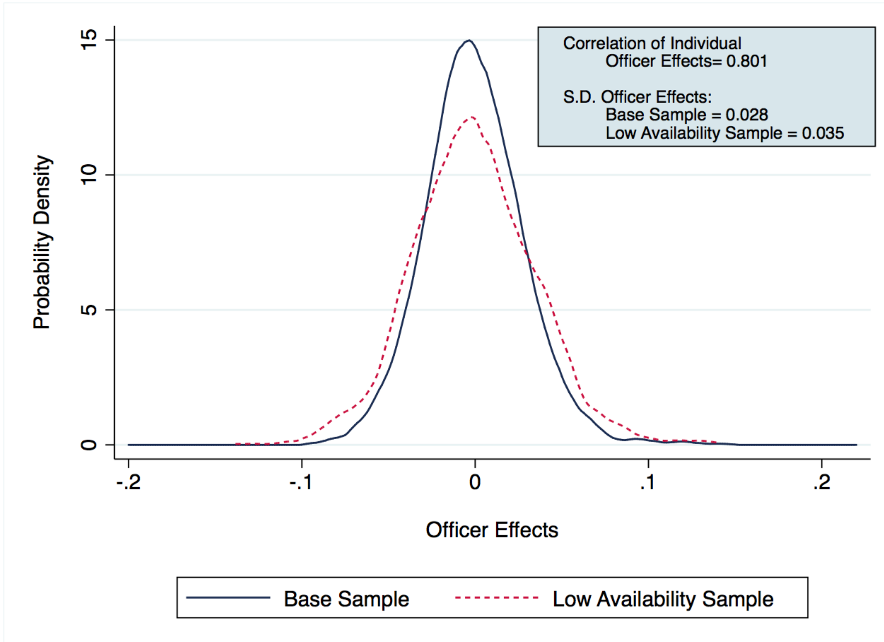
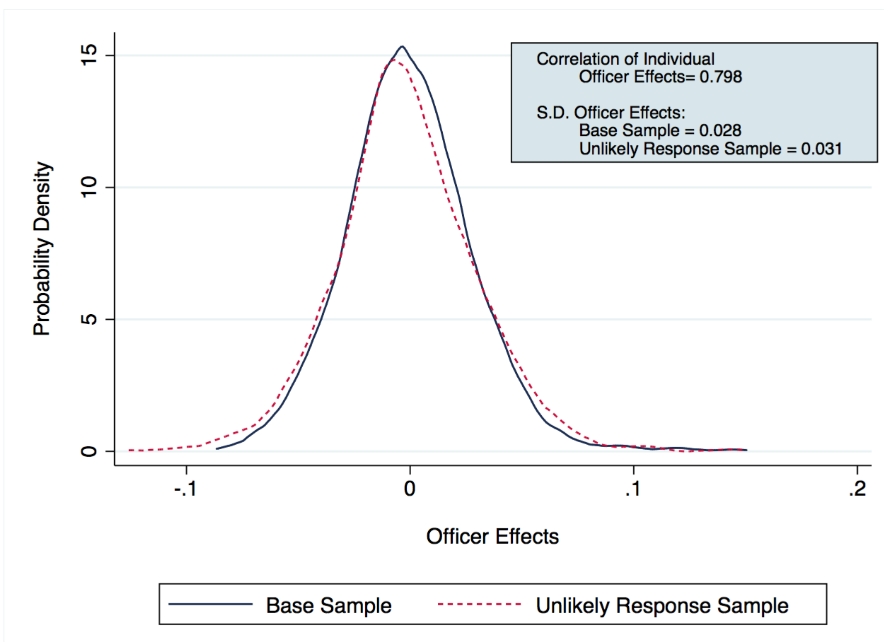


Figure 3.C: Officer Effects in Full Sample and Unlikely Response Sample



The “Low Availability” sample is determined by taking the set of observations where more officers are unavailable because they are responding to other offenses at the time an offense is dispatched, split at the median within patrol shift cells. I determine the “Unlikely Response” subsample by first estimating a linear probability model for each officer i that predicts whether i responded to each offense, conditional on the full set of covariates in the model. The “Unlikely Response” uses a sub-sample of observations where officers have a lower predicted probability of responding to calls, restricting the sample to observations with below median predicted response likelihood among actual responses for each officer. The analysis is restricted to officers with at least 50 observations in the sub-sample. The corresponding base sample benchmark is estimated over the full set of responses for the same officer group.

Table 2: Officer Effects and Officer Demographics

Outcome: Officer Effect	(1) Full Sample	(2) Low Availability Responses	(3) Unlikely Responses
Black	-0.0043**	-0.0099***	-0.0054**
	(0.0017)	(0.0023)	(0.002)
Hispanic	-0.005**	-0.0083***	-0.0062**
	(0.0018)	(0.0024)	(0.0022)
Other Race	-0.0022	-0.0067	-0.0027
	(0.0031)	(0.0043)	(0.0044)
Female	-0.0028	-0.0054*	-0.0032
	(0.0017)	(0.0024)	(0.0021)
Age	-0.00003	-0.0002	-0.0001
	(0.0001)	(0.0002)	(0.0002)
Trainee	-0.0034	-0.0023	-0.0075*
	(0.0025)	(0.0035)	(0.0035)
Sergeant	-0.005	-0.0002	0.0032
	(0.0055)	(0.0112)	(0.0099)
Experience	0.001**	0.0009+	0.0013**
	(0.0004)	(0.0005)	(0.0004)
Experience^2	-0.00003**	-0.00002	-0.00003**
	(0.00001)	(0.00001)	(0.00001)
Observations	1,800	1,352	1,328
R-squared	0.019	0.025	0.032
Fixed Effect Mean	-0.001	-0.001	-0.002
Fixed Effect S.D.	0.029	0.035	0.031
Outcome Mean	0.094	0.093	0.080
Outcome S.D.	0.292	0.290	0.271

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.1$

This table shows regression results of officer effects measured using the arrest outcome, $\hat{\theta}_i$, regressed on fixed officer characteristics, at the officer level. Robust standard errors are in parentheses. The analysis in columns (2) and (3) is restricted to officers with at least 50 observations in each sub-sample. Other race officers are the omitted race category. Officers without demographic information are excluded from the regressions.

Table 3: Racial Bias Test, Officer and Arrestee Race

		(1)	(2)	(3)	(4)
		Arrest	Arrest Black	Arrest Hispanic	Arrest White
A. Officer Level, Outcome=Officer Effects					
Black Officer		-0.0062**	-0.0030**	0.0002	0.0001
		(0.0021)	(0.0011)	(0.0006)	(0.0006)
Hispanic Officer		-0.0039+	-0.0022*	0.0006	0.0001
		(0.0022)	(0.0011)	(0.0008)	(0.0007)
Black=Hispanic:	F-Test	0.932	0.398	0.240	0.001
	P-Value	0.334	0.528	0.624	0.969
Black=White:	F-Test	8.45	7.49	0.107	0.046
	P-Value	0.004	0.006	0.743	0.830
Hispanic=White:	F-Test	3.02	3.99	0.563	0.020
	P-Value	0.083	0.046	0.453	0.887
Observations		1,303	1,303	1,303	1,303
Arrest Mean		0.078	0.025	0.009	0.009
B. Event Level, Outcome=First Stage Arrest					
Black Officer		-0.0044*	-0.0014	0.0001	-0.0002
		(0.0019)	(0.0012)	(0.0008)	(0.0007)
Hispanic Officer		-0.0042*	-0.0021+	0.0003	-0.0002
		(0.0021)	(0.0013)	(0.0008)	(0.0008)
Black=Hispanic:	F-Test	0.006	0.263	0.055	0.000
	P-Value	0.938	0.608	0.815	0.989
Black=White:	F-Test	5.16	1.38	0.018	0.058
	P-Value	0.023	0.239	0.895	0.809
Hispanic=White:	F-Test	3.91	2.72	0.147	0.056
	P-Value	0.048	0.099	0.701	0.813
Observations		165,845	165,845	165,845	165,845
Arrest Mean		0.077	0.025	0.009	0.009

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.1$
This table shows regressions of officer specific effects on officer demographics, where officer effects are derived from four different first stage outcomes, general arrests, and whether the arrestee was Black, White or Hispanic. Each arrestee race outcome is defined unconditionally as 1 if an individual of that race was arrested and 0 otherwise. Robust standard errors are in parentheses. The full sample, Panel (A), is restricted to observations where responding officers have a single race, and each officer has more than 50 observations within this restriction. Panels (B) and (C) represent the overlap of Panel (A) with the "Low Availability" and "Unlikely Response" samples, where the sample is restricted to officers with at least 50 observations within each subset. The interaction of arrestee outcome race and officer race through the regression coefficients represents a test of officer and arrestee race interaction effects. F-Tests measure whether officers of different races are more likely to make arrests of individuals of different races.

The bottom panel conducts a similar test directly in the first stage of analysis. Here, the regressions omit officer fixed effects, θ_i, and measure the direct effect of officer demographic characteristics in the first stage regression. Each regression controls for demographic characteristics of i and includes fixed effects for co-responders, θ_j. Standard errors are clustered at the level of the focal officer, i, and the shift cell, δ_{gt}, to account for error correlations within officers and shifts. Each arrestee race outcome is defined unconditionally as 1 if an individual of that race was arrested and 0 otherwise. Standard errors are clustered at the focal officer level, θ_i, and the shift cell, δ_{gt}.

