Are "Bad" Cops Better Police?

The Illusory Tradeoff Between Officer Aggression and Public Safety

Roman Rivera*

Princeton University

September 24, 2024

[Preliminary. Do not cite or redistribute]

Abstract

This paper explores the trade-off between police officer behaviors and crime, focusing on the relationship between productive law enforcement, crime reduction, and costly actions (e.g., misconduct and use of force). I estimate the causal effects of individual Chicago Police officers on stops, arrests, use of force, misconduct, and crime. The results reveal little evidence that officers who engage in more costly actions are more productive and reduce crime, despite significant behavioral differences across officers. Furthermore, officer-level activities, such as stops and low-level arrests, have minor effects on crime. These results challenge the belief that aggressive policing improves public safety.

*For feedback, comments, and advice, I thank Bocar Ba, Patrick Kline, Dorothy Kronick, Raffaele Saggio, Jon Moreno-Medina, Emma Harrington, Brian Bell, Sam Norris, David Card, Steven Raphael, Conrad Miller, Steve Mello, Sandra Black, and the participants at the UC Berkeley Labor Seminar, UC Berkeley Goldman Research Seminar, University of British Columbia Economics Seminar, the Texas Economics of Crime Workshop, Princeton IR Section's Northeast Labor Symposium for Early Career Economists, and the Conference on the Economics of Crime and Justice. I thank the Institute for Research on Labor and Employment, UC Berkeley Center for Labor Economics, and Princeton University Industrial Relations Section for their support. This work received support from the Institute for Humane Studies under grant number IHS017836. Email: romgariv@gmail.com

1 Introduction

Is there a trade-off between a police officer's effectiveness in reducing crime, enforcing the law, and their tendancy to engage in costly behaviors, such as misconduct and excessive force? This question addresses a fundamental tension in law enforcement, where balancing the costs and benefits of policing remains central to both public debate and research. Media and public discourse often assume that more aggressive officers who engage in these costly activities achieve better results (Linnemann (2022), Denman (2023)). Similarly, much public debate and academic work assumes a positive relationship between effective law enforcement and aggressive policing (Corman & Mocan (2005), Mac Donald (2016)), biased officers (Prendergast (2007)), and misconduct (Shi (2009)). Recent research and policy have shifted focus to reducing these costly behaviors in policing (Owens et al. (2018), Dube et al. (2023)). If officers who engage in costly policing behaviors actually are the most effective at enforcing the law and reducing crime, then policies aimed at curtailing these behaviors could reduce officer productivity and increase crime. However, the actual relationship between these dimensions of policing remains uncertain.

This paper studies the potential trade-off between an officer's ability to prevent crime, their productivity in enforcing the law, and their tendency to engage in overly aggressive policing behaviors, such as the use force, misconduct, and low-level arrests. I begin by estimating the effect of individual Chicago police officers on multiple outcomes, including stops, arrests, use of force, misconduct complaints, and crime, and I find significant amounts of variation across officers. Then, I assess the degree to which officers who cause more costly outcomes are also more productive in making arrests for serious crimes and conducting stops by estimating covariances across outcomes. I find relatively weak correlations between these productive activities and the costly behaviors, suggesting the absence of a substantial productivity-cost trade-off. Next, I estimate the effect of officer activities on crime rates and find small effects, indicating that more aggressive and socially costly officers are not better at preventing crime, inconsistent with a large crime-cost trade-off. I also find small elasticities

between police activity (stops, low-level arrests) and property and violent crime. Finally, I explore the relationship between officer characteristics and their effects on activities and crime-reduction, finding that demographics tend to have a weak relationship with crime-reduction despite very a strong relationship with officer activity.

To estimate the causal effects of individual officers, I employ an extensive data set of millions of officer-shift records from the Chicago Police Department between 2009 and 2019, covering their daily assignments and start and end times. I combine this data with records on almost every officer-level outcomes ('activity') studied in prior work, including street and traffic stops, arrests, use of force, complaints of misconduct, and court outcomes of arrests. With this data, I extend the canonical worker-firm model of Abowd et al. (1999) to separate the influence of the officer from the influence of their assignments — when, where, and in what role the officer is working — using officers' movement between assignments for identification. I then quantify the degree of officer heterogeneity within and across these outcomes as the (co)variance in officers' causal effects. In contrast with prior work on officer effects on officer activities using Empirical Bayes (Rozema & Schanzenbach (2019), Adger et al. (2022), Weisburst (2022)), I apply Kline et al. (2020)'s heteroskedasticity-robust bias correction to variance and covariance components which does not involve biasing officer effects or a reliance distributional assumptions, and allows for bias-corrections on cross-outcome covariances (Lachowska et al. (2023b)).

I find that differences across officers have substantial influence in determining the officer-level activity metrics studied. One standard deviation (SD) in officer effects is generally close to or exceeds 75% of the outcome's mean for stops and arrests, equating to around 33 street stops, 16 traffic stops, 10 nonindex (lower-level) arrests, and 1.2 index (serious) arrests over 100 shifts (around half an officer-year). For use of force and misconduct, 1 SD in officer effects equates to close to or more than 100% of the mean, or around 0.4 uses of force and between 0.12 and 0.23 complaints per 100 shifts. Notably, officer effects explain more variation than granular assignment effects for all outcomes other than index arrests which

are less discretionary. Despite the substantial officer heterogeneity, I find little evidence for a significant trade-off between an officer being more productive (more serious arrests, stops) and engaging in higher cost activities (use of force, misconduct). Cross-outcome correlations tend to be weak. For example, the correlation between officer effects on complaints and serious arrests tend to be < 0.08, complaints and stops < 0.03, while the correlation between use of force and stops and use of force and serious arrests tend to be < 0.17 and < 0.3. Even correlations between low-level arrest effects and complaints (< 0.24) and use of force (< 0.38) effects are modest, though larger than others. Furthermore, I do not find evidence that more costly officers are making higher quality arrests, measured by court outcomes of their arrests (Weisburst (2022), Rivera (2022)), conditional on their arrest quantity. Overall, while officer heterogeneity is large for activity metrics, the evidence does not support the existence of a significant productivity-cost trade-off.

To measure the effect of individual officers on crime, I construct 'teams' of officers working in the same police district in the same day-hour and link it with crime data at the district-hour level. Building on Bonhomme (2021), I model a team's influence on crime as the mean of its members' individuals' effects on crime. I leverage variation in which officers are working together on a team due to when individual officers start and stop working during their shifts and who is working on a given day based on the CPD's rotation calendar, to identify individual officers' effects on crime and employ both fixed effects and first difference estimators. Holding team sized fixed, the officer's causal effect on crime amounts to officer quality in crime-reduction, in contrast with the well-documented effects of officer quantity on crime (Chalfin & McCrary (2017)).[1] I find that team composition has a non-negligible influence on violent and property crimes, as a 1 SD increase in a team's average crime reduction effect decreases crime by 16-50% of the respective means. These effects are comparable in magnitude to the effects of officer quantity from Chalfin & McCrary (2018) and Mello (2019).

[1]See also, Levitt (2002), Di Tella & Schargrodsky (2004), Klick & Tabarrok (2005), Evans & Owens (2007), Draca et al. (2011), Donohue et al. (2013), Weisburst (2019),Mastrobuoni (2019), Chalfin & McCrary (2018), Cheng & Long (2018), Mello (2019), Cox & Cunningham (2021), Chalfin et al. (2022).

Furthermore, team membership also influences the frequency of 911 calls, suggesting these effects are measuring underlying influence on crime, with additional results most consistent with officers being differentially effective at deterrence.

Given the significance of officer heterogeneity for both policing activity and crime, I estimate the effect of officers' activity propensities on crime. I regress crime on a team's average propensity for each activity and control for highly granular unit-time fixed effects, exploiting differences in team composition within the same unit, year-month, day of week, and hour. To overcome any concerns of endogeneity, I instrument for actual team composition with predicted team composition based on their rotation calendar which is determined in the prior year, extending the analysis of Gudgeon et al. (2023). Despite a strong first stage, I do not find that teams with higher average complaint effects or force effects reduce violent or property crime. Indeed, effect of most officer activities crimes tends to be small. By instrumenting the number of stops and arrests occurring within a district-hour with predicted team propensity, I estimate the elasticity of crime with respect to these policing activities. Elasticities tend to be very small and sufficiently precise to rule out elasticities of violent and property crime with respect to stops larger than -0.15 and -0.08 and with respect to nonindex arrests larger than -0.37 and -0.1, respectively, which are smaller than the most of the elasticities of officer quantity on crime.

Finally, I document the relationship between a broad set of observable officer characteristics and officer effects on activities and crime reduction.[2] By projecting estimated officer effects onto officer characteristics, I confirm some results from prior work (Ba, Knox, et al. (2021), Goncalves & Mello (2021), Hoekstra & Sloan (2022), Gudgeon et al. (2023)), such as that Black and female officers make fewer stops, arrests, and uses of force, and I also contribute many new benchmark estimates. For example: Black and female officers receive fewer complaints; officer education does not have a significant impact on most outcomes; and officers who have had active service in the military are more active (stops and arrests) but

[2]These include officer race, gender, starting age, tenure, education level, military experience, and language ability.

do not receive more complaints. While most officer characteristics have a strong relationship with activity, they tend to have relatively small effects on crime-reduction. Officer tenure has large negative effects on nearly every dimension of officer activity, and estimates suggest that while high tenure officers ($>$ 15 years of experience) may reduce violent crime, they also allow for more property crimes, patterns consistent with the structural estimates of Ba, Bayer, et al. (2021).

This paper is closely related to the literatures on measuring the importance of heterogeneity across agents in labor economics. I apply canonical Abowd et al. (1999) model to the setting of police officers to quantify officer heterogeneity across almost all previously studied dimensions of officer behavior, building on work using the AKM model to study worker and firm effects on wages and inequality (Card et al. (2013), Bonhomme et al. (2022)) and bureaucratic productivity (Best et al. (2023)). In a similar vein, other work studies the role of heterogeneity in teacher 'value add' on various outcomes (Kane & Staiger (2008), Chetty et al. (2014), Jackson (2018), Rose et al. (2022)). I estimate and cross-correlate measures for officer 'value-add' on multiple behavioral dimensions, extending prior work on the causal effects of individual officers on smaller sets of outcomes (Rozema & Schanzenbach (2019), Weisburst (2022), Adger et al. (2022)), and apply improved cross-outcome correlations and bias-correction. These measures reflect officers' discretionary behaviors and thereby contribute to the literatures on both officer motivation, skill, and discretion (Mas (2006), Feigenberg & Miller (2022)).[3] This paper also builds on recent work on team production (Jaravel et al. (2018), Chan (2021), Y. Chen (2021), Bonhomme (2021)) by modeling crime prevention as a function of team composition, allowing me to recover a new dimension of officer heterogeneity — their influence on crime.

This paper advances our understanding of the costs and benefits of policing. Existing work documents the direct and indirect consequences of more aggressive policing (Mello (2021), Ang (2021), Ang & Tebes (2023), Ang et al. (2023), Agan et al. (2023)), while

[3]See also Mastrorocco & Ornaghi (2020),Chalfin & Goncalves (2021), Stashko (2022), Gonçalves & Mello (2023), and Grosjean et al. (2023).

other work studies the effects of procedural justice and cognitive trainings (Owens et al. (2018), Dube et al. (2023)), sustaining complaints (Rozema & Schanzenbach (2023)), and body-worn cameras (Lum et al. (2020), Kim (2023)) on specific officer behaviors. I find that there is little evidence for a trade-off between reducing costly policing activities and reducing productivity or allowing more crime. As a result, this paper suggests that interventions aimed at reducing aggressive behaviors in policing may not significantly diminish overall policing effectiveness.

Finally, this paper also contributes to the literature connecting policing activity and crime. Existing work documents how crime and policing outcomes at the departmental level change in response to oversight or public outrage (Prendergast (2001), Shi (2009)),[4] compensation disputes (Mas (2006)), media attention (Mastrorocco & Ornaghi (2020), Rackstraw (2023)), traffic stop interventions (Parker et al. (2024)), and police force size (Chalfin et al. (2022)), but, in general, they do recover the direct effect of policing activity on crime. By leveraging officer-level differences across policing activity, I am able to directly estimate the elasticity of crime with respect to stops and arrests, and I am generally able to rule out elasticities smaller than those of officer quantity (Chalfin & McCrary (2018), Mello (2019)). This is similar to Cho et al. (2023), which infers elasticities using arrest declines following officer deaths, whereas I estimate the direct impact of a wider range of officer activities on crime.

The paper proceeds as follows. Section 2 discusses the setting, background, and data. Section 3 introduces the main econometric frameworks for recovering officer effects on officer-level outcomes and crime. Section 4 presents variance estimates for officer-level outcomes and crime. Section 5 explores the relationship between officer activity and crime. Section 6 connects officer effects to officer characteristics. Section 7 concludes.