Figure 4: Distribution of Officer Effects across Officer Race

Figure 4.A: Officer Effects by Officer Race, Arrest Outcome

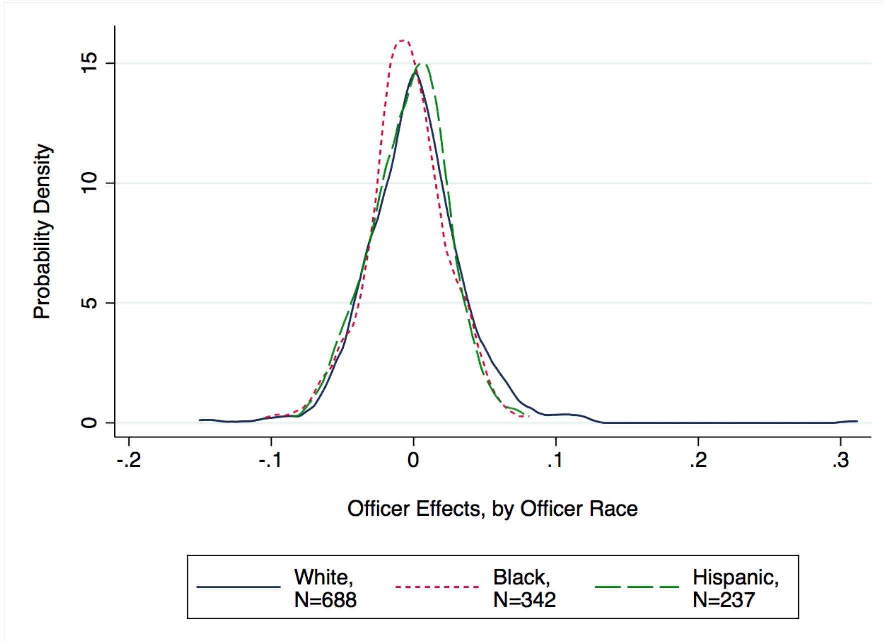
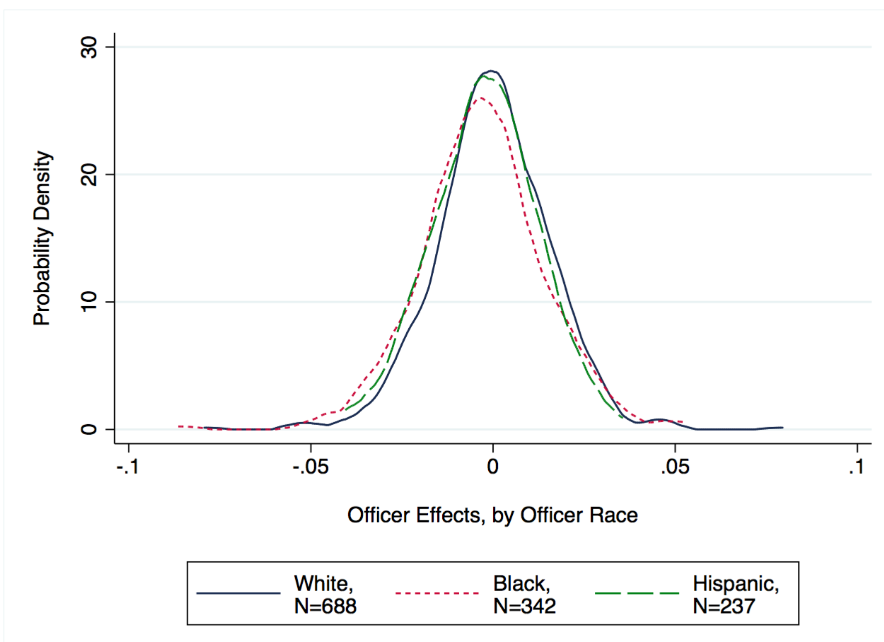


Figure 4.B: Officer Effects by Officer Race, Arrestee Black Outcome



These graphs the distribution of officer effects by officer race for different arrest demographic outcomes. The officer effects correspond to the outcomes used in Table 3. These graphs are shown for the full racial bias test sample, which restricts to observations where the responding officers have a single race and each officer within this restriction has at least 50 observations. Each arrestee race outcome is defined unconditionally as 1 if an individual of that race was arrested and 0 otherwise. Each graph shows the density overlap of officer effects for the different arrest outcomes.

Figure 4.C: Officer Effects by Officer Race, Arrestee Hispanic Outcome

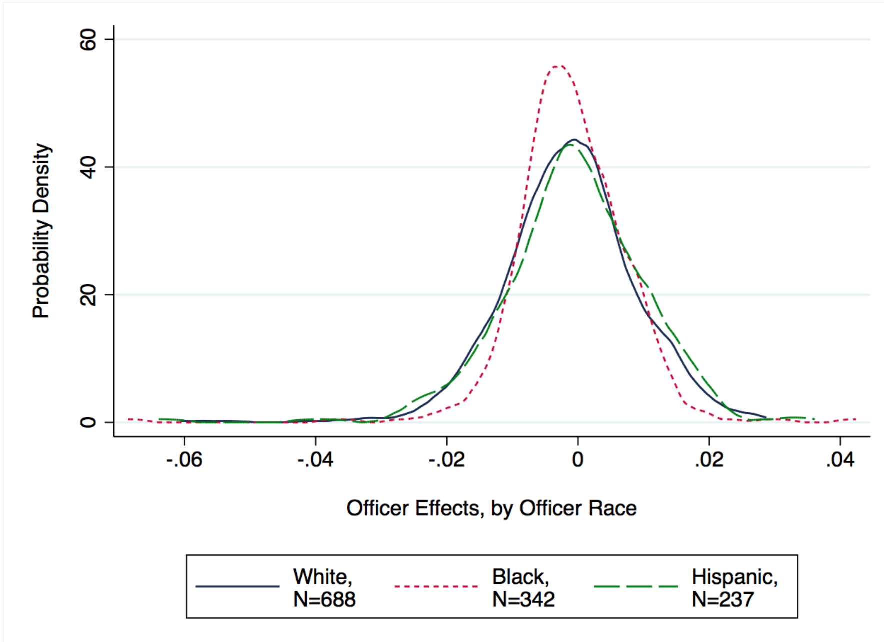
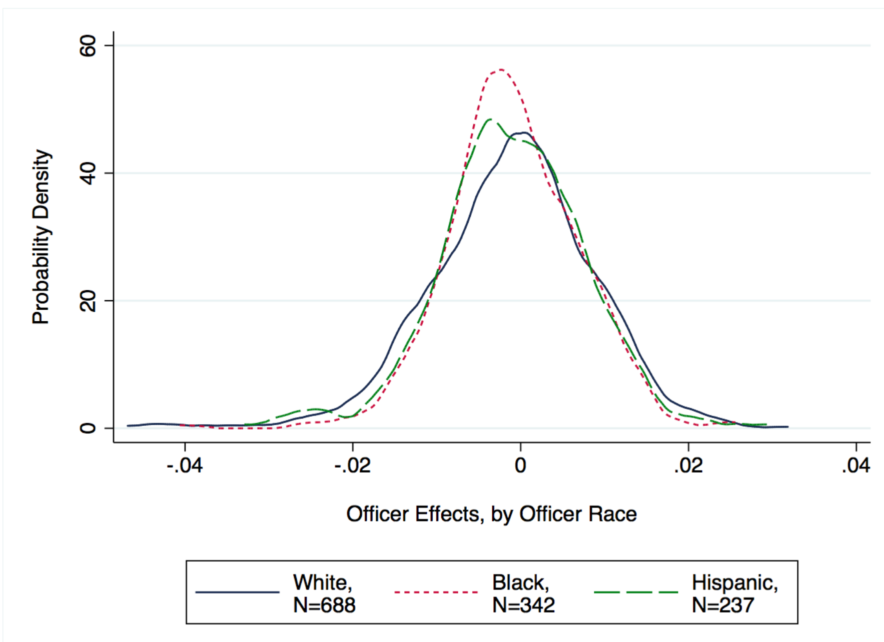


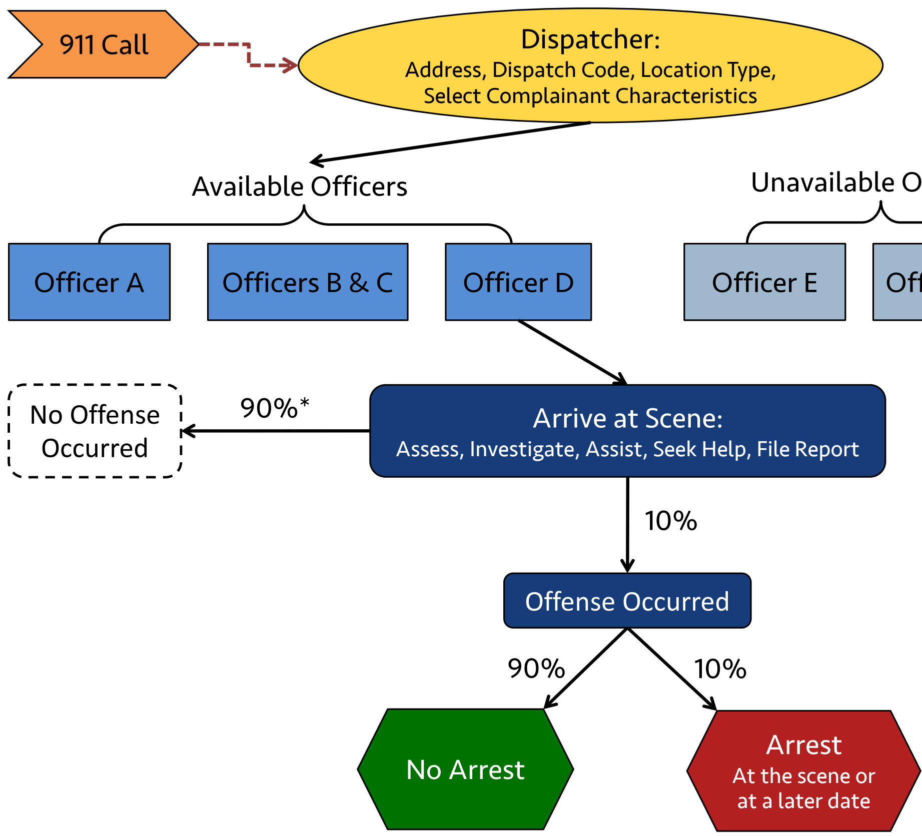
Figure 4.D: Officer Effects by Officer Race, Arrestee White Outcome



These graphs the distribution of officer effects by officer race for different arrest demographic outcomes. The officer effects correspond to the outcomes used in Table 3. These graphs are shown for the full racial bias test sample, which restricts to observations where the responding officers have a single race and each officer within this restriction has at least 50 observations. Each arrestee race outcome is defined unconditionally as 1 if an individual of that race was arrested and 0 otherwise. Each graph shows the density overlap of officer effects for the different arrest outcomes.

A1 Appendix Tables and Figures

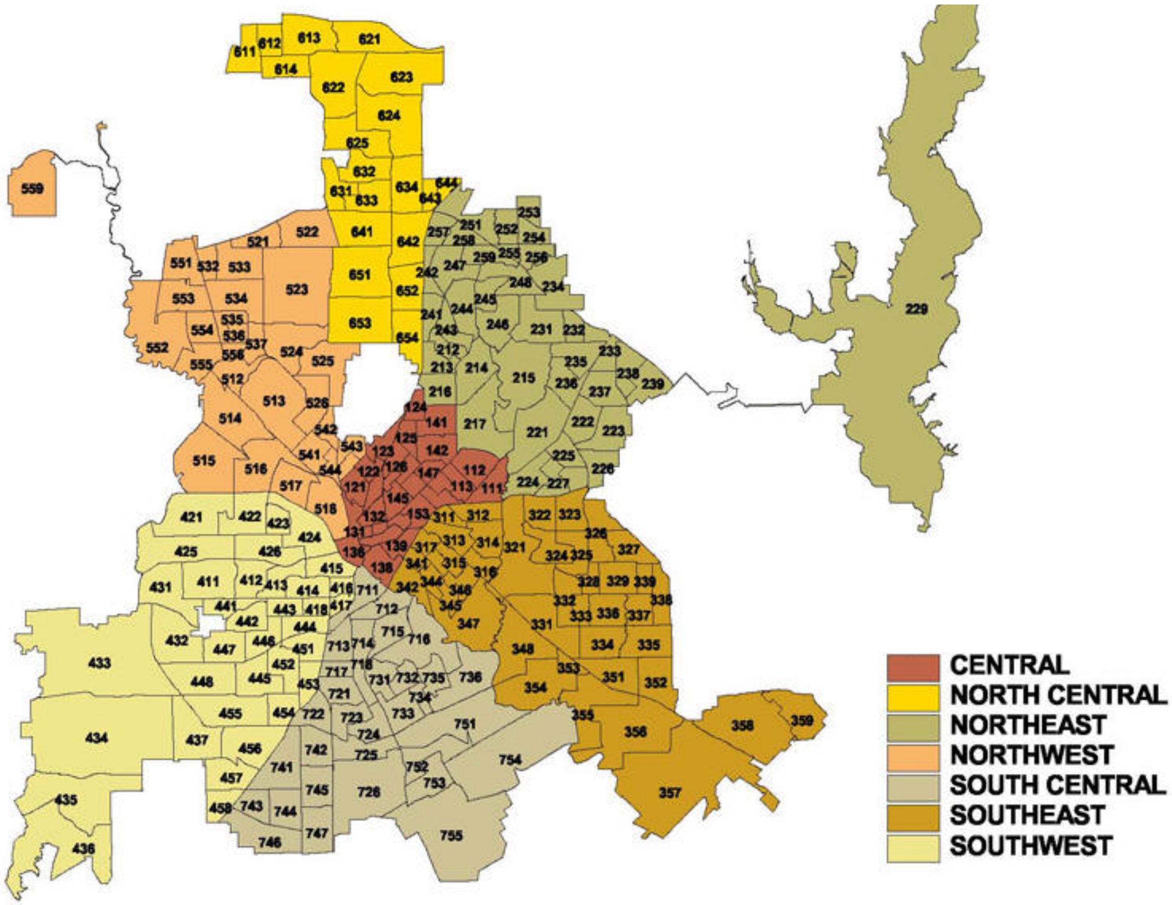
Figure A1: Steps involved in a Response to an Offense Reported through a 9-1-1 Call



This figure displays an outline of an incident response path at the Dallas Police Department. Information on call response protocols was obtained through conversations with officers and dispatchers at the department.

*The number of offenses in the analysis sample represents 10% of the total dispatch calls recorded. The dispatch response sample may be inflated in the sense that it includes some officer-initiated incidents or officer "mark-outs" to dispatchers, responses to assist officers on existing responses and responses to multiple calls that are later determined to be associated with a single incident. Offenses without a match in the dispatch data are excluded from the analysis (See Data Appendix A5 for additional details on data cleaning and sample construction).

Figure A2: Police Beats and Police Divisions in Dallas, TX



This figure shows a map of the 234 police beats contained in the 7 police divisions in Dallas. Police sectors are geographic units that are collections of beats within police divisions (35 total sectors). Map obtained from the North Dallas Neighborhood Alliance: `http://www.ndna-tx.org/crimeWatch/dallasPolice/DivMap.aspx`.

Table A1: Summary Statistics, Officer Sorting Robustness Samples

Table A1.A: Summary Statistics: Outcomes, Officers, and Complainants

	Analysis Sample		*Low Availability Sample*		*Unlikely Response Sample*	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total Observations	322,280		142,488		141,297	
Total Incidents	234,336		102,749		118,901	
Total Officers	1,858		1,353		1,331	
Outcomes						
Arrest	0.09	(0.29)	0.09	(0.29)	0.08	(0.27)
Arrestee Black	0.03	(0.17)	0.03	(0.16)	0.02	(0.15)
Arrestee Hispanic	0.01	(0.10)	0.01	(0.10)	0.01	(0.09)
Arrestee White	0.01	(0.11)	0.01	(0.11)	0.01	(0.10)
Arrests with Race Demographics	0.05	(0.22)	0.05	(0.21)	0.04	(0.20)
Officer Characteristics						
Officer Arrest Rate	0.09	(0.05)	0.09	(0.05)	0.09	(0.05)
Two Responders	0.54	(0.50)	0.55	(0.50)	0.32	(0.47)
Total Incidents	241.52	(142.28)	259.29	(136.23)	264.25	(136.97)
Trainee	0.08	(0.28)	0.09	(0.28)	0.07	(0.25)
Sergeant	0.01	(0.09)	0.00	(0.07)	0.00	(0.07)
Salary ($10,000s)	5.83	(1.12)	5.79	(1.11)	5.84	(1.12)
Years of Experience	11.91	(9.14)	11.61	(8.99)	12.03	(9.16)
Age	37.96	(10.17)	37.71	(10.12)	38.16	(10.19)
Female	0.16	(0.36)	0.15	(0.36)	0.15	(0.36)
Black	0.24	(0.43)	0.24	(0.42)	0.24	(0.43)
Hispanic	0.22	(0.41)	0.22	(0.41)	0.22	(0.41)
White	0.50	(0.50)	0.50	(0.50)	0.50	(0.50)
Officer has Race Demographics	0.98	(0.14)	0.99	(0.10)	1.00	(0.06)
Complainant Characteristics						
Victim with Injury	0.11	(0.31)	0.12	(0.33)	0.10	(0.30)
Number of Complainants	1.08	(0.33)	1.08	(0.34)	1.07	(0.33)
Female	0.47	(0.50)	0.47	(0.50)	0.47	(0.50)
Black	0.35	(0.48)	0.35	(0.48)	0.34	(0.47)
Hispanic	0.28	(0.45)	0.28	(0.45)	0.29	(0.45)
White	0.33	(0.47)	0.33	(0.47)	0.33	(0.47)
Complainants with Race Demographics	0.79	(0.41)	0.80	(0.40)	0.80	(0.40)

These tables display summary statistics for covariates used in analysis. The first column, "Analysis Sample", summarizes the primary analysis sample. The second column, "Low Availability Sample," consists of observations where fewer officers are available to respond to at the time of the incident. The third column, "Unlikely Response Sample," consists of observations where the predicted probability that an officer responds to the incident is low, given other observables in the model. More details on the construction of these samples can be found in Section 4.3. Officer arrest rate and number of incidents is calculated over all observations in the raw data. Robustness samples are restricted to officers with at least 50 observations within the relevant sub-sample.

Table A1.B: Summary Statistics: Incident Urgency, Location Type, and Dispatch Code

	Analysis Sample		***Low Availability Sample***		***Unlikely Response Sample***	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total Observations	322,280		142,488		141,297	
Total Incidents	234,336		102,749		118,901	
Total Officers	1,858		1,353		1,331	
Call Urgency						
Time to Dispatch (Minutes)	24.45	(27.94)	24.80	(27.79)	25.75	(28.62)
Location Type						
Apartment	0.14	(0.35)	0.15	(0.35)	0.14	(0.35)
Residence Other	0.17	(0.37)	0.16	(0.37)	0.17	(0.37)
Bar/Club/Entertainment	0.03	(0.18)	0.03	(0.18)	0.03	(0.18)
Retail	0.08	(0.27)	0.08	(0.27)	0.07	(0.26)
Business Other	0.05	(0.22)	0.05	(0.22)	0.05	(0.22)
Govt/Health/School/Religion	0.01	(0.10)	0.01	(0.09)	0.01	(0.10)
Motor Vehicle	0.02	(0.14)	0.02	(0.13)	0.02	(0.13)
Parking Lot	0.21	(0.41)	0.21	(0.41)	0.23	(0.42)
Street	0.16	(0.37)	0.17	(0.37)	0.17	(0.37)
Outdoor Other	0.05	(0.22)	0.05	(0.22)	0.05	(0.22)
Other Location	0.07	(0.26)	0.07	(0.26)	0.07	(0.25)
Dispatch Code Type						
Criminal Assault, High Priority	0.01	(0.12)	0.02	(0.13)	0.01	(0.11)
Armed Encounter/Active Shooter, High Priority	0.02	(0.14)	0.02	(0.15)	0.02	(0.13)
Robbery, High Priority	0.07	(0.25)	0.07	(0.26)	0.06	(0.24)
Burglary of Business, High Priority	0.01	(0.09)	0.01	(0.09)	0.01	(0.08)
Burglary of Business, Low Priority	0.05	(0.21)	0.04	(0.20)	0.05	(0.21)
Burglary of Residence, High Priority	0.03	(0.17)	0.03	(0.17)	0.03	(0.16)
Burglary of Residence, Low Priority	0.09	(0.28)	0.09	(0.28)	0.10	(0.30)
Burglary of Vehicle, High Priority	0.02	(0.15)	0.02	(0.15)	0.02	(0.14)
Burglary of Vehicle, Low Priority	0.14	(0.35)	0.13	(0.33)	0.16	(0.36)
Unauthorized Use of Vehicle, High Priority	0.01	(0.09)	0.01	(0.09)	0.01	(0.09)
Unauthorized Use of Vehicle, Low Priority	0.04	(0.19)	0.04	(0.19)	0.04	(0.20)
Theft, High Priority	0.02	(0.15)	0.02	(0.15)	0.02	(0.15)
Theft, Low Priority	0.05	(0.22)	0.05	(0.21)	0.06	(0.23)
Criminal Mischief, High Priority	0.01	(0.09)	0.01	(0.10)	0.01	(0.09)
Criminal Mischief, Low Priority	0.06	(0.24)	0.06	(0.23)	0.07	(0.25)
Major Disturbance, High Priority	0.14	(0.35)	0.15	(0.36)	0.12	(0.33)
Accident, High Priority	0.03	(0.18)	0.04	(0.19)	0.03	(0.18)
Accident, Low Priority	0.06	(0.24)	0.06	(0.24)	0.07	(0.25)
Injured Person, High Priority	0.00	(0.06)	0.00	(0.06)	0.00	(0.06)
Injured Person, Low Priority	0.01	(0.11)	0.01	(0.11)	0.01	(0.11)
Other, High Priority	0.06	(0.24)	0.06	(0.24)	0.05	(0.22)
Other, Low Priority	0.06	(0.24)	0.06	(0.24)	0.06	(0.24)

These tables display summary statistics for covariates used in analysis. The first column, "Analysis Sample", summarizes the primary analysis sample. The second column, "Low Availability Sample," consists of observations where fewer officers are available to respond to at the time of the incident. The third column, "Unlikely Response Sample," consists of observations where the predicted probability that an officer responds to the incident is low, given other observables in the model. More details on the construction of these samples can be found in Section 4.3. Robustness samples are restricted to officers with at least 50 observations within the relevant sub-sample.