[4]See also Devi & Fryer (2020), Rivera & Ba (2023), Premkumar (2022), Mikdash & Zaiour (2023), and Campbell (2024).

2 Setting and Data

The goal of this paper is study is to measure the relationship between individual officers and their effects on both directly observed outcomes (e.g., arrests, uses of force) and their effects on crime rates where and when they work. For this, I use an extensive dataset on the Chicago Police Department (CPD), the 2nd largest police department in the U.S. in a major metropolitan city, which polices a large and diverse population of around 2.7 million people, between 2009 and 2019. The size of the CPD, the breadth of data on policing and crime in Chicago, and a variety of institutional features make this a ideal setting.

The primary data for this study (officer rosters, demographics, daily shifts, arrests, stops, use of force, complaints, salary, unit history, education, language ability, military experience) were obtained via Freedom of Information Act (FOIA) requests by the author or were obtained from or in collaboration with the Invisible Institute, or from the Invisible Institute's public data repository (see here).

2.1 Background on CPD

There are over 10,000 sworn officers in the CPD at a given time, and they perform a variety of functions, with many hundreds working across the city at any given time. In this paper, I focus on police officers, which are the lowest and most common (80%) type of sworn member of the CPD, as opposed to higher ranks such as detectives, sergeants, or lieutenants.

The main organizational structure within the CPD is the 'unit'. Each unit performs different functions, ranging from general policing of a specific area ('patrol units') to detective units, SWAT units, and training units. Officers rarely work in units other than their main assigned unit within a given year. The most common units are patrol units or 'geographic' units which map to a specific police district — a large area of the city (around 10 square miles) similar to precincts in other cities — and officers are generally responsible for patrolling, answering 911 calls, and serving on special and tactical teams within that district. For

example, unit 14 corresponds to being a patrol officer in district 14 — and this is true for all units numbered 1 - 25. During this study, Chicago is divided into 22 (2013-2019) or 25 (2009-2012) police districts and corresponding patrol units (see Figure A.1 for a map of CPD districts and neighborhoods from Markovic (2009)).[5] Other units can map to more generalized areas encompassing multiple districts (e.g., detective or gang enforcement units) or no specific location (e.g., legal affairs or training).[6]

This study focuses on a sample of 54 units in which officers are sufficiently active, in that they engage frequently in policing activity (stops, arrests, etc.), and Figure A.3 displays these units by name and number along with the number of officer-shift (day) observations per unit in the sample. Other than the patrol units, the non-patrol units in the sample include the Narcotics Division, Gang Enforcement, and the Traffic Section, but exclude units such as legal affairs and the police academy (among others).

Next, an officer is assigned to a regular 'watch' which generally corresponds to when they work, with the vast majority of observations belong to watches 1-4 (see Figure A.3), and most shifts lasting 9 hours. Watches 1 (starting around 10 pm), 2 (starting around 7 am), and 3 (starting around 2pm) correspond to the regular shifts. Watch 4 is for tactical team officers, as studied by Kim (2024), and I subdivide Watch 4 into 2 groups 4-morning and 4-night based on when the officer starts their shift, and these are similarly common though half as common as Watch 1. A small share of observations correspond to other watches, which I pool together. Together, an officer's unit and watch assignment determines where, when, and what function they serve when they are working.

Finally, the days an officer is actually working is determined by their day off group (DOG). DOGs determine the cycling of present for duty officers working in a unit during a

[5]The 2009 district map changed slightly throughout 2012, during which three units, 13, 21, and 23, were absorbed by larger districts, primarily, 12, 2, and 19, respectively. Districts are subdivided into 3 to 5 sectors which are subdivided into geographic beats (usually 4, each around 1 square mile).

[6]Switches between units are not very frequent on a day-to-day basis, with most officers not moving more than 2 or 3 times over their careers, and the bidding process is mainly based on tenure and discussed more in Ba, Bayer, et al. (2021). When a new officer enters the department, they are initially assigned to units based on departmental demand and openings, and they can bid for vacancies in different units as openings appear and before the start of the calendar year.

specific watch over the course of the year. Most officers work on a 4-2 rotation (4 days on duty followed by 2 days off).[7] Importantly, DOGs are predetermined, and officers cannot easily or systematically manipulate days on duty. Furthermore, because of overlapping DOGs, the groups of officers working will change from day to day.[8]

Within a given year, an officer has a stable unit, watch, and day off group (DOG), which are determined at the end of the prior year. Changes to an officer's stable unit/watch/DOG occur through a bidding process, the outcome of which is mainly based on tenure. CPD officers are protected against major violations of their contracts by a strong union (see Ba, Bayer, et al. (2021) for more discussion).

When an officer is on duty, they are assigned to a specific 'beat code', which is an additional level of specificity for their role and/or specific location on that day beyond their general role, location, and time as captured by unit-watch. For example, an officer may be a 'beat' officer, meaning they patrol a specific geographic beat (around 1 square mile area within a district);[9] or they may be a rapid response officer, responding to high-priority 911 calls within a specific sector within a district; or they may be a bike officer, special mission officer, school officer, or assigned to desk duty. Officers have some influence on which beat they work within their unit and watch. However, they generally work the same or a small set of beats on a day-to-day basis, and the exact assignment of officers to beats is determined by upper commanders within the unit and watch (e.g., sergeants and lieutenants). Because there are tens of thousands of exact beat codes in the CPD, I classify similar beat codes into beat categories, which are displayed in Figure A.3.

[7] In 2009 and before, officers generally worked on 6-2 rotations (6 days on followed by 2 days off), this was switched to shorter rotations (most commonly 4-2) in 2010 (JG0304 (2009)).

[8] For example, DOGs A, B, and C may be active on one day while DOGs B, C, and D will be active the next day, and DOGs C, D, and A will be active the day after next. See A.2 for an example of the CPD's operations calendar.

[9] Assigned beats can map directly to geographic beats, but this is not the norm. Instead, an assigned beat denotes a general area and a role.

2.2 Data

The primary data for this study is daily assignment and attendance records for all units in the CPD between 2009 and 2019. This data contains officer-by-shift information, where a shift is a working spell by the officer generally for around 9 hours. This data indicates whether the officer was or was not present for duty, the reason for their absence (e.g., it was their regular day off, a vacation day, they were in training), their start and end times, their title (e.g., police officer, sergeant, detective), and their assigned beat, watch, and unit, among other information.

I take the subset of assignment data observations (at the officer-shift level, where officers work at most a single shift per day), where officers are present for duty with a known assignments, are the lowest and most common rank of "police officer" (around 80% of the CPD), on regular assignments in which officers are regularly active (e.g., not in training or desk duty), along with a few other restrictions. These filters result in 13,928,210 officer-shift observations, with 13,400 unique officers, in 54 units (25 of which are patrol units that correspond to districts). As shown in Column (1) of the officer-shift level summary statistics in Table 1, within a single year, the average officer works 182 in-sample shifts, given they work at least 1 day and officers work in an average of 15.4 unique assignments (GABs), with 45.7% of an officer's shifts being in their most common assignment. I link the assignment data with officer demographic data (age, race, gender, tenure, education, military experience), which are also summarized in Panel A of Table 1 and discussed in Appendix E.

To measure the many dimensions of officer-level activity, I collected and linked data on nearly all possible officer-level activity dimensions considered in policy and academic research (stops, arrests, use of force, and misconduct) to the assignment data. Panel D of Table 1 summarizes the means of the primary outcomes per 100 shifts. First, stops are classified as either 'street' stops or 'traffic' stops. Both types of stops can be considered a metric for contact with suspects and non-suspects alike, and they are more frequent than other outcomes averaging 35 per 100 shifts and 13 per 100 shifts for street and traffic stops,

respectively. Notably, in 2016 major reforms resulted in a significant decline in street stops and a rise in traffic stops (Hausman & Kronick (2023)).

Next, I link data on officers' arrests from 2009 to 2019. Arrests are the primary metric for enforcement are and often used to measure both departmental and officer productivity. Arrest data include both arrestee information, crime types, and arresting officer type (primary, secondary, and auxiliary/supporting). In the main results, I credit an officer for an arrest if they are listed as a primary or secondary arresting officer, and I divide arrests into two general categories: arrests for index crimes, 'index arrests', (violent: murder, robbery, rape, aggravated assault; property: burglary, motor vehicle theft, larceny/theft, arson) and arrests for non-index crimes, 'nonindex arrests', (e.g., drugs, prostitution, simple battery, vandalism, municipal code violations, traffic offenses). Nonindex arrests are considered more discretionary, often do not have a victim, and are more common at 11 per 100 shifts, where as all index crimes have a victim, are considered less discretionary, and are less common at around 1.6 per 100 shifts.

Use of force and misconduct are highly costly activities that officers also engage in. I collect uses of force by officers from 2009 to 2019 which contain all instances when officers report using force on a subject, ranging from relatively minor force (e.g., pressure points, defensive techniques) to gunshots and tasers (which are relatively rare). Uses of force occur at around 0.36 per 100 shifts.

I measure misconduct using complaint data which contains complaints filed against police officers by either civilians or fellow officers — though civilian complaints comprise the vast majority of complaints ($> 85\%$). The data on complaints is sourced from two different oversight regimes. The complaint data which covers 2009 to 2015 was obtained in November of 2016 by the Invisible Institute, which I refer to as 'Complaints [2016]', and occur at a mean of 0.24 per 100 shifts between 2009 and 2015. I obtained an extended data set on complaints ('Complaints [2019]') from the CPD which contains the full breadth of the data. The 2019 data include fewer complaints overall than the 2016 data, with a mean of 0.096 per 100 shifts.

Finally, to link officers with reported crime, I construct 'teams' of police officers who are working together within the same location (district) at the same time (hour). The unit of observation is the district-hour (e.g., district 12 on June 15th, 2013, at 2pm), and I allocate officers to 'teams' (the set of officers working in a specific district at a specific hour) based on the start and end times of their shifts when they are present for duty in a specific district (units 1-25), which is recorded in the assignment data. For team composition, I only count officers working as police officers (the focus of the prior sample) among other filters (see Appendix E), though I do record the number of other ranks active by unit-hour to use as controls (e.g., number of detectives). In the analysis sample, there are 2,032,449 teams (district-hour observations), 11,686 unique officers, and the average officer being observed on 8,485.699 unique teams. The average team contains 48.8 officers, however the size of teams can vary significantly within the same district over the course of the day as shown in Figure B.3. Summary statistics for the teams data are displayed in Table 2.

Reported crimes (either cleared by an arrest or uncleared) are collected from Chicago's public data portal. I aggregate crimes at the district-hour level such that a specific police team experiences a specific number of reported crimes. Table 3 displays the average crimes per district-hour, clearance rates, and summary statistics for the crime data. As with arrests, crimes are categorized as either index or nonindex, with each being broken into smaller categories. Notably, index property crime (motor vehicle theft, theft, burglary, arson) are rarely cleared at only 9% but are fairly common at almost 0.5 per district-hour, index violent (murder, robbery, rape, and aggravated assault) occur at a rate of 0.14 per district-hour and cleared 17%, and nonindex drug offenses are driven by discretionary arrests as their clearance rate is almost 100% and are fairly common at 0.15 per district-hour.

3 Econometric Framework for Identifying Officer Heterogeneity

This study aims to quantify the importance of differences across individual officers in influencing both officer-level outcomes and crime rates and the relationship between them.

I employ two separate approaches because unlike use of force or an arrest, which can be directly attributed to an officer, a crime cannot be directly linked to an individual officer, rather a crime can be assigned to a location and time during which the officer is working.

3.1 Officer-Level Activity

Officer-level outcomes include any metric of performance or activity that can be attributed to an individual officer during their shift, such as contacts with civilians and suspects (street and traffic stops), enforcement of laws (arrests), and undesirable and socially costly outcomes (use of force and misconduct). I denote an officer i's 'type' or effect on officer-level outcomes with the vector θ_i. Elements of θ_i, θ_i^k, capture the individual influence on outcome type k when on duty regardless of other factors, such as where, when, or in what role they are working. For example, θ_i^{Arrest} is the officer's influence on arrests made on that day compared to some reference officer.

I employ an Abowd et al. (1999) -style (AKM) framework to disentangle the influence of the individual officer from their where, when, and in what role they are working (their 'assignment'), time varying factors (such as the exact date they are working), and unobserved factors influencing their outcomes. Given officer i working in assignment J on day t, the outcome $y_{i,t}^k$ (e.g., number of index arrests made by i) for officer i on date t is the sum of officer-specific effect θ_i^k, assignment-specific effect $\psi_{J(i,t)}^k$, time-varying observable factors $\beta^k x_{i,t}$, and unobserved factors $\epsilon_{i,t}^k$:

$$y_{i,t}^k = \theta_i^k + \psi_{J(i,t)}^k + \beta^k x_{i,t} + \epsilon_{i,t}^k \tag{1}$$

θ_i^k is all of the officer-specific factors that influence the relevant outcome, such as their personal biases and beliefs, friendliness, general aggression, how civilians see the officer, their intrinsic motivation, and importantly their individual 'discretion'. $\psi_{J(i,t)}^k$ is all assignment-specific factors which influence outcomes, such as the mission of the job (answering 911 calls,

or actively patrolling an area on foot, a tactical team, or traffic enforcement), the time of day they are working, and the composition of civilians and criminal activity. Time varying factors, such as date effects may also influence outcomes separate from the fixed influence of officers or assignments.

Identifying variation for the parameters in this model,$\theta^k_{1,...,N}, \psi^k_{1,...,J}, \beta^k$, come from officers working in different assignments over time, and identification requires the following mean independence condition is satisfied:

$$E(\epsilon^k_{i,t}|x_{1,1}...x_{N,T}, J(1,1)...J(N,T), \theta^k_1...\theta^k_N, \psi^k_1...\psi^k_J) = 0 \tag{2}$$

that conditional on officers, assignments, and time-varying observables, the unobserved influences ('shocks') on outcomes, $\epsilon^k_{i,t}$, must be independent of past and future assignments and shocks. This allows for officers to sort into assignments based on their fixed characteristics, but does not allow officers to choose assignments based on shocks, and officers cannot choose when to be on duty based on shocks (as it would be analogous to selecting the 'off duty' assignment based on $\epsilon^k_{i,t}$). I probe the plausibility of these assumptions in Section 3.3.

The main specification for recovering officer and assignment effects in equation (1) with officer effects and GAB (assignment) effects uses date-specific fixed effects in $x_{i,t}$. Because of the breadth of time periods and regimes which are captured in the data, I follow Card et al. (2013) and Lachowska et al. (2023a) by estimating equation (1) for rolling time periods (p): 2009 - 2010, 2010 - 2011, ..., 2017 - 2018, 2018 - 2019. This allows us to ensure that officer effects are permitted to evolve over time (e.g., as they gain tenure and experience) and assignment effects can change as the CPD's structure shifts over time. A number of alternative specifications for robustness are discussed in Appendix D.4. Additionally, the ability for officers to extend their shifts using overtime poses some issues, but I do not find much evidence for significant concerns (see Appendix D.4.3).

3.2 Crime and Team Outcomes

In contrast with officer-level outcomes, an officer's influence on crime is more difficult to measure because a crime occurring (or not) cannot be assigned to an individual officer. While crimes occur at in a specific area at a specific time, officers are not generally assigned to specific locations and the total police force in an area work as a team to reduce crime, so measuring the effect of a specific officer on the occurrence of crime is empirically difficult. To overcome these challenges, I model how the number of crime occurring within a police district at a specific time (day-hour) are influenced by the team of officers working in that district at that time. Despite not being directly observed, the influence of an individual officer on the crime rates where they work is evidenced by annecdotes by police (Moskos (2008)).[10]

The number of crimes y of some type c (e.g., the number of uncleared robberies) which occur in district d in time (hour) h as the effect of the team τ of officers working in district d in time h, $\Omega^c_{\tau(d,h)}$, the environment-specific effects, $\gamma^c_{w(d,h)}$, and additional observable factors $X_{d,h}$, and unobserved factors $\epsilon^c_{d,h}$:

$$y^c_{d,h} = \Omega^c_{\tau(d,h)} + \beta^c X_{d,h} + \gamma^c_{w(d,h)} + \epsilon^c_{d,h} \tag{3}$$

The team's effect on crime and other aggregate outcomes, $\Omega^c_{\tau(d,h)}$, is a function of the individual officers who are working together at that time. While it is possible to model the team production (or prevention) of crime to be a highly non-linear function, involving individual officers, peer effects, and hierarchical features, I make the simplifying assumption that a team's effect on crime is an additive function of its composite members, similar to Bonhomme (2021). Specifically, I model $\Omega^c_{\tau(d,h)}$ as an additive production function:

$$\Omega^c_{\tau(d,h)} = \frac{1}{N_{\tau(d,h)}} \sum_{i \in \tau(d,h)}^{N_{\tau(d,h)}} \omega^c_i \tag{4}$$

[10] In a letter to his superior, Baltimore police officer Moskos states: "I believe the effectivness [sic] of my work can best be judged by the low level of crime on my posts during my working hours" (Moskos (2008), page 156).

where the effect of a team on the y^c is the sum of its members' individual effects on the outcome (ω_i^c) divided by the number of officers on the team ($N_{\tau(d,h)}$). ω_i^c is the influence of officer i on y^c relative to some reference officer if they were both working alone in the same environment. $\gamma^c_{w(d,h)}$ captures fixed assignment-time factors (e.g., the fixed influence of district 2 at night on Tuesdays in December of 2013), which capture both where, when, and in what time of year the team is working. $X_{d,h}$ controls for time varying factors, such as polynomials of total team size, total officers working, the number of sergeants on duty, and date-specific fixed effects.[11]

Similar to the model in equation (1), the identifying assumption for equation (3) is that unobserved shocks to crime rates are conditionally independent of observable factors $E[\epsilon^c|X, \Omega^c, \gamma^c] = 0$. The central concern is that officers strategically select when to work based on shocks to underlying crime rates within their district on a given day. However, institutional features significantly reduces the likelihood of such systematic selection, as discussed in Appendix B along with other concerns.

The main specification for estimating this fixed effect ('FE') specification in equation (3) controls for assignment-time effects with $\gamma_{w(d,h)}$ being the interaction between district, year, quarter (or month), day of week, and 8-hour time window (3 in the day). As a result, γ controls for sorting across primary assignments (unit and watch) as well as day-of-week interacted with the month (or quarter), meaning the identifying variation comes from differences in which officers work on different Tuesdays, in unit 10, between 10pm and 6am, in the month of July 2016 (or quarter 2 of 2016). In $X_{d,h}$ for both the fixed effects and first difference designs, I include second degree polynomials of the number of officers in the team, date-specific fixed effects, and second degree polynomials of other policing counts (number of officers working

[11] In matrix form, combining equation (3) and equation (4) becomes:

$$Y^c = B\omega^c + G\gamma^c + X\beta^c + \epsilon^c \tag{5}$$

with N district-day-hour observations and K officers, B is a $N \times K$ matrix with each row comprised of indicators for whether an officer was working scaled by the $1/N_{d,h}$, such that $B\omega^c = \Omega^c$, with $\omega^c = \{\omega_i^c\}_{\forall i/\ i^0}$'s being identified relative to an excluded reference officer. G is a matrix of indicators with one per row and γ contains the assignment-time fixed effects.

on tactical teams, working overtime on specific teams, working as desk officers, detectives, sergeants, detectives, etc.) — these officers (e.g., probationary officers, detectives, desk officers, overtime task forces) are excluded from the team composition. To account for the changes in the CPD and Chicago over the sample period, while also taking into account data limitations, I split the data into three multi-year periods — 2010-2012, 2013-2015, and 2016-2019.[12]

Finally, I employ an additional estimation strategy by leveraging variation solely in the team composition produced by officers starting and ending their shifts estimated using a first difference ('FD') estimator. This removes the need for inclusion of unit-time fixed effects and simplifies estimation at the cost of only using the variation produced in officers' first and last hours of work for identification:

$$\Delta y^{c}_{d;h,h-1} = \Delta \Omega^{c}_{\tau(d;h,h-1)} + \beta^{c} \Delta X_{d;h,h-1} + \Delta \epsilon^{c}_{d;h,h-1} \tag{6}$$

This removes unit-time fixed effects and it alleviates concerns over the $\gamma_{w(d,h)}$ not being granular enough. Additionally, it simplifies estimation at the cost of only using the variation produced in officers' first and last hours of work for identification. For additional robustness, I use the predicted composition of the team based on the rotation calendar and their predicted start/end times in combination with the first difference estimation, analogous to a reduced form relationship between predicted team composition and changes in crime.

3.3 Internal Validity

Given the data and setting of this paper, identification of the parameters of interest require multiple conditions and assumptions to be satisfied. I discuss these conditions briefly in this section, while Appendix B discusses them in more detail.

First, there must be sufficient identifying variation such that estimates of the parameters

[12] I exclude 2009 in these main estimates because of the change in rotations after 2009 and that officers' unit-watch assignments are more stable at the year level.

can even be recovered — see Appendix B.1 for additional discussion. For the officer and assignment effects for officer-level outcomes, this condition is satisfied by focusing on the largest connected set of officers and assignments (Abowd et al. (2002)). To reduce concerns over limited mobility bias (Andrews et al. (2008), Jochmans & Weidner (2019), Bonhomme et al. (2022)), I follow the current literature pruning the network to find the largest connected component that is robust to the removal of any single officer, as in Kline et al. (2020) and Bonhomme et al. (2022), which I refer to as the "LLC". As shown in Table B.1, in practice, the LLC contains the vast majority of observations ($> 95\%$).

Similarly, the team model also require that officers and assignment-time effects are connected and that officer effects can be identified separately from one another (e.g., if two officers always work together, then their individual ω_i's cannot be identified). I employ the pruning algorithm of Bonhomme (2021), and it removes a small share of observations and officer/district-time indicators for FE and changes in officer indicators for FD (see Table B.2), with generally over 98% of observations remaining in the largest identified set, indicating that the degree of variation in team composition is sufficiently large.[13]

Second, given that estimates can actually be recovered, the primary threat to identification of causal effects of officers and assignments is that officers are endogenously selecting which days they work based, such that, for example, officers who prevent more crime and make more arrests (all else equal) also choose to work on days when crime and arrest opportunities are higher. Fortunately, the days an officer is going to work in a given year is determined at the end of the prior year by their selection of a day off group (DOG) and their fixed rotation (e.g., 4 days on, 2 days off). This prevents officers from systematically choosing which days to work or not based on (unobserved to the econometrician) shocks to crime rates or other factors.

To directly test the strength of the operations calendar, I follow Gudgeon et al. (2023),

[13]Bonhomme (2021) provides a pruning algorithm to iteratively remove indicators (e.g, officers) for whom effects cannot be identified, then removing any team in which they are on, until convergence. To ensure non-singularity for all fixed effects in FE, I prune the matrix containing officer indicators and district-time fixed effects to ensure both ω and γ are identified.

and Appendix B.2 contains more details and additional tests. Each year, I construct a predicted rotation for each officer based on the first set of observations, which follow a 4-2 (or, in a minority of cases, 5-2) rotation. Then I project this predicted rotation backward (if the first 4-2 is observed in the middle of January, for example) and forward throughout the rest of the year. This prediction is when the officer 'should' be working and 'should' be off based on a crude usage of the rotation system, ignoring all vacation days, injuries, training, potential for switching units, or changing DOGs. Nevertheless, as shown in Figure 1, which displays the share of correctly predicted days in each week of the year, this crude prediction is quite accurate at around 85% correctly predicted early on which decreases to 70% by the end of the year. For reference, if officers were randomly working 4 out of 6 days, the share of correct predictions would be less than 50%. This consistency in predictive power of even a crude rotation is evidence of the binding nature of the calendar and the difficulty in officers deviating, even if they could observe shocks to local crime conditions on a day-to-day basis.

Third, the final major threat to identification is that movements of officers between assignments is "endogenous mobility" (Card et al. (2013)). In this environment, endogenous mobility would be evidenced by officers may be sorting into assignments based on match-specific values, that match-specific components 'drift' as officers learn about their assignments over time, or that officers are choosing when to switch based on unobserved shocks. I test for match-specific components and drift effects using the traditional event study plots (Figure 2) following Card et al. (2013), and do not find patterns consistent with these concerns, and I test for shock-based movements in Figure B.6, which provides little evidence of outcomes one day increasing the likelihood of switching assignments the next day.[14] See Appendix B.3 for more details.

[14]I include multiple additional specifications to probe the robustness of my results to these concerns, such as sampling only officer-assignment spells which last for at least 4 days, sampling only says officers are predicted to work based on the calendar, and dropping the first and last day of a officer-assignment spell, among others. Furthermore, Appendix D.3 compares primary results to the use of complete match effects model with interacted officer and GAB effects.