Table A2: Tests of the Importance of Officer Sorting to Officer Effect Distribution

	Analysis Sample	*Low Availability Sample*		*Unlikely Response Sample*	
	(1)	(2)	(3)	(4)	(5)
		Corresponding	Low Availability	Corresponding	Unlikely
	Full Sample	Full Sample	Responses	Full Sample	Responses
Primary Results					
Contribution of Officer Effects					
Relative % of R-2 from Officer Effects	15.6%	15.2%	17.0%	15.1%	15.4%
Relative % of Adj. R-2 from Officer Effects	12.5%	12.4%	13.0%	12.4%	11.6%
Distribution of Officer Effects					
S.D. of Officer Effect	0.030	0.028	0.035	0.028	0.031
% Change: 1 S.D. Increase in Officer Effect	31.5%	31.0%	38.2%	31.1%	39.3%
Gap: 10th to 90th Percentile in Officer Effect	0.071	0.069	0.087	0.069	0.078
% Change: 10th to 90th Percentile in Officer Effect	76.3%	75.1%	94.0%	75.0%	97.7%
Auxilary Results					
Distribution of Officer Effects: Model w/o Incident Controls					
S.D. of Officer Effect	0.037	0.036	0.042	0.036	0.037
% Change: 1 S.D. Increase in Officer Effect	39.9%	39.1%	45.1%	39.3%	47.1%
Gap: 10th to 90th Percentile in Officer Effect	0.092	0.089	0.101	0.088	0.091
% Change: 10th to 90th Percentile in Officer Effect	98.0%	96.7%	109.3%	96.0%	114.9%
Correlation of Officer Effects					
Full model & Model w/o Incident Controls	0.913	0.911	0.924	0.909	0.918
Sub-sample & Corresponding Full Sample			0.801		0.820
Correlation of Officer Effects & Number of Observations	0.029	-0.002	0.010	-0.004	0.066
Mean of Outcome	0.094	0.092	0.092	0.091	0.080
S.D. of Outcome	0.291	0.288	0.290	0.288	0.271
Total Officers	1,858	1,353	1,353	1,329	1,329
Total Observations	322,280	284,453	142,488	283,237	141,297

This table summarizes the main analysis arrest results and tests of officer sorting. The contribution of officer effects shows the relative proportion of R^2 due to including officer effects in the model. The "Correlation of Officer Effects" compares the base model officer effects, $\hat{\theta}_i$, to officer effects, $\tilde{\hat{\theta}}_i$, that are estimated from a model that does not include offense characteristics and police beat fixed effects, X_{kt} and ϕ_g.

Figure A3: Bootstrap Benchmark Test: Distribution of Results under Assumption of No "True" Officer Effects (or all Officer Effects are Jointly Zero)

Figure A3.A: S.D. of Officer Effect Distribution

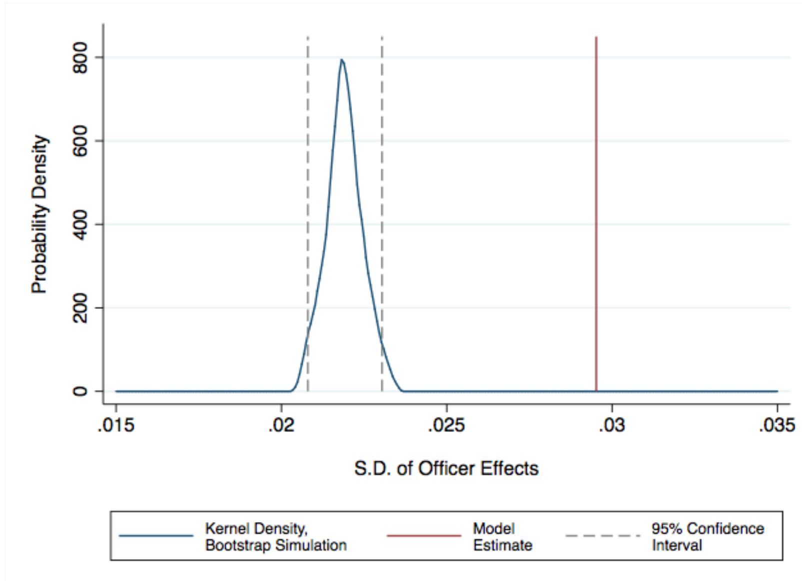
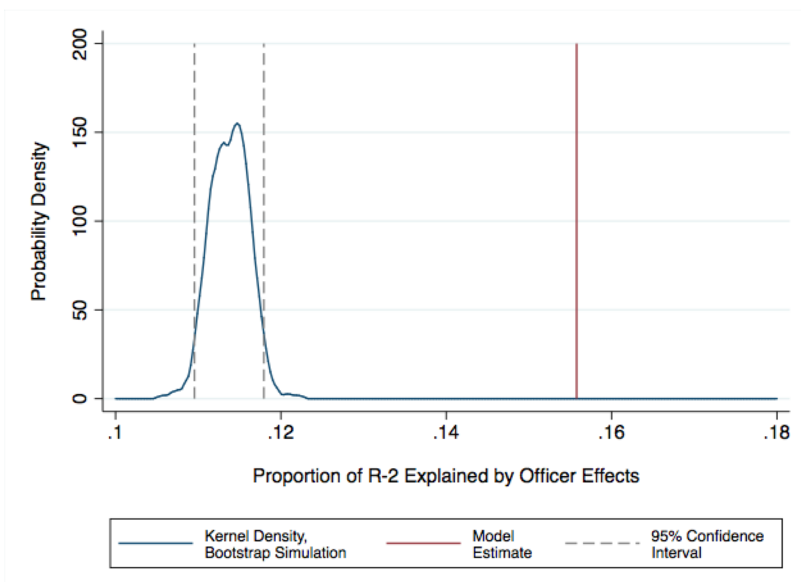


Figure A3.B: Proportion of R^2 Explained by Officer Effects



Each graph shows the residual bootstrap test distribution for the S.D. of the officer effect distribution as well as the relative proportion of R^2 explained by officer fixed effects. Each bootstrap iteration is obtained as follows: (1) Residuals and predicted outcomes are obtained from a first stage model that does not include Officer FE (under the null hypothesis that these variables are jointly zero), (2) Estimated residuals are assigned a wild bootstrap weight of $w \in \{1, -1\}$ that is constant within shift clusters δ_{gt}, with equal probability for each shift group, and these residuals are added to the predicted outcomes from (1), (3) Using these simulated outcome variables, the full model, including Officer FE, is estimated to obtain each statistic of interest. Post-estimation Empirical Bayes adjustments are made to the estimates after each iteration. Each test is based on 250 bootstrap replications.

I have also conducted this test imposing the restriction that $\tilde{Arrest}_b$ is binary in each iteration. To do this, I set the highest values of the outcome variable equal to one such that the mean of $\tilde{Arrest}_b$ equals the mean of $Arrest$ (approximately the top decile given an arrest mean of 10% in the sample. The results of this bootstrap test are similar and are available on request.

Table A3: Robustness Specification Tests

	(1) Base Model	(2) Add Reporting Area Fixed Effects	(3) Add Sector*Month Fixed Effects	(4) Add Individual Shift Effects	(5) Add Full Set of Dispatch & Location Codes	(6) First Stage Fixed Effects: >100 N	(7) First Stage Fixed Effects: Weighted by N	(8) First Stage Fixed Effects: Unadjusted
% of R-2 from Officer Effects	15.6%	14.7%	15.7%	9.5%	10.2%	15.1%	15.6%	15.6%
% of Adj. R-2 from Officer Effects	12.5%	11.8%	12.9%	9.5%	7.4%	12.4%	12.5%	12.5%
S.D. of Officer Effect	0.030	0.030	0.030	0.036	0.027	0.033	0.033	0.037
% Change: 1 S.D. Increase in Officer Effect	31.5%	31.7%	32.4%	38.1%	28.9%	36.1%	35.4%	40.0%
Gap: 10th to 90th Percentile in Officer Effect	0.071	0.072	0.073	0.084	0.065	0.079	0.077	0.087
% Change: 10th to 90th Percentile in Officer Effect	76.3%	76.4%	78.2%	90.1%	69.5%	86.4%	82.2%	92.5%
Correlation to Arrest Officer Effect		0.996	0.994	0.951	0.964	0.988	0.991	0.987
Mean of Outcome	0.094	0.094	0.094	0.094	0.094	0.092	0.094	0.094
S.D. of Outcome	0.291	0.291	0.291	0.291	0.291	0.288	0.291	0.291
Total Officers	1,858	1,858	1,858	1,858	1,858	1,331	1,858	1,858
Total Observations	322,280	321,907	322,280	321,651	321,887	283,435	322,280	322,280

This table summarizes the analysis results across robustness specifications. Column (1) replicates the results from the primary specification. In column (2), police beat FE, ϕ_g, are replaced with reporting area FE, a finer geographic unit. In column (3), police beat FE, ϕ_g, are replaced with Sector*Month FE, which interact the 35 geographic police sectors with month indicators. Column (4) includes individual shift indicators or Date*Division*8 hour Shift fixed effects rather than shifts aggregated by month. Column (5) inserts the full set of 119 dispatch codes and 34 location type codes available in the data, rather than the broader 22 dispatch codes and 11 location type codes used in the preferred specification. Column (6) reports the dispersion in unadjusted officer fixed effects from the first stage, using a sample restricted to officers with more than 100 observations. Column (7) calculates the dispersion metric as a standard deviation of unadjusted officer fixed effects from the first stage that is weighted by the number of observations for each officer. In column (7) the correlation to base model officer effects is weighted by the number of observations for each officer. Column (8) reports dispersion in the unadjusted officer fixed effect estimates. The specifications in column (7) in Panel (A) are removed from Panel (B) because it involves weighted data. Correlations shown in the bottom panel are calculated for overlapping observations across different specifications.