3.4 Variance Components

For officer-level and team-level outcomes, the relative influence of different officers, or the importance of officer heterogeneity, in producing different policing and crime outcomes can be measured by $Var(\theta^k)$ and $Var(\Omega^k)$, respectively, and the total variation in the outcome (after low-dimensional controls are partialled out, e.g., $\tilde{Y}^c = Y^c - X\beta^c$) can be decomposed into:

$$\begin{aligned} Var(\tilde{y}^k_{i,t}) &= Var(\theta^k_i) + Var(\psi^k_{J(i,t)}) + 2Cov(\theta^k_i, \psi^k_{J(i,t)}) + Var(\epsilon^k_i) \\ Var(\tilde{y}^c_{d,h}) &= Var(\Omega^c_{\tau(d,h)}) + Var(\gamma^c_{w(d,h)}) + 2Cov(\Omega^c_{\tau(d,h)}, \gamma^c_{w(d,h)}) + Var(\epsilon^c_{d,h}) \\ Var(\Delta\tilde{y}^c_{d;h,h-1}) &= Var(\Delta\Omega^c_{\tau(d;h,h-1)}) + Var(\Delta X_{d;h,h-1}\beta^c) + Var(\Delta\epsilon^c_{d;h,h-1}) \end{aligned} \tag{7}$$

However, even the estimates $\hat{\theta}^k_i$ and $\hat{\omega}^k_i$ are unbiased, $Var(\hat{\theta}^k)$ and $Var(\hat{\Omega}^k)$ ('plug-in' variance estimates) are biased upwards due to measurement error (Krueger & Summers (1988), Kline et al. (2020)). This will cause officer heterogeneity to be overstated. Similarly, plug-in variances of assignment effects will also be biased upwards, and covariances between officer and assignment effects (e.g., $Cov(\theta^k_i, \psi^k_{J(i,t)})$) and cross-outcome covariances (e.g., $Cov(\Omega^{Violent}_\tau, \Omega^{Property}_\tau)$) are biased downwards. This will understate the degree to which officer effects on different outcomes are related, so it can understate the trade-off between, for example, misconduct and productivity if $0 > Cov(\theta^{Arrest}_i, \theta^{Misconduct}_i) > Cov(\hat{\theta}^{Arrest}_i, \hat{\theta}^{Misconduct}_i)$. Furthermore, plug-in correlations will be biased in an indeterminate direction (attenuated by the denominator and biased downward by the numerator).

I bias-correct all variance and covariance components using Kline et al. (2020)'s (KSS) heteroskedastic-consistent bias correction. I apply the Lachowska et al. (2023b) extension of KSS to construct bias-corrected estimates of cross-outcome covariances.[15] Additionally, while

[15]The leverages for the KSS correction for the officer-level analyses and alternative specifications use match-level (data aggregated to one row per officer-assignment after partialling out time-varying controls with proper weights) to avoid concerns over serial correlation in longer panels; non-match-level results are included in the robustness figures.

Bonhomme (2021) proposes a homoskedasticity robust correction and utilizes team-specific scaling factors in the team production model, in this environment the KSS assumption of the number of officers and number of teams tending to infinity is reasonable.

4 Effect of Officers on Enforcement, Force, Misconduct, and Crime

This section explores the influence of individual officers on outcomes which are directly observed at the officer-level — enforcement (stops, arrests for low-level and serious crimes), uses of force, and complaints of misconduct — and aggregate outcomes which may not be observed at the officer-level — crime, 911 calls. If estimated differences across officers are small and account for little influence on the observed outcomes, then the question of a trade-off between undesirable aggression and effective crime-fighting is mute as officers, themselves, would play no significant role. If differences across officers do matter for some outcomes, then the relationship between officer effects on different outcomes has significant implications.

4.1 Officer Effects on Officer-Level Outcomes

Table 4 displays bias corrected standard deviations of officer effects scaled by the mean outcome (Column (1)) and per 100 shifts (Column (2)), along with the and the total amount of variation explained by officer and assignment effects (Column (3)), and the share of each component of equation (7) explained by officer effects, assignment effects, and the sorting between the two (Columns (4)-(6)). Each set of rows corresponds to a different outcome. Because there are ten periods and six outcomes, the table displays the median value across the periods for each outcome with the minimum and maximum values in brackets below. Appendix D.1 discusses the importance of the bias-correction.

Focusing on Columns (1) and (2), differences across individual officers are economically significant. Over 100 shifts, 1 SD increase in officer effects for each respective outcome would

increase street stops by 32.9 (36.6 pre-2016 and 16.4 post-2016) or 87.5% of the mean, traffic stops by 16.4 (15.8 pre-2016 and 28 post-2016) or 195.1% of the mean, nonindex arrests by 9.93 (88.6% of the mean), index arrests by 1.22 (77.5% of the mean), uses of force by 0.421 (116.3% of the mean), and accusations of misconduct by 0.231 and 0.124 (99.3% and 123.3% of their respective means), for the 2016 and 2019 complaints data respectively.

We can compare some of these estimates to prior work: Adger et al. (2022) finds a 1 SD increase in use of force propensity equates to a 138% increase compared with the mean among field training officers using 911 call data from Dallas, and Rozema & Schanzenbach (2019) finds a 1 SD increase in officer effects for civilian allegations of misconduct is around 140% of the mean. Both are somewhat larger than estimates in this study, but qualitatively similar. In contrast, Weisburst (2022) finds a 31% increase in arrests in response to a 1 SD increase in officer effects using 911 call data from Dallas, though this smaller relative effects may be due to focusing on arrests in 911 calls, pooling arrest types, and assigning arrests to large groups of officers. Notably, the KSS bias correction proves to be very important, which is more robust than the Empirical Bayes methods used in the prior studies, as shown in Figure D.1. Plug-in variances would overestimate $Var(\theta_i^k)$, particularly for the less common outcomes, as the share of bias in the plug-in variance estimates ranging tends to be between 40-70% for index arrests and use of force, 60 to over 80% for complaints [2016] and [2019], 15-27% for nonindex arrests, and 20-40% for street and traffic stops.

For robustness, Figure 3 displays the standard deviations in outcomes per shift scaled by the means across the various specifications, and the results are highly consistent with the main results. Additionally, Figure C.1 displays standard deviations of officer effects scaled by the mean dependent variables for additional outcomes, including subtypes of arrests, stops, force, and complaints. Notably, officer heterogeneity is present and significant when filtering for stops and arrests for which the officer is a 'primary' officer (while the main specification allows for two officers per arrest and stop). Some subtypes display clearly larger degrees of heterogeneity compared with the main outcome, in particular, arrests for drug crimes,

municipal code violations, and traffic offenses, compared with pooled nonindex arrests.

Using Columns (4) and (5), we can compare the degree of officer heterogeneity with that of heterogeneity in assignments. Recall that assignment effects encapsulate a wide range of different assignments and roles officers can work in. Despite this, the magnitudes of officer heterogeneity ($Var(\theta^k)$) tends to be as large as assignment effect heterogeneity ($Var(\psi^k_J)$). For street stops and nonindex arrest variance in officer effects is around 1.17 and 1.35 times the size of variance of assignment effects, while for traffic stops officer heterogeneity is nearly twice as large as assignment heterogeneity (1.86). While officer heterogeneity is smaller than assignment effect heterogeneity for index arrests, this is expected as index arrests are considered less discretionary, nevertheless $Var(\theta)$ is around 0.559 the size of $Var(\psi)$. Officer heterogeneity is significantly larger than assignment heterogeneity for use of force (2.81), but less so for complaints (1.41 and 1.5 for [2016] and [2019]), suggesting that differences across officers are a more significant factor in these highly costly outcomes compared with differences in assignments. Nevertheless, officer and assignment effects explain small portions of the overall variation in the more rare outcomes $< 1\%$ and around 2% for index arrests, though they explain a significant portion of the more common and highly discretionary activities (14-22% of overall variation).

Finally, sorting between officers into assignments that match their behavioral propensities actually explains very little of the overall variation and is negative for complaints and near zero for almost all outcomes other than for nonindex arrests for which it is positive. This indicates that to the extent that officers do sort into assignments, it is most consistent with sorting on opportunities for low-level arrests rather than highly rare outcomes such as force or misconduct. Furthermore, it rejects any notion that 'bad apples' (high force or high misconduct officers) are sorting into assignments where they can abuse their power, producing more misconduct and use more force. One implication of this is that the raw use of force rates and complaint rates among officers, unconditional on their assignments, has a strong

relationship with their causal effect on complaints and force.[16]

4.2 Assessing the Productivity-Cost Tradeoff

One advantage of the many outcomes available for this study is to recover the degree of cross-outcome correlations in officer and assignment effects. Focusing on officer effects, the correlations between officer effects for different outcomes inform us on the mixture of officers who work in the CPD and their behavioral preferences. For example, Shi (2009) models misconduct is a byproduct of officers' effort in enforcing the law — if this is the case, we expect officer effects for arrests and misconduct to be strongly correlated.

Table 5 displays the cross-outcome correlations across officer effects. Interestingly, I do not find any very large (> 0.6) correlations between outcomes of different categories. Correlations between officer effects for index and non-index arrests and non-index arrests have the strongest correlation at around 0.4 and 0.37, while officer effects for index arrests and street stops are modestly correlated at 0.26. However, officer effects for traffic and street stops are only correlated at 0.17.

Most relevant to a productivity-cost trade-off, officer effects for enforcement activity are essentially uncorrelated with their effects on misconduct: street stops are generally negatively correlated with complaints; traffic stops and index arrests are essentially uncorrelated with complaints; and index arrests and stops are only modestly correlated with use of force (0.33

[16]An alternative explanation is that the results sorting are due to officers of different demographic groups having different working environment preferences and also different average officer effects. However, as shown in Figure C.3 there do not appear to be highly significant sorting patterns across demographic groups. White and male officers appear to sort more based on nonindex arrests compared with Black, Hispanic, and female officers, and negative sorting is smaller among white and male officers for other outcomes such as use of force, stops, but is more negative for index arrests. However, these differences are not significantly pronounced indicating that the general sorting patterns are consistent across race and gender groups. For tenure, as officers gain more tenure, sorting becomes increasingly negative though less pronounced for nonindex arrests. Furthermore, Figure C.2 displays the correlations across the other specifications, and these results are consistent. Finally, the degree of sorting across the main categories is partially driven by differential sorting on subtypes. Figure C.4 displays correlations across outcome subtypes. For nonindex arrests, there is positive sorting on the main subtypes (e.g., drug arrests, warrant arrests, municipal code violations, and others), while there is, on average, no sorting for weapons arrests and generally negative sorting on simple violent arrests. Officers negatively sort on index arrests, both violent and property, uses of force involving injuries, and all complaint types, suggesting these results are not due to aggregating outcome types.

and 0.11). Officers' propensities to make nonindex arrests are positively correlated with complaints for misconduct at around 0.14 and 0.11 and more strongly correlated with use of force at around 0.36. Complaints are, however, most strongly correlated with officers' propensity to use force at around 0.44 and 0.49.[17] The trade-off between each dimension can be quantified by determining the expected change in officer behaviors given a 1 SD increase in a specific behavior based on the variance-covariance relationships. Figure C.5 displays how officer behavior changes with a 1 SD change in a specific dimension. Consistent with the correlations, a single SD decrease in officer propensity to receive complaints has, at worst, a small negative effect on nonindex arrests but a relatively large negative effect on use of force (around 50% of the mean). A single SD decrease in use of force effects results in modest declines in nonindex and index arrest effects, though only around 25% of the mean and less than 20% for traffic and street stops, while reducing complaints by around 50% of the mean.

Finally, while the quantity of enforcement does not appear to be related to officers' propensities to use force or receive complaints, it is possible that the higher cost officers are making more impactful arrests in terms of quality. As a proxy for higher quality arrests, I classify arrests which resulted in a guilty finding (through trial or a plea) and resulted in the defendant/arrestee being sentenced to incarceration. However, $\theta^{Incar.}$ measures the officer's effect on the number of high quality arrests, I measure quality as the residual variation in $Var(\theta^{Incar})$ which cannot be explained by arrest quantity. Following Rose et al. (2022), I use the variance-covariance structure to mimic the regression of high quality arrests on arrest quantity and other activity metrics, as shown in Table 6. As shown in Column (1), Arrest quantity (using four crime types) explains 75% of the variation in $Var(\theta^{Incar})$, suggesting that variation in higher quality enforcement is still sizable beyond quantity of enforcement as the standard deviation of the residual variation in θ^{Incar} is around 60% of the mean of high quality arrests. Column (2) adds street stop and traffic stop propensities and Column (3) adds complaints [2016] and use of force propensities. The median implied R^2 increases

[17]Note that these cross-outcome correlations do not significantly differ across demographic groups, and most differences are not persistent across subsamples (as in Figure C.6).

by 0.0027 in Column (2), off of 0.75 in Column (1), and by an additional 0.0075 in Column (3). This suggests that beyond the quantity of officer activity, higher cost officers have no significant relationship with higher quality of enforcement.

4.3 Crime Effects

Table 7 summarizes the primary statistics of interest in this section using the main specification for estimating equation (3) — see Appendix D.6 for alternative specifications — and the first difference specification (equation (6)). Column (1) displays the mean of the dependent variable from the fixed effect ('FE') specification, which is used to scale the standard deviations in team effects for the FE specification and the standard deviations of team effects based on first difference ('FD') specifications in Columns (2) and (3) for comparability. Overall, the FD estimates are smaller than the FE estimates, with FD standard deviations generally being around one-fourth the size of FE standard deviations. However they are significantly more precise over the different time periods and are highly consistent across both actual and predicted team compositions (which are pooled in the table), where as the FE estimates produce more varying estimates across samples and occassionally reach negative $Var(\Omega)$ values due to the bias correction taking in a much larger number of parameters. Bias correction proves to be massively important, with the majority of plug-in variance estimates being over 60% bias, as shown in Figure C.7, and many being over 90% bias for reported violent crimes.

For violent crime, a 1 SD increase in team effects results in a 60% (FE) and 16% (FD) increase, with similar effects on both robbery and aggravated assaults. In terms of variation explained, team effects account for around 4.6% of $Var(\Omega)/Var(Y)$ and only 0.2% of $Var(\Delta\Omega)/Var(\Delta Y)$, though in both cases total variation explained is similar to or less than variation explained by just team effects. For property crime, a 1 SD increase in team effects results in a 51% (FE) and 16% (FD) increase with these effects being primarily driven by very similarly sized effects on theft. Burglary and motor vehicle theft, using FD estimates,

experience 22% and 23% increases in response to a 1 SD increase in team effects. Variation explained by teams is larger for property crime than violent crimes, with $Var(\Omega)/Var(Y)$ being around 10% for FE and 0.6% for FD, driven by variation in thefts (11% for FE and 0.5% for FD), while for burglary and motor vehicle theft teams explain 0.8% and 4% for FE and 0.2% and .017% for FD.

To better quantify the magnitude of these effects, we can focus on how a 1 SD better (crime-reducing) officer changes crime outcomes compared with adding a single officer to a team. As much of the crime literature studies the elasticity between the quantity of officers and crime, I use elasticity estimates from Mello (2019) and Chalfin & McCrary (2018) (CM) (which includes standard errors) to compute these quantity effects of adding a single *average* officer to the average team — going from 49 to 50 officers, or increasing officer quantity by 2%. I compare this to the effect of holding team size fixed but increasing a single officer by 1 SD in terms of crime-reduction quality using the FE and FD estimates. Figure 4 provides a comparison for the influence of the quantity and the 'quality' of officers. For violent crime, FE quality estimates are similarly sized to CM quantity estimates, while FD estimates are around 1/4th the size, with a single officer increasing in quality by 1 SD producing around 0.3% to 1.2% decrease in violent crime, compared with quantity effects of 0.6% for CM and 2.5% for Mello. For property crime, quality effects are consistently as large or larger than quantity estimates, though generally of comparable magnitudes — e.g., between 0.3% ad 1% for quality and 0.4% for CM and 1.5% for Mello for pooled property crimes. Overall, quality effects are a comparable size to quantity effects across index crime types.

Finally, for nonindex crimes, team effects tend to explain a larger share of variation than for index crimes under both FE and FD specifications: 15% and 1% for drug crimes, 6.5% and 0.3% for simple violent crimes, 3% and 0.1% for weapon offenses, and 6% and 1.1% for all other nonindex crimes. As a result, standard deviation increase in team effects using FE (FD) equates to a 123% (43%) increase in drug crimes, 48% (13%) increase in simple violent crimes, 126% (31%) increase in weapon offenses, and 44% (23%) increase in all other offenses.

These patterns are expected, as these crime types are more reliant on discretionary arrests made by officers.

Despite the importance of team heterogeneity, it is not clear why officers differentially influence crime. First, officers may be reducing crime through contemporaneous deterrence (Mello (2019), Chalfin et al. (2022)) through higher quality positioning, tactics, and visibility some officers reduce crime by increasing the perceived probability of detection. Alternatively, officers may be reducing crime by arresting and thus incapacitating offenders who would otherwise commit (additional) crimes (Mastrobuoni (2019)). Finally, it is possible that officers are influencing crime reports through increased reporting of crime or suppressing crime reports but not influencing underlying crime rates. In this section, I attempt to disentangle these potential mechanisms.

First, if officers are reducing crime through arrests and incapacitation, we would expect that teams which tend to make more arrests should also reduce unsolved crimes, implying a negative covariance between team effects for unsolved and arrests for the same crime types (if criminals broadly specialize within crime types). Specifically, because Crime = Cleared + Uncleared the variance of team's on any crime type can be decomposed into:

$$Var(\Omega^{Crime}) = Var(\Omega^{Cleared}) + Var(\Omega^{Uncleared}) + 2 \times Cov(\Omega^{Cleared}, \Omega^{Uncleared}) \quad (8)$$

Figure D.12 displays the components of equation (8) normalized by the overall variance of the crime effects. The primary takeaway being that officer effects for uncleared crime and arrests are generally positively correlated, though the covariance between the two is relatively negligible compared with the the variance in arrests for drug crimes and the variance in uncleared offenses for property crime. This suggests that for property crime, the influence of officers on property crime is driven by an influence on uncleared offenses rather than changing arrest rates. Additionally, officers who make more arrests of a given type are not those who

reduce uncleared crime rates, which is inconsistent with incapacitation and more consistent with officers being differentially good at deterring crime.

Finally, to determine if officers are simply influencing reporting of crime rather than crime itself, I use 911 calls as a proxy for underlying crime which cannot be easily manipulated by officers. Figure C.8 displays SDs of team effects scaled by the mean on 911 calls of different types and comparable crime types. The magnitudes for the influence of team heterogeneity on calls is consistent with their effects on crimes of similar types (e.g., calls about theft and reported thefts, assaults, robberies, etc.), which is consistent with officers actually influencing underlying crime rates and suggests that results are not driven by manipulation of reporting by officers.

Appendix D.5 discusses cross-crime correlations and correlations between team effects on crimes and related 911 calls. Overall, team effects are positively, but modestly, correlated across crime types, suggesting some teams are not better at preventing one type at the expense of the others. However, drug crime effects are generally very weakly correlated with more serious crime types, suggesting these discretionary arrests do not prevent other crimes.

5 Assessing the Activity-Crime Tradeoff

Given that differences across officers do influence crime rates where and when they are working, which officer behaviors contribute to crime reduction? This addresses the trade-off between effective crime prevention and costly behaviors by officers. In this section, I connect the officer-level effects on observable activities to their influence on crime to determine the extent of the trade-off between policing benefits and costly activity.

5.1 Empirical Strategies

One approach to estimate the relationship between officer characteristics and their influence on crime is to project estimated officer crime effects, $\hat{\omega}_i$, onto officer characteristics, including

demographics Z_i and their officer effects θ_i. As officer effects θ_i are proxied by estimates $\hat{\theta}_i$, recovered in Section 4:

$$\hat{\omega}_i^c = \hat{\theta}_i \eta^c + \mu_i^c \tag{9}$$

However, crime effects are imprecise due to large amounts of measurement error in $\hat{\omega}_i^c$ making it difficult to estimate a precise effects. To more directly relate officer characteristics to crime, I regress the crimes in each district-hour on the composition of officers, rather than individual officer effects, similar to equation (3) but replacing the team of fixed effects, Ω, with a vector of average behavioral propensities (e.g., mean officer effect on street stops), $\overline{\hat{\theta}}_{d,h} = \frac{1}{N_{\tau(d,h)}} \sum_{i \in \tau(d,h)}^{N_{\tau(d,h)}} \hat{\theta}_i$:

$$y_{d,h}^c = \overline{\hat{\theta}}_{d,h} \eta^c + \beta^c X_{d,h} + \gamma_{w(d,h)}^c + \epsilon_{d,h}^c \tag{10}$$

I include the same controls as in equation (3) in $X_{d,h}$ (polynomials of officer counts, exact date fixed effects, etc.), and I employ more conservative district-time fixed effects ($\gamma_{w(d,h)}$) than in the prior section. Specifically, I use District-Year-Month-Day of Week-Hour fixed effects (e.g., Mondays in district 7 at 2pm in July of 2013), which corresponds to at most 5 observations per $\gamma_{w(d,h)}^c$. Table C.1 summarizes means and standard deviations of team demographic compositions (Panel A), team average officer effects (Panels B and C), and the controls used (Panel D) across samples. Panel B summarises $\overline{\hat{\theta}}_{d,h}$ using 'raw' fixed effects while Panel C uses 'shrunken' fixed effects, which are shrunken toward the mean as random intercepts to reduce measurement error using the mixed models method, similar to Rozema & Schanzenbach (2019).

To overcome any concerns over endogeneity of team composition, I also instrument for each dimension of team composition using the predicted composition based on the predicted rotation of officers discussed in Section 3.3. Recall that the predicted rotation is determined at the start of the year and thus unrelated to any contemporaneous or short-run shocks to

crime rates. The first stage equation for each dimension of team composition (for either raw or shrunken θ's), where superscript p denotes the variable is predicted based on the rotation, is:

$$\overline{\hat{\theta}}_{d,h} = \overline{\hat{\theta}}^{p}_{d,h}\Gamma^1 + \Gamma^2 X_{d,h} + \Gamma^3_{w(d,h)} + \mu^p_{d,h} \tag{11}$$

Consistent with the strength of the predicted rotation matching with the actual rotation, predicted team composition measurements have a strong first stage relationship with actual team composition (see Panel A of Table C.2).

5.2 Results

Beginning with the projection approach, there are few statistically significant coefficients, coefficients tend to be inconsistently signed across sample periods, and confidence intervals tend to be large which is consistent with the large amount of bias in $Var(\widehat{\Omega})$, as shown in Figure C.12. While higher drug arrest propensities, unsurprisingly, have a large positive relationship with drug crime effects, the general lack of precision means we cannot discern underlying relationships between effects on crime and officer behavior.

Table 8 displays the results of estimating equation (10) using the pooled 2010-2019 sample violent crime, property crime, and drug crimes. Beginning with drug crime (almost entirely drug arrests) as a test of consistency with the results in the prior sections, the average propensity of a team's officers to make drug arrests has the clearly largest influence over reported drug crimes: a 1 SD increase in the team's average drug arrest propensity increases reported drug crimes by between 15-30% for OLS and 15-40% for IV coefficients, and the coefficient is even larger when using shrunken officer effects to reduce measurement error. Other arrest propensities have smaller and negative relationships, and use of force propensities have a consistently positive effect: 1 SD increase results in a 3-7% increase in drug crimes even conditional on officer propensities to make drug arrests.

For violent, the OLS specifications have no statistically significant effects, and the largest effect is from 1 SD increase in propensity to make violent arrests, which causes less than a 1% increase in violent crimes. Using IV, coefficients tend to be larger, with negative effects from teams with higher propensities for non-drug nonindex arrests, more complaints, and more street stops, and positive effects from more traffic stops. However, only the effect of non-drug nonindex arrest propensities remains statistically significant when using shrunken fixed effects, with 1 SD increase in non-drug nonindex arrests reducing violent crimes by 7% and allowing us to reject reductions less than 1.8%. The coefficients for other effects become significantly smaller though standard errors tend to decrease as well.

For property crime, coefficients tend to be inconsistent between OLS and IV specifications, often with sign reversal. Directionally, the most consistent effects are the negative but economically and statistically insignificant effects of drug arrests, the economically significant but highly imprecise effect of use of force and traffic stops. Most notably, the OLS specification using raw fixed effects suggests that 1 SD increase in street stop propensities reduces property crimes by 0.6% and rejecting decreases larger than 1.2%. While the magnitude and direction of this effect is highly consistent across OLS and IV using both raw and shrunken fixed effects, with reductions ranging from 0.47% to 0.76% per 1 SD, the effects are not statistically significant.

Figure C.9 displays these effect across samples and for crime subtypes for the IV specifications. Overall, we can generally reject economically significant relationships between officer propensities for specific activities and their ability to reduce crimes, with few exceptions. For violent crimes, the larger effect of non-drug nonindex arrest propensities is driven by effects on aggravated assaults (though point estimates are negative for robberies as well), and it is actually consistent in terms of magnitude (around 5-10% decrease per 1 SD increase) across time periods. For property crimes, there are more significant effects, but they are inconsistent across samples. For example, on in the 2010-2012 sample, drug arrest propensities reduce theft (by 5-7% per 1 SD); and only in the 2016-2019 sample, traffic stops reduce each

property crime subtype (by 3-10% per 1 SD) but are not statistically significant and use of force reduces burglaries (by 10% per 1 SD). Finally, the effects of drug arrests on drug crimes is consistent across samples, and teams with higher propensities for non-drug nonindex arrests tend to increase other nonindex crimes. Higher nonindex non-drug arrests have a large negative effect on weapon offenses (30-40% per 1 SD) in 2010-2012, as do traffic stops in the 2016-2019 (20% per 1 SD, but not statistically significant).