Table A4: Summary Statistics, Racial Bias Test Sample

Table A4.A: Summary Statistics: Outcomes, Officers, and Complainants

	Racial Bias Test Sample		*Black Officers*		*Hispanic Officers*		*White Officers*	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total Observations	165,845		42,015		27,842		91,687	
Total Incidents	135,793		35,629		23,425		72,546	
Total Officers	1,303		342		237		688	
Outcomes								
Arrest	0.08	(0.27)	0.07	(0.26)	0.06	(0.24)	0.09	(0.28)
Arrestee Black	0.02	(0.15)	0.03	(0.17)	0.02	(0.13)	0.02	(0.15)
Arrestee Hispanic	0.01	(0.10)	0.01	(0.08)	0.01	(0.10)	0.01	(0.10)
Arrestee White	0.01	(0.10)	0.01	(0.08)	0.01	(0.09)	0.01	(0.10)
Arrest has Race Demographics	0.04	(0.20)	0.04	(0.20)	0.04	(0.19)	0.05	(0.21)
Officer Characteristics								
Officer Arrest Rate	0.08	(0.05)	0.07	(0.04)	0.06	(0.04)	0.09	(0.05)
Two Responders	0.36	(0.48)	0.30	(0.46)	0.31	(0.46)	0.41	(0.49)
Total Incidents	175.56	(112.07)	161.54	(85.37)	161.69	(108.82)	187.12	(123.27)
Trainee	0.04	(0.19)	0.01	(0.12)	0.04	(0.20)	0.04	(0.21)
Sergeant	0.01	(0.08)	0.01	(0.10)	0.00	(0.06)	0.00	(0.07)
Salary ($10,000s)	5.99	(1.12)	6.01	(1.11)	5.83	(1.09)	6.02	(1.15)
Years of Experience	13.37	(9.47)	13.62	(8.89)	11.65	(9.23)	13.87	(9.87)
Age	39.45	(10.29)	40.82	(10.20)	37.41	(10.24)	39.53	(10.35)
Female	0.15	(0.36)	0.24	(0.43)	0.14	(0.34)	0.12	(0.32)
Black	0.25	(0.43)	1					
Hispanic	0.17	(0.37)			1			
White	0.55	(0.50)					1	
Officer has Race Demographics	1		1		1		1	
Complainant Characteristics								
Victim with Injury	0.10	(0.30)	0.09	(0.29)	0.10	(0.30)	0.11	(0.32)
Number of Complainants	1.07	(0.33)	1.06	(0.30)	1.07	(0.33)	1.08	(0.34)
Female	0.49	(0.50)	0.51	(0.50)	0.49	(0.50)	0.48	(0.50)
Black	0.35	(0.48)	0.47	(0.50)	0.31	(0.46)	0.31	(0.46)
Hispanic	0.33	(0.47)	0.28	(0.45)	0.40	(0.49)	0.32	(0.47)
White	0.33	(0.47)	0.25	(0.43)	0.31	(0.46)	0.37	(0.48)
Complainants with Race Demographics	0.80	(0.40)	0.82	(0.39)	0.81	(0.39)	0.79	(0.40)

These tables display summary statistics the racial bias test analysis. The first column, "Racial Bias Test Sample", summarizes the primary analysis sample. This sample is restricted to observations where there is a single race for responding officers (if there are multiple officers) and each officer has more than 50 observations. Officer arrest rate and number of incidents is calculated over all observations in the raw data. Columns (2) - (4) describe characteristics of observations for Black, Hispanic, and White officers within the "Racial Bias Test Sample."

Table A4.B: Summary Statistics: Incident Urgency, Location Type, and Dispatch Code

	Racial Bias Test Sample		*Black Officers*		*Hispanic Officers*		*White Officers*	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total Observations	165,845		42,015		27,842		91,687	
Total Incidents	135,793		35,629		23,425		72,546	
Total Officers	1,303		342		237		688	
Call Urgency								
Time to Dispatch (Minutes)	25.34	(28.34)	25.64	(28.37)	26.67	(29.10)	24.62	(27.98)
Location Type								
Apartment	0.13	(0.34)	0.15	(0.35)	0.12	(0.33)	0.13	(0.34)
Residence Other	0.16	(0.37)	0.18	(0.39)	0.19	(0.39)	0.15	(0.36)
Bar/Club/Entertainment	0.03	(0.17)	0.03	(0.16)	0.03	(0.17)	0.03	(0.18)
Retail	0.07	(0.25)	0.07	(0.25)	0.06	(0.23)	0.07	(0.26)
Business Other	0.05	(0.23)	0.04	(0.20)	0.05	(0.22)	0.06	(0.24)
Govt/Health/School/Religion	0.01	(0.10)	0.01	(0.10)	0.01	(0.10)	0.01	(0.10)
Motor Vehicle	0.02	(0.14)	0.03	(0.17)	0.02	(0.13)	0.02	(0.12)
Parking Lot	0.25	(0.43)	0.23	(0.42)	0.24	(0.43)	0.26	(0.44)
Street	0.17	(0.37)	0.13	(0.34)	0.17	(0.37)	0.19	(0.39)
Outdoor Other	0.05	(0.22)	0.07	(0.25)	0.07	(0.25)	0.04	(0.19)
Other Location	0.05	(0.22)	0.07	(0.25)	0.05	(0.21)	0.05	(0.21)
Dispatch Code Type								
Criminal Assault, High Priority	0.01	(0.11)	0.01	(0.10)	0.01	(0.10)	0.01	(0.11)
Armed Encounter/Active Shooter, High Priority	0.02	(0.13)	0.02	(0.13)	0.02	(0.13)	0.02	(0.13)
Robbery, High Priority	0.06	(0.23)	0.05	(0.22)	0.06	(0.23)	0.06	(0.24)
Burglary of Business, High Priority	0.01	(0.08)	0.01	(0.07)	0.01	(0.09)	0.01	(0.09)
Burglary of Business, Low Priority	0.05	(0.22)	0.04	(0.21)	0.05	(0.22)	0.05	(0.22)
Burglary of Residence, High Priority	0.03	(0.16)	0.02	(0.15)	0.03	(0.16)	0.03	(0.16)
Burglary of Residence, Low Priority	0.10	(0.30)	0.11	(0.32)	0.11	(0.31)	0.09	(0.28)
Burglary of Vehicle, High Priority	0.02	(0.15)	0.02	(0.14)	0.03	(0.16)	0.02	(0.15)
Burglary of Vehicle, Low Priority	0.17	(0.38)	0.17	(0.37)	0.18	(0.38)	0.17	(0.37)
Unauthorized Use of Vehicle, High Priority	0.01	(0.09)	0.01	(0.09)	0.01	(0.09)	0.01	(0.09)
Unauthorized Use of Vehicle, Low Priority	0.04	(0.20)	0.04	(0.21)	0.05	(0.21)	0.04	(0.20)
Theft, High Priority	0.02	(0.15)	0.03	(0.16)	0.02	(0.15)	0.02	(0.15)
Theft, Low Priority	0.06	(0.23)	0.06	(0.25)	0.06	(0.24)	0.05	(0.22)
Criminal Mischief, High Priority	0.01	(0.08)	0.01	(0.08)	0.01	(0.09)	0.01	(0.09)
Criminal Mischief, Low Priority	0.07	(0.25)	0.08	(0.27)	0.07	(0.26)	0.06	(0.24)
Major Disturbance, High Priority	0.11	(0.31)	0.13	(0.34)	0.10	(0.30)	0.11	(0.31)
Accident, High Priority	0.03	(0.18)	0.02	(0.15)	0.03	(0.18)	0.04	(0.20)
Accident, Low Priority	0.07	(0.25)	0.05	(0.21)	0.06	(0.24)	0.08	(0.27)
Injured Person, High Priority	0.00	(0.06)	0.00	(0.06)	0.00	(0.06)	0.00	(0.05)
Injured Person, Low Priority	0.01	(0.11)	0.01	(0.12)	0.01	(0.10)	0.01	(0.10)
Other, High Priority	0.05	(0.22)	0.04	(0.20)	0.05	(0.21)	0.05	(0.23)
Other, Low Priority	0.05	(0.23)	0.06	(0.23)	0.04	(0.21)	0.06	(0.23)

These tables display summary statistics the racial bias test analysis. The first column, "Racial Bias Test Sample", summarizes the primary analysis sample. This sample is restricted to observations where there is a single race for responding officers (if there are multiple officers) and each officer has more than 50 observations. Columns (2) - (4) describe characteristics of observations for Black, Hispanic, and White officers within the "Racial Bias Test Sample."

Figure A4: Power of Racial Bias Test, Bootstrap Simulation

Figure A4.A: Power of the Test, Deviations in Black Officer Race Coefficient

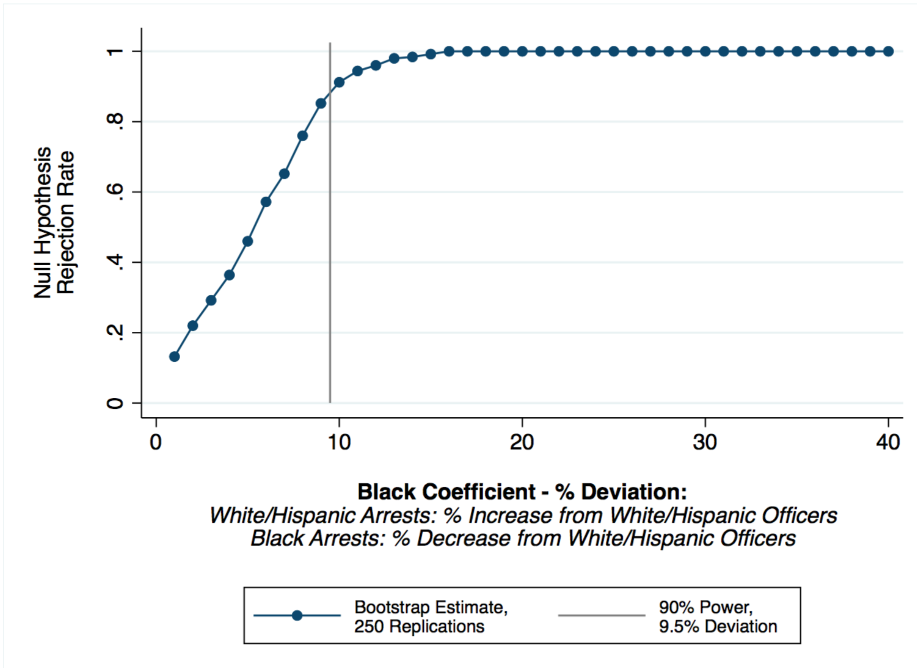
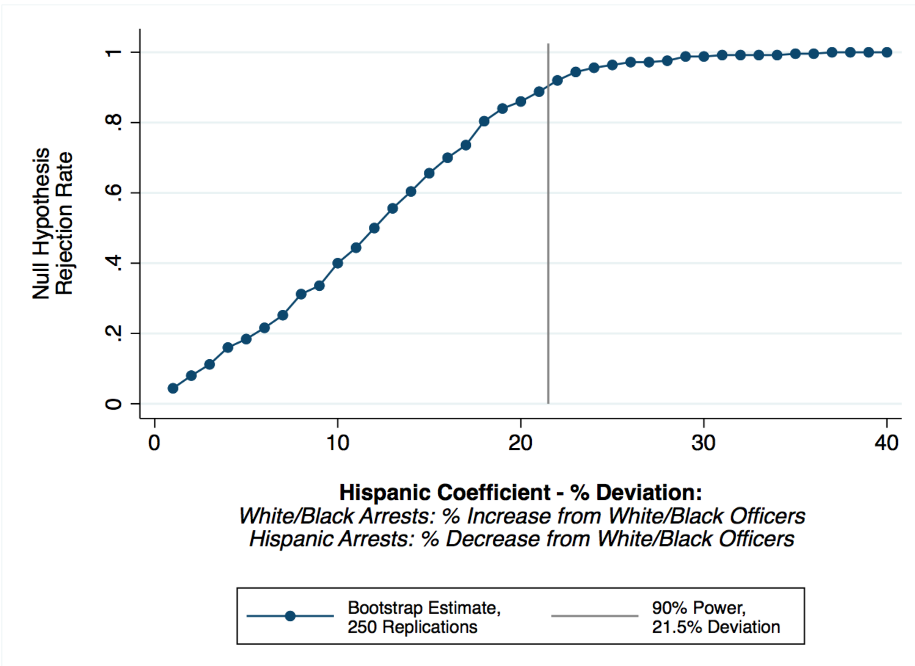


Figure A4: Power of Racial Bias Test, Bootstrap Simulation

Figure A4.B: Power of the Test, Deviations in Hispanic Officer Race Coefficient



These graphs plot the rejection rate of the racial bias test when an alternative hypothesis of racial bias is imposed in a bootstrap simulation, or the power of the test. The X-axis portrays different alternative hypotheses, expressed as a constant percentage deviation in arrest race outcomes caused by the altered officer race coefficient, relative to the arrest average for officers of other races. The deviation is an increase for arrestee outcomes when the officer and arrestee are different races, and a decrease when the officer and arrestee are the same race.