Overall, these results provide very little evidence for a strong relationship between officer-level behaviors and differences in their effectiveness at preventing crime. While differential arresting behavior and perhaps stopping frequency may matter for some crime types, I find essentially no evidence that officers who receive more complaints or use more force are also reducing high cost crime.

5.3 Effect of Stops and Arrests on Crime

As a final strategy to connect officer behaviors to crime, I directly regress crimes on the number of officer-level outcomes (y^k, e.g., number of street stops) that occur at the district-hour level:

$$y^c_{d,h} = y^k_{d,h}\kappa^{k,c} + \beta^c X_{d,h} + \gamma^c_{w(d,h)} + \epsilon^c_{d,h} \tag{12}$$

and I instrument for the officer-outcome using the average predicted propensity of the team for that activity:

$$y^k_{d,h} = \overline{\hat{\theta}^k}^p_{d,h}\Gamma^1 + \Gamma^2 X_{d,h} + \Gamma^3_{w(d,h)} + \mu^p_{d,h} \tag{13}$$

There is potential for exclusion restriction violations because, for example, officers' propensities to make stops may also increase use of force which then influences crime. I borrow from the judge-IV literature and, in an additional specification, I add the predicted team average officer effects for the other outcomes $\overline{\hat{\theta}^{\setminus k}}^p_{d,h}$ as controls (Mueller-Smith (2015)).[18]

[18]For example, I use propensities for street stops as the instrument for total stops and controlling for

The first stage results are shown in Panels B (endogenous variable only) and C (additional prediction controls), with generally strong first stages.

The results are displayed in Figure C.10 for property and violent crime for endogenous officer-activities street and traffic stops, all arrests, nonindex arrests, and the most discretionary types of arrests for the lowest levels of crime, 'quality-of-life' (QOL) (Cho et al. (2023)), across samples and specifications.[19] I scale the coefficient κ^c by the mean of the endogenous variable over the mean of the outcome, $\frac{\overline{y^h_{d,h}}}{\overline{y^c d,h}}$, to compute the elasticity between crime and officer-level activities. Consistent with small effects of officer behaviors on crimes and despite the strong first stage between predicted team propensities and actual outcomes, I nearly all small and not statistically significant effects across samples and results are similar across specifications.

Table 9 displays results for the pooled sample using shrunken officer effects to construct team propensities. For violent crimes, we can reject elasticities larger (more negative) than -0.15 for street stops, -0.087 for traffic stops, -0.52 for all arrests, -0.37 for nonindex arrests, and -0.24 QOL arrests. The point estimates tend to be even smaller and only occasionally negative – specifically elasticities are a minimum of -0.04 for street stops, -0.16 for nonindex arrests, and -0.12 for QOL arrests, while traffic stops have a small positive effect. For property crime, implied elasticities are more precise, allowing us to reject elasticities larger than -0.079 for street stops, -0.086 for traffic stops, -0.099 for all arrests, -0.092 for nonindex arrests, and -0.06 QOL arrests. We can compare these lower bounds to the smallest across recent estimates for the elasticity of crime with respect to police quantity of -0.34 for violent and -0.174 for property crime from Chalfin & McCrary (2018).[20] Furthermore, Cho et al. (2023) calculates implied elasticities of violent and property crimes with respect to arrests at 0.42 and -0.15, but with confidence intervals that cannot reject elasticities between -0.4 and 0.3

predicted non-street stop propensities. See also, Bhuller et al. (2020), Norris et al. (2021), and Humphries et al. (2023).

[19]Quality of life crimes include drug, liquor, prostitution, disturbing the peace, miscellaneous nonindex, and municipal code violations.

[20]Note that Chalfin & McCrary (2018) elasticities are the smallest compared with other relevant estimates from Mello (2019), Weisburst (2019), Evans & Owens (2007), and Lin (2009).

for either. However, we also can reject modestly sized positive elasticities, larger than +0.2, between any of these officer activities and crime. Overall, we can reject even relatively small effects of officer-level activities, in particular street and traffic stops and low-level arrests, on reducing property and violent crime.

6 Effects of Officer Characteristics

Given that individual officer behaviors have a small direct influence on crime, a key question is whether demographics and easily observable characteristics of officers are linked to their effects on both crime reduction and officer activity. While prior work has documented how race, gender, and other officer characteristics influence stops, arrests, and use of force (Ba, Knox, et al. (2021), Goncalves & Mello (2021), Hoekstra & Sloan (2022), Weisburst (2022), Rivera (2022)), we do not know if these groups also allow for increased crime or if they are actually better at preventing crime. In this section, I expand on prior work exploring the effects of not only officer race, gender, and tenure, but also officer education, military experience, language ability, on the primary officer activities and on crime when they work.

6.1 Effect of Characteristics on Officer-Level Activities

First, to estimate the effect of officer characteristics on officer activities and crime, I project estimates officer effects on activities ($\hat{\theta}_i^k$) onto officer characteristics (Z_i):

$$\hat{\theta}_i^k = Z_i \alpha^k + \mu_i^k \tag{14}$$

The results of projecting characteristics onto officer effects for officer-level activities (equation (14)) are summarized in Table 10, displaying the median estimate, scaled by the mean of the raw outcome, along with the minimum and maximum estimates across sample periods, where stars correspond to p-values using Kline et al. (2020) -bias corrected heteroskedasiticty robust standard errors. Figure C.11 displays all point estimates along with

95% confidence intervals for all main and sub-outcomes. Across outcomes, the amount of variation explained by these characteristics is highest for street stops, nonindex arrests, and uses of force with medians of at least 8%, and around 4% for index arrests, 3% for traffic stops (though this shifts over time to a peak of 23%), and less than 3% for complaints.

Consistent with prior literature, Black officers make fewer street stops (-29%), nonindex arrests (-18%), index arrests (-14%), uses of force (-39%), but they also receive fewer complaints (-24% and -15% for [2016] and [2019] complaints respectively) and make fewer traffic stops (-12%), though the differences in Complaints 2019 effects are not statistically significant for the median. By comparison, Hispanic officers are only slightly different from white officers making around less than 10% fewer stops and arrests, but they use force 18% less and receive complaints around 12% less — though this shifts over time. As shown in Figure C.11, while Hispanic officers tend to receive fewer or the same amount of complaints for serious and service related misconduct, Black officers receive fewer complaints for serious misconduct and generally more complaints for service related misconduct.

Female officers make fewer stops, traffic stops, and index arrests (medians smaller than -18%) compared with male officers, but they make fewer nonindex arrests (-23%) and use force (-45%) and receive complaints at a significantly lower rate (-32% - -37%). As a result, while they engage in more or similar numbers of stops and index arrests compared with Black officers, they use less force, make fewer low-level arrests, and receive fewer complaints. As shown in Figure C.11, the differences in complaints for female versus male officers comes from fewer serious misconduct complaints rather than service complaints.

The age officers start in the CPD also matters, and when conditioning on tenure this effectively compares officers of the same experience but who are older or younger due to start age. As officers start slightly older, above the average age of 29, and much older, above 33 years, they make slightly fewer street stops, nonindex arrests, and index arrests, and they make more traffic stops. However, older-starting officers use substantially less force and receive many fewer complaints.

For life experiences, Spanish-speaking officers are negligibly different compared with non-Spanish speaking officers. Education proves to be a negligible factor as well for most outcomes. Officers with more than a BA tend to make more stops than those with missing educational information and those with less than a BA by around 6% of the mean, but few differences across other outcomes. Notably, the lack of results for education prove to consistent across subtypes.

Officers with proxied military experience more aggressively than non-military. Though point estimates differ using the 'military' and 'military (active)' classifications, the trends are the same: military-related officers make more stops (+16%, +7%) and traffic stops (+18%, 37%), index and nonindex arrests ($\geq$ +20%), and use more force (+20%, +29%), but unlike all other groups with more aggressive policing, they actually receive fewer complaints on average (-26% and -7% for [2016] and -13% and -27% for [2019]) and this is generally consistent across complaint subtypes. So, despite being more aggressive, military officers do not on average receive more accusations of misconduct.

Finally, officer tenure proves to be the largest factor in changing all outcomes. Officers with 5 - 10 years of experience make 22-31% fewer street and traffic stops, and -9% fewer index and nonindex arrests, and use 41% less force compared with officers with less than 5 years of experience. While not statistically significant at the median, they also receive 9-20% fewer complaints. However, these differences increase with higher tenure officers from 10-15, 15-20, and $>$ 20 years of experience compared with $<$ 5 years: differences in street stops, traffic stops, nonindex arrests, and index arrests exceed -50% for $>$ 20 years of experience. While tenure reduces enforcement activity it has a significantly large influence on use of force and complaints of misconduct, and these effect sizes tend to be larger than any other demographic influence.

6.2 Comparing Activity and Crime Effects

Given these large effects, I estimate the effect of officer characteristics on their influence on crime by projecting officer characteristics onto individual crime effects ($\hat{\omega}_i^c$):

$$\hat{\omega}_i^c = Z_i \alpha^c + \mu_i^c \tag{15}$$

Additionally, I modify equations (10) and (11) by replacing (predicted) team average officer fixed effects with (predicted) team average demographic composition ($\overline{Z}_{d,h} = \frac{1}{N_{d,h}} \sum_{i \in \tau(d,h)} Z_i$):[21]

$$y_{d,h}^c = \overline{Z}_{d,h} \eta^c + \beta^c X_{d,h} + \gamma_{w(d,h)}^c + \epsilon_{d,h}^c \tag{16}$$

$$\overline{Z}_{d,h} = \overline{Z}_{d,h}^p \Gamma^1 + \Gamma^2 X_{d,h} + \Gamma_{w(d,h)}^3 + \mu_{d,h}^p \tag{17}$$

Because the two types of officer-effects are estimated at different frequencies — i.e., θ_i is the effect of officer i at the officer level over a 9-hour shift, while ω_i is the effect of officer i as one member of a team of some size over a single hour — I convert elements of $\hat{\alpha}^k$ and $\hat{\alpha}^c$ to both be the effect of replacing a single officer with the baseline characteristics (white, male, < 5 years of tenure, missing education, non-military, starting at an age < 25, and non-Spanish speaking) with another officer who is different on a specific dimension on 200 9-hour shifts (roughly 1 year), estimated over the same sample period. Table 11 displays the estimated effect of an officer's characteristics relative to the baseline officer on policing activity and crime outcomes over a single year (200 9-hour shifts) across sample periods for a subset of demographics. Figure C.13 displays effects scaled by the means of the outcome variables for all demographics.

Consistent with the prior results, replacing a white officer with a Black officer, male officer with a female officer, and a less experienced officer with a more experienced officer

[21] See Panel D of Table C.2 for the relevant first stage F-statistics.

produces significantly different officer-level activities. For example, a Black officer will make at least 2 fewer nonindex arrests, about 0.4 fewer index arrests, and around 0.2 fewer uses of force; however, crime will not significantly increase, with estimates of between -0.61 to 0.67 violent crimes and between -0.97 to 0.39 property crimes. Hispanic officers have no effect on crime but lower activities compared with white officers, though differences are smaller compared than the Black-White gap. Replacing a male officer with a female officer will result in significantly fewer arrests (similar to the Black-white gap), uses of force (at least 0.26 fewer), complaints (at least 0.13 for [2016] and at least -0.039 for [2019]), while all point estimates suggest the reduce violent crimes by -0.38 to -1.3 though none are statistically significant but they allow for more property crime between 0.041 ($p > 0.1$) and 3.8 ($p < 0.05$).

For officers of higher tenures, compared with new officers (< 5 years of experience), make significantly fewer stops, arrests, uses of force, and complaints. While point estimates suggest that very experienced officers (tenures between $15 - 20$ and > 20) reduce violent crime by -1.1 to -3.5 (both $p > 0.1$) and -1.9 ($p > 0.1$) to -4 ($p < 0.05$), higher tenure groups have consistently positive and generally statistically significant effects on property crime in the 2010-2012 and 2013-2015 samples: $15 - 20$ increase property crime by 10 ($p < 0.05$) and 12 ($p < 0.01$) and > 20 increase property crime by 9.4 ($p < 0.05$) and 10 ($p < 0.05$). Directionally, these effects on violent and property crime are broadly consistent with the effects of tenure on crime in Ba, Bayer, et al. (2021).

Overall, these effects are consistent with those using equations (16) and (17) to directly estimate the effect of characteristics on crime. Table 12 displays the OLS and IV coefficients from the pooled samples. While team composition have at best small effects on contemporaneous crimes — most coefficients suggesting that 1 SD increase in a team composition on a particular dimension (tenure, race, gender, etc.) has shifts crime outcomes by less than 1% of the mean, with almost all coefficients being not statistically significant. The exception is that higher tenure officers have positive effects on property crime, with 1 SD increases in the shares of 10-15, 15-20, and > 20 years of experience officers on a team all increasing property

crime by over 3% of the mean under IV (all $p < 0.1$).[22]

7 Conclusion

This paper studies the trade-off between costly police behaviors, such as using force and misconduct, and their effectiveness in enforcing the law and reducing crime. I find that differences across individual officers have significant implications for the number of stops, arrests, uses of force, complaints, and crimes reported when and where they work. However, I find little evidence of a significant relationship between officer's propensity to use more force or receive more complaints and either their propensity to make more arrests for serious crimes or their ability to reduce crime where and when they are working, which is contrary to the conception of aggressive officers being more effective police officers. Furthermore, I find small effects of stops and low-level arrests on crime rates, suggesting a weak link between these officer-level activities and reducing crime through deterrence. Finally, officer characteristics play a significant role in differences in officer activity but, with few exceptions, tend to have smaller effects on differences in crime-prevention ability. While these results suggest that there is little trade-off in training or recruiting less costly police officers in terms of increased crime or fewer important arrests, they also indicate that the features that make an officer good at preventing crime are not clearly linked to traditional metrics of officer activity and thus are difficult to measure.

[22]Additionally, drug crimes display multiple significant effects. For example, higher shares of Black and more tenured officers have statistically and economically significant negative effect, with 1 SD increase in share Black or officers with over 20 years of tenure both decreasing drug crimes by over 10% of the mean under the IV specification (both $p < 0.05$). Unsurprisingly, given the results in the prior section, the characteristics of the team have relatively small effects on crimes.

8 References

Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. *Longitudinal Employer-Household Dynamics Technical Papers*.

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*(2), 251–333.

ACLU. (2015). *Stop and frisk in chicago*. ACLU of Illinois.

Adger, C., Ross, M., & Sloan, C. (2022). *The effect of field training officers on police use of force* (p. 57) [Working Paper].

Agan, A., Doleac, J. L., & Harvey, A. (2023). Misdemeanor prosecution. *The Quarterly Journal of Economics*.

Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *171*(3), 673–697.

Ang, D. (2021). The effects of police violence on inner-city students. *The Quarterly Journal of Economics*, *136*(1), 115–168.

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2023). *Shots fired: Crime and community engagement with law enforcement after high-profile acts of police violence* [Working Paper]. Working Paper.

Ang, D., & Tebes, J. (2023). Civic responses to police violence. *American Political Science Review*, 1–16.

Ba, B., Bayer, P., Rim, N., Rivera, R., & Sidibé, M. (2021). *Police officer assignment and neighborhood crime* (Working Paper 29243). National Bureau of Economic Research.

Ba, B., Knox, D., Mummolo, J., & Rivera, R. (2021). The role of officer race and gender in police-civilian interactions in chicago. *Science*, *371*(6530), 696–702.

Best, M. C., Hjort, J., & Szakonyi, D. (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review*, *113*(8), 2121–2167.

Bhuller, M., Dahl, G. B., Løken, K. V., & Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, *128*(4), 1269–1324.

Bonhomme, S. (2021). *Teams: Heterogeneity, sorting, and complementarity* (arXiv:2102.01802). arXiv. http://arxiv.org/abs/2102.01802

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., & Setzler, B. (2022). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*.

Campbell, R. A. (2024). *What does federal oversight do to policing and public safety? Evidence from seattle* [Working Paper].

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, *128*(3), 967–1015.

Chalfin, A., & Goncalves, F. (2021). *The professional motivations of police officers* [Working Paper]. Working Paper.

Chalfin, A., Hansen, B., Weisburst, E. K., & Williams Jr., M. C. (2022). Police force size and civilian race. *American Economic Review: Insights*, *4*(2), 139–158.

Chalfin, A., & McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, *55*(1), 5–48.

Chalfin, A., & McCrary, J. (2018). Are u.s. Cities underpoliced? Theory and evidence. *The Review of Economics and Statistics*, *100*(1), 167–186.

Chan, D. C. (2021). Influence and information in team decisions: Evidence from medical residency. *American Economic Journal: Economic Policy*, *13*(1), 106–137.

Chen, J., & Roth, J. (2023). *Logs with zeros? Some problems and solutions*.

Chen, Y. (2021). Team-specific human capital and team performance: Evidence from doctors. *American Economic Review*, *111*(12), 3923–3962.

Cheng, C., & Long, W. (2018). Improving police services: Evidence from the french quarter task force. *Journal of Public Economics*, *164*, 1–18.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers i:

Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.

Cho, S., Gonçalves, F., & Weisburst, E. (2023). *The impact of fear on police behavior and public safety* (31392). National Bureau of Economic Research. https://www.nber.org/papers/w31392

Corman, H., & Mocan, N. (2005). Carrots, sticks, and broken windows. *The Journal of Law and Economics*, *48*(1), 235–266.

Cox, R., & Cunningham, J. P. (2021). Financing the war on drugs: The impact of law enforcement grants on racial disparities in drug arrests. *Journal of Policy Analysis and Management*, *40*(1), 191–224.

Denman, D. S. (2023). The cinematic universe of copaganda: World-building and the enchantments of policing. *Culture, Theory and Critique*, *0*(0), 1–16.

Devi, T., & Fryer, J., Roland G. (2020). *Policing the police: The impact of "pattern-or-practice" investigations on crime* (27324). National Bureau of Economic Research. https://www.nber.org/papers/w27324

Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, *82*(3), 991–1030.

Di Tella, R., & Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, *94*(1), 115–133.

Donohue, J. J., Ho, D. E., & Leahy, P. (2013). Do police reduce crime?: A reexamination of a natural experiment. In *Empirical legal analysis*. Routledge.

Draca, M., Machin, S., & Witt, R. (2011). Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *American Economic Review*, *101*(5), 2157–2181.

Dube, O., MacArthur, S. J., & Shah, A. K. (2023). *A cognitive view of policing* (31651). National Bureau of Economic Research. https://www.nber.org/papers/w31651

Evans, W. N., & Owens, E. G. (2007). COPS and crime. *Journal of Public Economics*, *91*(1),

181–201.

Feigenberg, B., & Miller, C. (2022). Would eliminating racial disparities in motor vehicle searches have efficiency costs? *The Quarterly Journal of Economics*, *137*(1), 49–113.

Goncalves, F., & Mello, S. (2021). A few bad apples? Racial bias in policing. *American Economic Review*, *111*(5), 1406–1441.

Gonçalves, F. M., & Mello, S. (2023). *Police discretion and public safety* (31678). National Bureau of Economic Research. https://www.nber.org/papers/w31678

Grosjean, P., Masera, F., & Yousaf, H. (2023). Inflammatory political campaigns and racial bias in policing. *The Quarterly Journal of Economics*, *138*(1), 413–463.

Gudgeon, M., Jordan, A., & Kim, T. (2023). *Do teams perform differently under black and hispanic leaders? Evidence from the chicago police department* [Working Paper]. Working Paper.

Hausman, D., & Kronick, D. (2023). The illusory end of stop and frisk in chicago? *Science Advances*, *9*(39).

Hoekstra, M., & Sloan, C. (2022). Does race matter for police use of force? Evidence from 911 calls. *American Economic Review*, *112*(3), 827–860.

Holz, J. E., Rivera, R., & Ba, B. A. (2023). Peer effects in police use of force. *American Economic Journal: Economic Policy*, *15*(2), 256–291.

Humphries, J. E., Ouss, A., Stavreva, K., Stevenson, M. T., & Dijk, W. van. (2023). *Conviction, incarceration, and recidivism: Understanding the revolving door.*

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107.

Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, *108*(4), 1034–1073.

JG0304. (2009, November 18). *New CPD schedule! Officer.com.*

Jochmans, K., & Weidner, M. (2019). Fixed-effect regressions on network data. *Econometrica*, *87*(5), 1543–1560.

Kane, T., & Staiger, D. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (w14607). National Bureau of Economic Research.

Kim, T. (2023). *Facilitating police reform: Body cameras, use of force, and law enforcement outcomes* [Preprint]. SSRN.

Kim, T. (2024). *Promotion incentives, career decisions, and police performance* [Working Paper]. Working Paper.

Klick, J., & Tabarrok, A. (2005). Using terror alert levels to estimate the effect of police on crime. *The Journal of Law and Economics*, *48*(1), 267–279.

Kline, P., Saggio, R., & Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, *88*(5), 1859–1898.

Krueger, A. B., & Summers, L. H. (1988). Efficiency wages and the inter-industry wage structure. *Econometrica*, *56*(2), 259.

Lachowska, M., Mas, A., Saggio, R., & Woodbury, S. A. (2023a). Do firm effects drift? Evidence from washington administrative data. *Journal of Econometrics*, *233*(2), 375–395.

Lachowska, M., Mas, A., Saggio, R., & Woodbury, S. A. (2023b). *Work hours mismatch* (31205). National Bureau of Economic Research. https://www.nber.org/papers/w31205

Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: reply. *American Economic Review*, *92*(4), 1244–1250.

Lin, M.-J. (2009). More police, less crime: Evidence from US state data. *International Review of Law and Economics*, *29*(2), 73–80.

Linnemann, T. (2022). *The horror of police*. University of Minnesota Press.

Lum, C., Koper, C. S., Wilson, D. B., Stoltz, M., Goodier, M., Eggins, E., Higginson, A., & Mazerolle, L. (2020). Body-worn cameras' effects on police officers and citizen behavior: A systematic review. *Campbell Systematic Reviews*, *16*(3).

Mac Donald, H. (2016). *The war on cops : How the new attack on law and order makes everyone less safe* (First American Edition.). New York: Encounter Books.

Markovic, J. (2009). Why neighborhoods matter: The importance of geographic composition.

Geography & Public Safety, *2*.

Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics*, *121*(3), 783–821.

Mastrobuoni, G. (2019). Police disruption and performance: Evidence from recurrent redeployments within a city. *Journal of Public Economics*, *176*, 18–31.

Mastrorocco, N., & Ornaghi, A. (2020). *Who watches the watchmen? Local news and police behavior in the united states* (tep0720). Trinity College Dublin, Department of Economics.

Mello, S. (2019). More COPS, less crime. *Journal of Public Economics*, *172*, 174–200.

Mello, S. (2021). *Fines and financial wellbeing* (p. 99) [Working Paper].

Mikdash, M., & Zaiour, R. (2023). *The impact of police shootings on gun violence and civilian cooperation* [Preprint]. SSRN.

Moser, W. (2012). Garry McCarthy's new chicago crime strategy: Social networks, 'hot people'. *Chicago Magazine*.

Moskos, P. (2008). Under arrest: Discretion in the ghetto. In *Cop in the hood* (STU - Student edition, pp. 111–157). Princeton University Press.

Mueller-Smith, M. (2015). *The criminal and labor market impacts of incarceration* (p. 59) [Working Paper].

Norris, S., Pecenco, M., & Weaver, J. (2021). The effects of parental and sibling incarceration: Evidence from ohio. *American Economic Review*, *111*(9), 2926–2963.

Owens, E., Weisburd, D., Amendola, K. L., & Alpert, G. P. (2018). Can you build a better cop? *Criminology & Public Policy*, *17*(1), 41–87.

Parker, S. T., Ross, M. B., & Ross, S. (2024). *Driving change: Evaluating connecticut's collaborative approach to reducing racial disparities in policing* (32692). National Bureau of Economic Research. https://www.nber.org/papers/w32692

Premkumar, D. (2022). *Public scrutiny, police behavior, and crime consequences: Evidence from high-profile police killings* (p. 132) [Working Paper].

Prendergast, C. (2001). *Selection and oversight in the public sector, with the los angeles*

police department as an example (w8664). National Bureau of Economic Research.

Prendergast, C. (2007). The motivation and bias of bureaucrats. *The American Economic Review; Nashville*, *97*(1), 180–196.

Rackstraw, E. (2023). *When reality TV creates reality: How "copaganda" affects police, communities, and viewers* (4592803). https://papers.ssrn.com/abstract=4592803

Rivera, R. (2022). *The effect of minority peers on future arrest quantity and quality* (SSRN Scholarly Paper 4067011). Social Science Research Network.

Rivera, R., & Ba, B. A. (2023). The effect of police oversight on crime and misconduct allegations: Evidence from chicago. *The Review of Economics and Statistics*, 1–45.

Rose, E. K., Schellenberg, J. T., & Shem-Tov, Y. (2022). *The effects of teacher quality on adult criminal justice contact* (30274). National Bureau of Economic Research. https://www.nber.org/papers/w30274

Rozema, K., & Schanzenbach, M. (2019). Good cop, bad cop: Using civilian allegations to predict police misconduct. *American Economic Journal: Economic Policy*, *11*(2), 225–268.

Rozema, K., & Schanzenbach, M. (2023). Does discipline decrease police misconduct? Evidence from chicago civilian allegations. *American Economic Journal: Applied Economics*, *15*(3), 80–116.