This simulation is conducted using the following steps (Black coefficient example, $\alpha_{Black,r}$). I first regress the second stage $\theta_{i,r}$ outcomes on officer demographics excluding $\alpha_{Black,r}$ and recover a predicted $\hat{\theta}'_{i,r}$ and residual $\hat{r}_{i,r}$. In each bootstrap iteration (of 250 iterations), I draw a wild bootstrap weight $w_b \in \{-1, 1\}$ with equal probability. I then impose an alternative hypothesis on the $\alpha_{Black,r}$ coefficient, and set $\tilde{\theta}^b_{i,r} = \hat{\theta}'_{i,r} + \Delta\alpha_{Black,r} + w_b\hat{r}_{i,r}$. I set Δ to be a constant percent increase for White and Hispanic ("other race") arrestee outcomes and an equivalent percent decrease in the Black ("own race") arrestee outcome. These percent changes are set relative to White or Hispanic officer averages for the total arrest outcome, using the max of the two officer groups (White or Hispanic) for percent increases and the min of the two for the percent decrease. I then regress these simulated values, $\tilde{\theta}^b_{i,r}$, on the full set of officer demographic variables and use F-tests to determine whether the ranking of officer race groups changes across arrestee race outcomes.

A2 Coefficients in the First Stage of Model (Online Appendix)

In the body of the paper, I restrict attention to aspects of officer effects because this paper focuses on estimating differences in officer arrest behavior, $\hat{\theta}_i$, and the importance of officers in predicting arrests. This appendix discusses other components of the arrest prediction model.

Table A5 shows the first stage coefficients for incident characteristics, X_{kt}. First, the probability of an arrest is increasing in call severity or urgency, at a decreasing rate. This is shown by the "Time to Dispatch" variables that measure the number of minutes that lapse between when a call is made by the complainant and when an officer is dispatched to the scene. An increase of 10 minutes in this time gap decreases the likelihood that an arrest is made by 2% (relative to the average time gap). The average call in the data has a time difference of 24 minutes between the call and dispatch time, which corresponds to a decrease the likelihood of arrest by 14% relative to an instantaneously dispatched call.

Next, the model includes direct controls for complainant demographic information, relative to the omitted category of complainants with no demographic data. The likelihood of arrest increases when there are more complainants. All race categories can be included as controls because some calls have multiple complainants of different races. Arrests are less likely when the complainant is female. Lastly, arrests are less likely when there is demographic information for the complainant and when there is a victim injury. Complainants without listed demographics are often businesses, so the negative complainant demographic information coefficient suggests that arrests are more likely when offenses occur in business establishments.

The third set of incident controls in the model are dispatch and location type codes. These variables are generally more positive for crimes that are more serious or are likely to have the evidence necessary to make an arrest. The omitted category of minor incidents (other minor) have the highest arrest rates, possibly because officers have the most discretion to make an arrest at the scene in these cases. A 1 standard deviation in the officer effect distribution is comparable to the difference in arrest likelihood moving from a “Low Priority Accident” to a “High Priority Robbery.” For location, incidents that occur in a business or public institution setting appear more likely to result in arrest. This may be related to security surveillance systems used in businesses.

Lastly, the model includes indicator variables for the hour within each shift. Interestingly, arrests are less likely in the sixth and seventh shift hour than in the last hour of a shift. This may relate to officer overtime pay incentives. If officers make arrests in the last hour of their shift, they may be more likely to receive overtime pay for activities related to filing the arrest, including booking the individual in the county jail and writing the arrest report for the incident. Officers are 9% less likely to make arrests in the sixth and seventh hour of their shift relative to the last hour of their shift.

Overall, the incident context controls are important predictors of whether an incident results in an arrest. As noted above, these variables collectively account for $\approx 50 - 65\%$ of the explainable variation in arrest outcomes.

Table A5: Covariate Coefficients in First Stage of Arrest Model

Call Urgency		***Dispatch Codes***		***Location Type Codes***	
Time to Dispatch (Minutes)	-0.0008***	Criminal Assault, High Priority	-0.0175+	Apartment	-0.0436***
	(0.0001)		(0.0096)		(0.0033)
Time to Dispatch (Minutes), Squared	0.00001***	Armed Encounter/Active Shooter, High Priority	-0.0531***	Residence	-0.0427***
	(0.0000)		(0.0079)		(0.0032)
		Robbery, High Priority	-0.1583***	Car	-0.0219***
Complainant Characteristics			(0.0058)		(0.0052)
Victim Injury	-0.0231***	Burglary of Business, High Priority	-0.1688***	Parking Lot	-0.0351***
	(0.003)		(0.0108)		(0.0031)
Number of Complainants	0.0103***	Burglary of Business, Low Priority	-0.2597***	Bar/Club/Entertainment	-0.003
	(0.0023)		(0.0058)		(0.0051)
Complainant, Black	0.03***	Burglary of Residence, High Priority	-0.1052***	Retail	0.0984***
	(0.0026)		(0.0061)		(0.0048)
Complainant, Hispanic	0.0222***	Burglary of Residence, Low Priority	-0.1724***	Business Other	0.0185***
	(0.0026)		(0.0047)		(0.0045)
Complainant, White	0.0306***	Burglary of Vehicle, High Priority	-0.1174***	Govt/Health/School/Religion	0.0144+
	(0.0025)		(0.0066)		(0.0084)
Complainant, Other	0.0143***	Burglary of Vehicle, Low Priority	-0.1929***	Street	-0.0414***
	(0.0036)		(0.0048)		(0.0032)
Complainant, Female	-0.0071***	Unauthorized Use of Vehicle, High Priority	-0.1062***	Outdoor Other	-0.0194***
	(0.0013)		(0.0103)		(0.004)
Complainant, Race Demographic Information	-0.0607***	Unauthorized Use of Vehicle, Low Priority	-0.1715***	*Omitted Category: Other*	
	(0.0026)		(0.0057)		
Omitted Category:		Theft, High Priority	-0.1663***		
No Complainant Demographics			(0.0068)		
		Theft, Low Priority	-0.1949***		
Hour Within Shift			(0.0051)		
First Hour	0.0023	Criminal Mischief, High Priority	-0.0356***		
	(0.0025)		(0.0107)		
Second Hour	0.0018	Criminal Mischief, Low Priority	-0.1775***		
	(0.0025)		(0.0048)		
Third Hour	0.0027	Major Disturbance, High Priority	-0.0205***		
	(0.0027)		(0.0053)		
Fourth Hour	-0.0023	Accident, High Priority	-0.1565***		
	(0.0027)		(0.0057)		
Fifth Hour	-0.001	Accident, Low Priority	-0.1883***		
	(0.0028)		(0.0049)		
Sixth Hour	-0.0084**	Injured Person, High Priority	-0.1519***		
	(0.0028)		(0.011)		
Seventh Hour	-0.0083**	Injured Person, Low Priority	-0.0971***	Observations	322,280
	(0.003)		(0.008)	R-Squared	0.221
Omitted Category:		Other, High Priority	-0.117***	Arrest Outcome Mean	0.097
8th Hour			(0.0055)	Arrest Outcome Standard Deviation	0.296
		Omitted Category: Other, Low Priority			
Other Controls in Model:	First Officer and Co-responder fixed effects, Shift (Day-of-the-week*8-Hour Shift*Month*Year) fixed effects, Police Beat fixed effects				

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1
This table shows the results of the first stage arrest model. Coefficients are shown for each covariate in the vector X_{kt}. Robust standard errors are clustered at the level of the focal officer, i, and the shift category, δ_{gt}.

A3 Empirical Bayes Shrinkage Estimates (Online Appendix)

As outlined in the text, the estimates of permanent officer arrest propensity are adjusted using Empirical Bayes techniques. Empirical Bayes techniques are useful when a statistician observes a large number of different estimates of parameters that are drawn from the same underlying distribution, and each estimate is measured with error. These techniques are detailed in work by Morris (1983), and are commonly employed in the economics of education literature on teacher value added (e.g. Guarino et al., 2015; Koedel et al., 2015; Chetty et al., 2014; Kane and Staiger, 2008; Aaronson et al., 2007). A number of different variants of Empirical Bayes techniques have been used in the prior literature, the estimation in this paper shares features with Guarino et al. (2015); Chetty et al. (2014); Aaronson et al. (2007). In robustness checks in the paper, I show that the results do not substantively change when a number of alternate precision adjustments are used.

In this paper, I observe sample estimates of officer arrest propensity, $\bar{r}_i$, which are derived from a first stage regression model. Each of these estimates is an approximation of a "true" officer arrest propensity, θ_i, though some officer estimates are derived from more observations and are thus more precise than others. The underlying parameters, θ_i, can also be thought of as random variables which are derived from a separate distribution of potential officer arrest propensities. Empirical Bayes techniques develop a "prior" distribution for the underlying distribution of θ_i that is estimated empirically from the data on all officers. The estimation constructs a weighted mean of the observational estimate and the "prior."

Each θ_i is assumed to be independent and identically distributed across G total officers. The underlying distribution of each $\bar{r}_i$ and the total distribution of θ_i across i are given by:

$$\begin{aligned} \bar{r}_i|\theta_i &\sim N(\theta_i, \frac{\sigma^2_{\varepsilon,i}}{N_i}) \\ \theta_i|\mu, \sigma^2_A &\sim N(0, \sigma^2_A) \end{aligned}$$

The mean of the distribution of θ_i is known to be 0 in this setting, given the normalization of the fixed effects in the model. Given a "prior" for the distribution of θ_i, the posterior distribution of $\theta_i|\bar{r}_i$ give the adjusted estimates of $\hat{\theta}_i^{EB}$ used in this paper:

$$\begin{aligned} \theta_i^{EB}|\bar{r}_i, \sigma^2_{\varepsilon,i}, \sigma^2_A &\sim N(B\bar{r}_i, B\frac{\sigma^2_{\varepsilon,i}}{N_i}) \\ where\ B &= \frac{\sigma^2_A}{\sigma^2_A + \frac{\sigma^2_{\varepsilon,i}}{N_i}} \end{aligned}$$

I derive estimates of officer arrest propensity, $\hat{\theta}_i^{EB}$, using the following steps:

1. Estimate the first stage of the model and calculate residuals, $\hat{r}_{ikgt}$, and their officer-

level average, $\bar{r}_i$. I include all officer fixed effects in the first stage regression to allow for arbitrary correlations between responding officers and the other covariates in the model to improve the estimation of the residuals. This procedure is similar to the first stage approach used in Chetty et al. (2014).

$$\begin{aligned} Arrest_{ikgt} &= \theta_i + \theta_j + \pi X_{kt} + \delta_{gt} + \phi_g + \varepsilon_{ikgt} \\ \hat{r}_{ikgt} &= \hat{\theta}_i + \hat{\varepsilon}_{ikgt} \\ \bar{r}_i &= \frac{1}{N_i} \sum_{N_i} \hat{r}_{ikgt} \end{aligned}$$

2. Calculate individual variance estimates, $\hat{\sigma}^2_{\varepsilon,i}$ and solve for a sample analog of the prior variance of θ_i, $\hat{\sigma}^2_A$.

$$\begin{aligned} \hat{\sigma}^2_{\varepsilon,i} &= \frac{1}{N_i - 1} \sum_{N_i} (\hat{r}_{ikgt} - \bar{r}_i)^2 \\ \sigma^2_A &= E[r^2_{ikgt}] - E[\varepsilon^2_{ikgt}] \\ \hat{\sigma}^2_A &= \frac{1}{N - G - K} \sum_{G} \sum_{N_i} \hat{r}^2_{ikgt} - \frac{1}{N - G} \sum_{G} N_i \hat{\sigma}^2_{\varepsilon,i} \end{aligned}$$

with $N-G-K$ are the degrees of freedom in the first stage regression, given G officers and K regressors in the first stage model.[28]

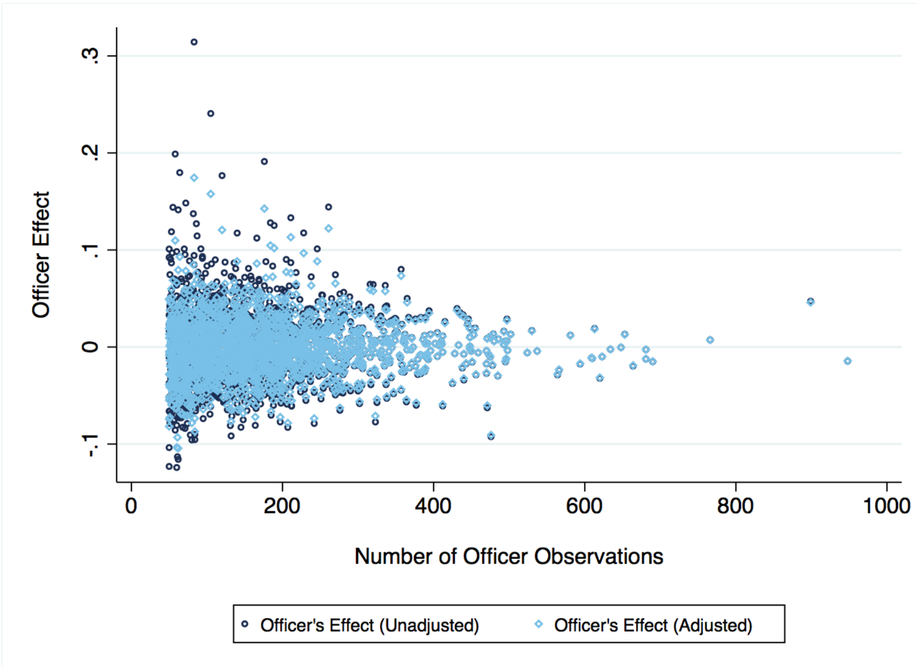
3. Calculate the posterior estimates $\hat{\theta}^{EB}_i$ by applying the shrinkage factor $\hat{B}$. The shrinkage factor is always less than 1 and is increasing in N_i and decreasing in $\sigma^2_{\varepsilon,i}$. This factor gives higher weight to police officer arrest propensity estimates that are more precisely measured and shrinks less precise estimates toward 0, the center of the distribution.

$$\hat{\theta}^{EB}_i = \frac{\hat{\sigma}^2_A}{\hat{\sigma}^2_A + \frac{\hat{\sigma}^2_{\varepsilon,i}}{N_i}} \cdot \bar{r}_i$$

[28]Given that I absorb four sets of fixed effects in the model, θ_i, θ_j, δ_{gt}, and ϕ_g, K contains the number of group categories in the non-focal fixed effects. In practice, the degrees of freedom must also be adjusted for the number of omitted reference categories in the model, or the number of "mobility groups", M. The actual degrees of freedom used is $N - G - K + M$.

The following figure displays the relationship between the unadjusted and adjusted estimates of officer effects and the number of observations per officer:

Figure A5: Adjusted and Unadjusted Officer Effects



The above figure shows unadjusted officer fixed effects overlaid with the adjusted officer effects estimates used in this paper. The correlation between the adjusted and unadjusted estimates is 0.987.

A4 Economic Model for Racial Bias Test (Online Appendix)

In this section, I outline the model used to test for the presence of racial bias among officers. The model is adapted from the test of racial bias in Anwar and Fang (2006) to the setting of police calls for service, where the econometrician does not directly observe officer effort choices or suspect race.

Relative Ranking of Arrest Rates by Officer Race

There are two races for officers and suspects in the illustrative model, $r \in \{M, W\}$. The model examines officer responses to incidents of a similar type, with the same observable characteristics. When officers arrive to respond to an incident they observe the suspect's race but this information is not observed by the econometrician.[29] The likelihood that a call has a suspect of a given race, r_s, is ψ^{r_s}. Note that $\psi^M + \psi^W \leq 1$ if some calls for service do not have any relevant suspect, which may occur for accidents or incidents that do not have a clear party that is at fault.

For each suspect race, r_s, π^{r_s} is the likelihood that an arrest is *feasible* if an officer exerts effort to respond. An arrest is feasible if there is a minimum basis for an arrest. Feasibility is a function of the total amount of evidence that is capable of being recovered for an incident, which may vary according to suspect characteristics and may be correlated with suspect race. The basis for an arrest could also be higher if a suspect has a criminal history that may be discovered during a response. Criminal history is a suspect characteristic that may also be correlated with suspect race.

More severe offenses will also have a higher basis for an arrest. The regression application of this test will directly control for observable components of the severity or arrest feasibility of incidents. However, there may be unobservable characteristics of incidents that affect arrest feasibility. If the composition of unobservable characteristics is correlated with suspect race, π^{r_s} will also differ across suspect race.

When an officer responds to an incident, he observes information related to the offense that provides a signal of whether an arrest is feasible. This information may include evidence immediately available at the scene, cooperation of the victim, location of the incident, etc. The information at the scene is summarized by an index $s \in [0, 1]$. If an arrest is feasible, s is randomly drawn from the distribution $f_a^{r_s}(s)$, while if the arrest is not feasible, s is randomly drawn from the distribution $f_n^{r_s}(s)$. These distributions are allowed to differ across suspect race, r_s, reflecting the fact that total information content in responses may differ for different suspect races.

The distributions $f_a^{r_s}(s)$ and $f_n^{r_s}(s)$ have the following properties:

- Both are defined over the full support of $s \in [0, 1]$

[29] The data does include some information about suspects. The data includes records of suspects identified by officers prior to 2017, as well as characteristics of suspects unknown to officers at the conclusion of a response. However, suspect information is always recorded by officers as part of a response to a call, so it is treated as an outcome of a response rather than a given characteristic of an incident.

- Monotone Likelihood Ratio Property: $\frac{f_a^{r_s}(s)}{f_n^{r_s}(s)}$ is strictly increasing in s. This implies that a higher s means an arrest is more likely to be feasible.
- Unbounded Likelihood Ratio: $\frac{f_a^{r_s}(s)}{f_n^{r_s}(s)} \to \infty$ as $s \to 1$. This implies that very high signals θ provide nearly certain information that an arrest is feasible.
- $F_a^{r_s}(s)$ first order stochastically dominates $F_n^{r_s}(s)$.

Officers make a discrete effort choice $E \in \{0, 1\}$ after viewing the suspect race and the signal, $\{r_s, s\}$. If an officer chooses to exert effort, the posterior likelihood of arrest is increasing in s and is given by Bayes' Rule:

$$P(A|r_s, s) = \frac{\pi^{r_s} f_a^{r_s}(s)}{\pi^{r_s} f_a^{r_s}(s) + (1 - \pi^{r_s}) f_n^{r_s}(s)}$$

Each officer will receive a benefit if an arrest is made that is normalized to 1 and faces a cost of effort that varies by both suspect race and officer race, $t(r_s, r_p) \in [0, 1]$. If the officer chooses not to exert effort, he receives a benefit of zero. Each officer maximizes his utility as a choice between effort and no effort:

$$max\{P(A|r_s, s) - t(r_s, r_p), 0\}$$

Officers will exert effort in response to an incident when $P(A|r_s, s) \geq t(r_s, r_p)$. As a result, it can be shown that officers will exert effort on a suspect of race, r_s, if the value of $s \geq s^*(r_s, r_p)$, where the threshold $s^*(r_s, r_p)$ satisfies $P(A|r_s, s^*(r_s, r_p)) = t(r_s, r_p)$. This effort threshold is monotonically increasing in $t(r_s, r_p)$.

In the setting of calls for service, both officer effort choices and suspect race are not observed in the data. Instead, the econometrician can observe the number of arrestees of a given race adjusted by the total number of incidents in the data, or "unconditional" arrestee race outcomes. Allow these unconditional arrestee race outcomes to be denoted as the "arrest rate" for an arrestee race group and officer race combination.

In terms of the model, the arrestee race outcome for arrestees of race, r_s, and officers of race, r_p, is:

$$A(r_s, r_p) = \psi^{r_s} \pi^{r_s} [1 - F_a^{r_s}(s^*(r_s, r_p))]$$

This arrest rate is decreasing in the signal threshold $s^*(r_s, r_p)$ and is also decreasing in the cost of effort $t(r_s, r_p)$.

The following definitions characterize officer race specific costs:

1. Racial Bias: Officers are racially biased with respect to suspects if for some officer race, r_p, $t(M, r_p) \neq t(W, r_p)$.
2. Monolithic Behavior: Officers are not monolithic in their behavior if officer costs differ across officer race for a given suspect race, or $t(r_s, M) \neq t(r_s, W)$.
3. Statistical Discrimination: Assume $t(M, r_p) = t(W, r_p)$, or officers are not racially biased. Then race r_p officers will exhibit statistical discrimination if $s^*(M, r_p) \neq s^*(W, r_p)$.

If officers are not racially biased and exhibit monolithic behavior across officer race, then $t(M, M) = t(M, W) = t(W, W) = t(W, M)$. It follows that arrest rates within suspect race will be constant across officer race, but that total arrest rates for different suspect races may differ if $s^*(M, r_p) \neq s^*(W, r_p)$, or there is statistical discrimination.

The test also allows officers to have differing total costs of effort that vary by officer race, or behave in a manner that is not monolithic. The first half of this paper assesses whether individual officers behave differently from one another in their responses to similar incidents, which can be interpreted as evidence that individual officers are not monolithic in their behavior. If officers do not exhibit monolithic behavior but are also not prejudiced, then the ranking of arrest rates across officer race within suspect race will be independent of suspect race, or constant across suspect race.

For example, allow minority officers to have a higher cost of effort than White officers for any race of suspect. Then:

$$
\begin{aligned}
t(M, M) > t(M, W) \quad &\& \quad t(W, M) > t(W, W) \\
t(M, M) = t(W, M) \quad &\& \quad t(M, W) = t(W, W) \\
s^*(M, M) \quad &> \quad s^*(M, W) \\
\& \;\; s^*(W, M) \quad &> \quad s^*(W, W) \\
A(M, M) \quad &< \quad A(M, W) \\
\& \;\; A(W, M) \quad &< \quad A(W, W)
\end{aligned}
$$

Or in this case, minority officers will be less likely to make arrests than White officers for both suspect races. In other words, the relative ranking of minority and White officers is the same for both suspect groups. We can conclude that if both races of officers are not biased, the relative ranking of arrest rates across officer race should be the same for incidents within each suspect race.

Generally, the test proposed in this paper allows total arrest rates to differ across arrestee race by focusing attention on relative rankings of officer arrest rates rather than total levels of officer arrest rates. This feature allows officers to behave in a manner that is consistent with statistical discrimination and isolates officer behavioral patterns associated with taste-based racial bias. Statistical discrimination will occur in this model if total arrest rates for one suspect group is higher than the other suspect group but the relative ranking of officer arrest rates is the same for both suspect groups. For example, it may be the case that minority suspects are more likely to have a criminal history and this causes the total signal threshold to be lower for incidents with minority suspects, $s^*(M, r_p) < s^*(W, r_p)$.

Conversely, a reversal in the rank order of arrest rates across officer race for different suspect race groups will provide evidence of taste-based racial bias.

This is illustrated by the following stylized example:

$$\begin{aligned}
t(M,M) > t(W,M) \quad \& \quad & t(W,W) > t(M,W) \\
\& \;\; t(W,W) = t(M,M) \quad \& \quad & t(W,M) = t(M,W) \\
s^*(M,M) \quad > \quad & s^*(M,W) \\
\& \;\; s^*(W,M) \quad < \quad & s^*(W,W) \\
A(M,M) \quad < \quad & A(M,W) \\
\& \;\; A(W,M) \quad > \quad & A(W,W)
\end{aligned}$$

If arrests are higher for White officers relative to minority officers when responding to incidents with minority suspects, $A(W,W) < A(W,M)$, and arrests are higher for minority officers relative to White officers when responding to White suspects, $A(M,M) < A(M,W)$, we can conclude that one or both officer race groups is biased. This opposing rank order violates the null hypothesis of no racial bias among officers.

A5 Data Appendix (Online Appendix)

Several different data files were used for this project. This Appendix summarizes the decisions made in cleaning and constructing the data set used for this project.

Incident Offense Data The base file used in this project is the DPD "Police Incidents" file accessed through the Dallas Open Data portal. Data sets compiled and released through this portal are updated daily, with each new data set consisting of a moving time window of records. Because old records are replaced with new records in this interface, I have periodically downloaded new versions of the data (on an approximate monthly basis), updating the existing records with new downloads. Throughout this project, I use the most recently updated record for each incident in the data when there are duplicate records for the same incident across downloaded versions. I allow a grace period of one month to pass before using an incident record, as records may be corrected retrospectively. This procedure allows me to use the most complete set of information for each observation and increase fidelity in comparisons of incidents over time, as some records may be collected, modified, or updated retrospectively. This data updating process is important because DPD releases the records as they occur and some records may be incomplete or inaccurate. The data in this project covers the time window of June 2014 through mid-October 2018.

The data includes all incidents that are reported to DPD by a reporting party or complainant, with the exception of sexually oriented offenses, offenses involving juveniles, and social service referral offenses. While the data encompasses calls for service, it also includes other incidents reported by complainants through other means, as well as some officer-initiated interactions. I use information included in several of the data fields to narrow the set of observations to include only records that are both "highly likely" to correspond to calls for service and are relatively complete.

First, I exclude observations that do not have information on the time that a call was received or dispatched. Second, I exclude calls that do not have a listed police division where the call occurred or have a missing address.

Third, I exclude incidents that were dispatched more than 2 hours after a call was received. I am unable to distinguish between 9-1-1 calls and 3-1-1 calls in the data. While 3-1-1 calls are placed for less serious incidents than 9-1-1 calls, they are connected to the same call-taker lines as emergency 9-1-1 calls and follow the same protocols for response if a response is warranted. Because I only consider calls that received a response and were dispatched within 2 hours that the call was received, the observations are weighted toward 9-1-1 calls. Additionally, by including a control for the urgency of a call in the estimation model I am able to generally distinguish between these two types of calls in the model.

Fourth, I exclude any incident that does not include a complainant record, or that has a complainant listed as the City of Dallas, the Dallas Police Department, or another city, or police, sheriff or fire department, as these calls are unlikely to involve a civilian complainant. Next, through conversations with officers and dispatchers at DPD, I identify a set of dispatch codes that were unlikely to originate with a call from a complainant. These include calls where an officer is not the first responder but is called to join another officer on an existing call, such as assisting another officer that needs help or is injured, assisting in a chase or foot pursuit, assisting an off-duty officer, providing warrant service for an individual interacting with police, and responding to a fire or aiding a fire department response at their request. Further, I do not include calls to respond to a suspect operating a car that was planted or is being monitored by DPD ("bait" car or ETS activation responses). I also exclude calls that originate with complainants walking into a police station to alert police about an incident as well as 9-1-1 hang-up calls. Lastly, I exclude calls where an officer may have discovered an incident or complainant during patrol that was unlikely to be called in and dispatched to a larger set of officers. These dispatched call types are routine investigations, traffic stops, and public park checks.

Lastly, I exclude offenses that cannot be merged to a dispatch call record using the dispatch data (described below).

To clean the data, I conduct the following steps. I calculate the difference between the time a call is made and the time that the call is dispatched in minutes to form a call severity or call urgency variable, using time stamps in the data.

While there are codes for shifts or "watches" in the data, these often do not align with the general time slots for shifts. I construct a more strict and usable definition of shifts to eliminate shift overlap; these shifts are 12am-8am, 8am-4pm, and 4pm-12am. Next, I combine the remaining 119 dispatch codes into 22 groupings, to increase power and remove very small categories. Similarly, I combine the 34 location type codes in the data into 11 groupings.

Throughout the analysis, I use dispatch codes as incident controls rather than offense types, because dispatch codes are available to officers before they respond to incidents while incident offense types are designated by patrol officers after they arrive at the scene of an incident, and are therefore a choice variable.

Dispatch Data I supplement the offense data with records of all events dispatched by DPD during the sample period, obtained through an Open Records Request to the city of Dallas. This data includes address, police beat, sector and division, dispatch code, and time stamps for each call, dispatch assignment, response arrival and response conclusion. Like the offense data, the dispatch data includes records for events that were initiated by officers, such as officer "mark outs" of their location when responding to an incident identified through patrol, or dispatches to assist other officers. I am able to merge these records to offenses using offense incident ids included in the data.

I use this data in two ways. First, I exclude all offense records that cannot be merged to the dispatch data (about 10% of the offense sample after it is cleaned using other restrictions described above). This restriction assures that offenses in the sample originated through a call for service. Second, I use the dispatch data to determine the proportion of officers observed working on a shift that are available at the time of each offense in the sample. This availability rate is used to construct the "Low Availability" subsample used as a robustness check in the paper, or the sample of offenses with availability rates below the median in the offense sample.

Persons Involved and Arrest Data I use three DPD data files to supplement information on arrests and complainants in the main "Police Incidents" file: "Police Arrests", "Police Arrest Charges", and "Police Person". Additionally, I supplement these files with an open records request for DPD arrest records over the sample time period. These files offer different coverage of arrests than the "Police Incidents", and I use the differences across the files to create a liberal and comprehensive measure of arrests. The difference in coverage across the files may be related to when the data pulls were obtained and internal records updates. These files contain more detailed records of arrestees, including the time and place of arrest, arrestee name and demographics, and arrest charge. The "Police Person" data includes names and demographic information for arrestees and complainants associated with incidents.

The main outcome in this paper is whether any arrest occurred in association with a particular incident. I consider an arrest to have occurred if there is a record of an arrest in the "Police Incidents" file or any of the three supplementary arrests files.

I use complainant records in the "Police Person" to supplement the complainant records in the main incident file, and use these combined records to exclude records that do not have a civilian or business complainant. When there are multiple complainants, arrestees or suspects, associated with a call, demographic characteristics are measured as the maximum of these variables across the relevant group. For example, the indicator for a Black complainant is set to 1 if any of the complainants associated with the call are Black. I calculate whether there was a victim injury in a similar way, using information on complainants from both the "Police Person" and incident files.

Officer Demographics Data Lastly, I complement the data files available from DPD with officer information obtained through an open records request to the city of Dallas. Through this request, I acquired records of all police department employees from 2014 to the present that include officer names, badge number, job title, hire date, leave date (if applicable), ethnicity or race, gender, age, and salary. I match the request records to the incident file using officer badge numbers, and match officers by name if badge numbers are not available.

Because the officer request file includes employee title and badge number, I also use this information to exclude observations with responding officers that are civilian police employees, as these incidents likely involved only a phone response and did not entail a physical patrol officer response to the scene. I also exclude observations for individuals that are not employed by DPD (e.g. county police, firefighters, volunteer police officers).