Shi, L. (2009). The limit of oversight in policing: Evidence from the 2001 cincinnati riot. *Journal of Public Economics*, *93*(1), 99–113.

Skogan, W. G. (2022). Prospects for reform? The collapse of community policing in chicago. *The University of Chicago Law Review*, *89*(2), 383–404.

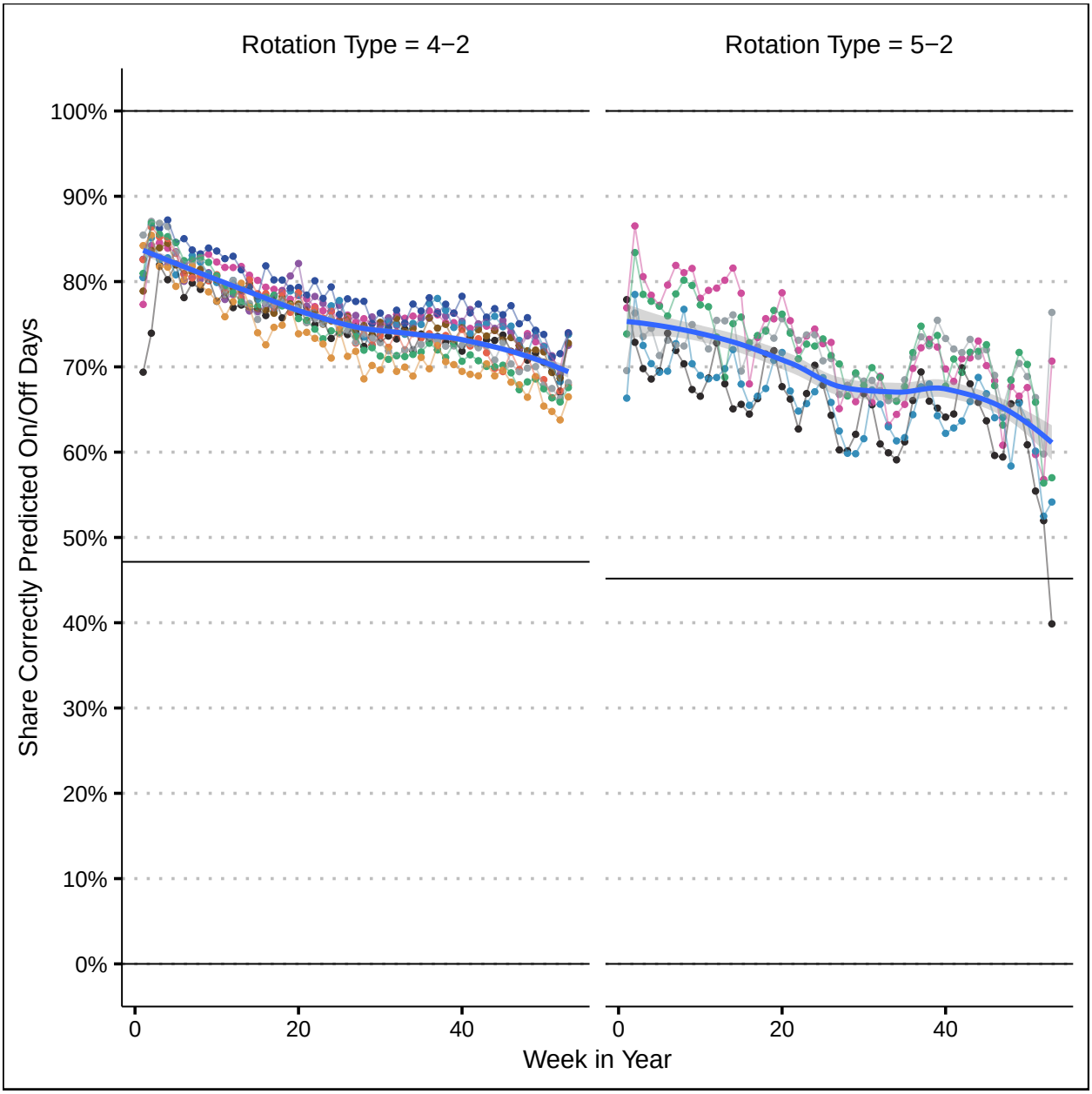
Stashko, A. (2022). Do police maximize arrests or minimize crime? Evidence from racial profiling in u.s. cities. *Journal of the European Economic Association*.

Weisburst, E. K. (2019). Safety in police numbers: Evidence of police effectiveness from federal COPS grant applications. *American Law and Economics Review*, *21*(1), 81–109.

Weisburst, E. K. (2022). "Whose help is on the way?": The importance of individual police

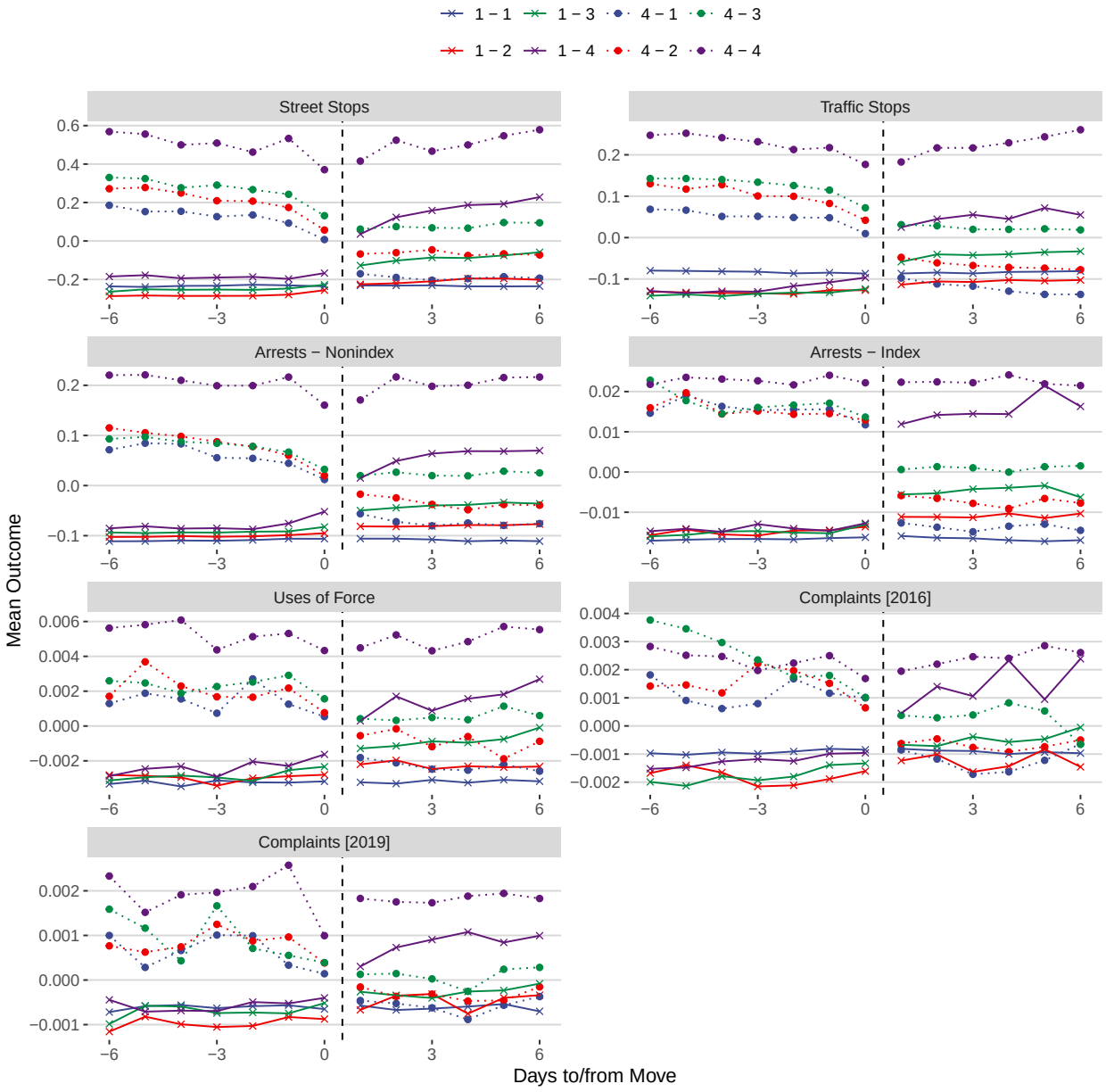
officers in law enforcement outcomes. *Journal of Human Resources*.

Figure 1: Accuracy of Predicted Days On and Off



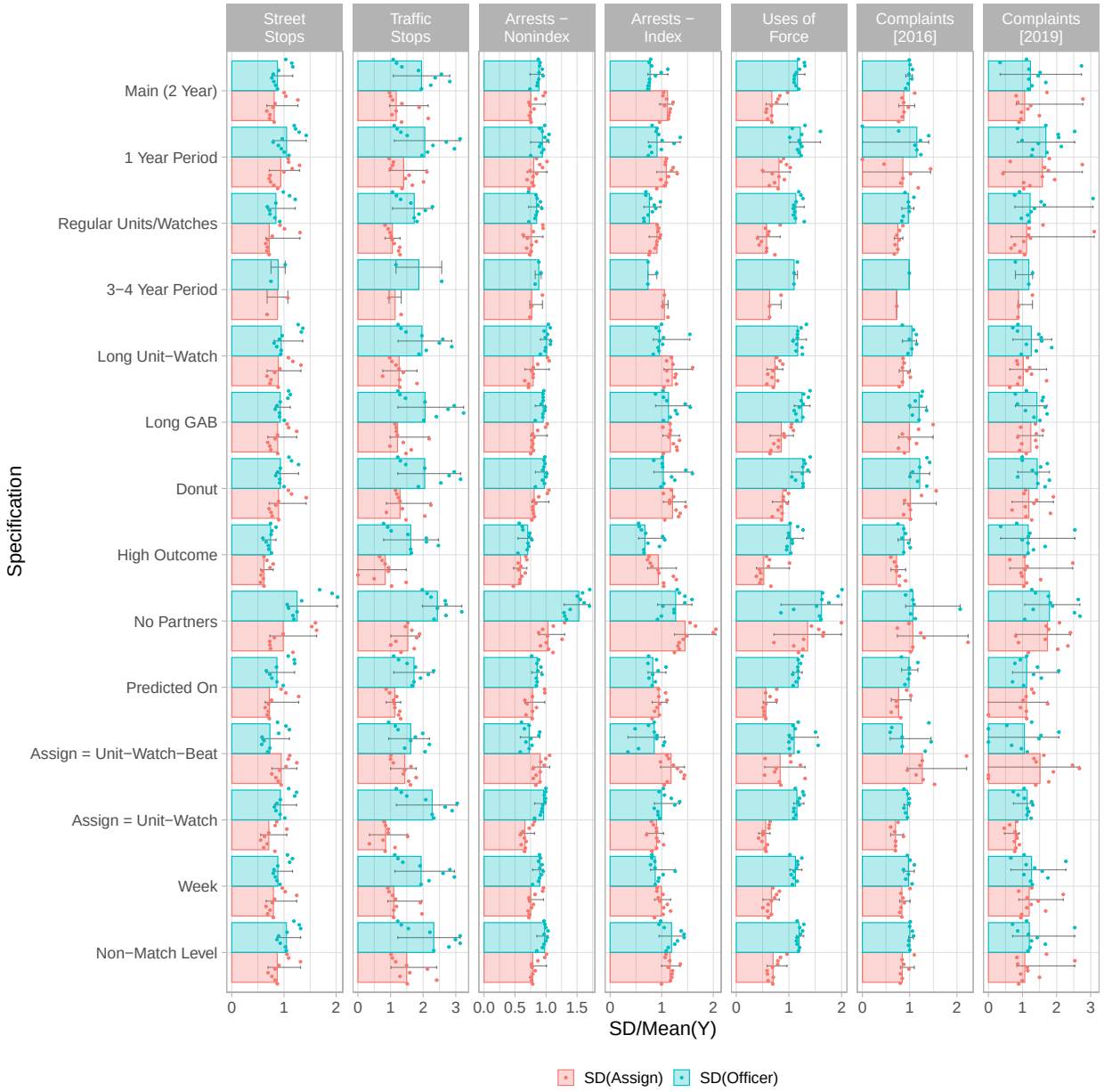
Note: Figure displays the share of correctly predicted days on and days off using based on projecting backward and forward the first 4 days on and 2 days off (4-2) and 5-2 spells, with average correct predictions aggregated at the year-week level and each color denoting a different year of data. The horizontal line corresponds to the accuracy of a random set of days given the spell duration.

Figure 2: Mean Outcomes of Assignment Switchers by Quantile of Mean Outcome in Assignment



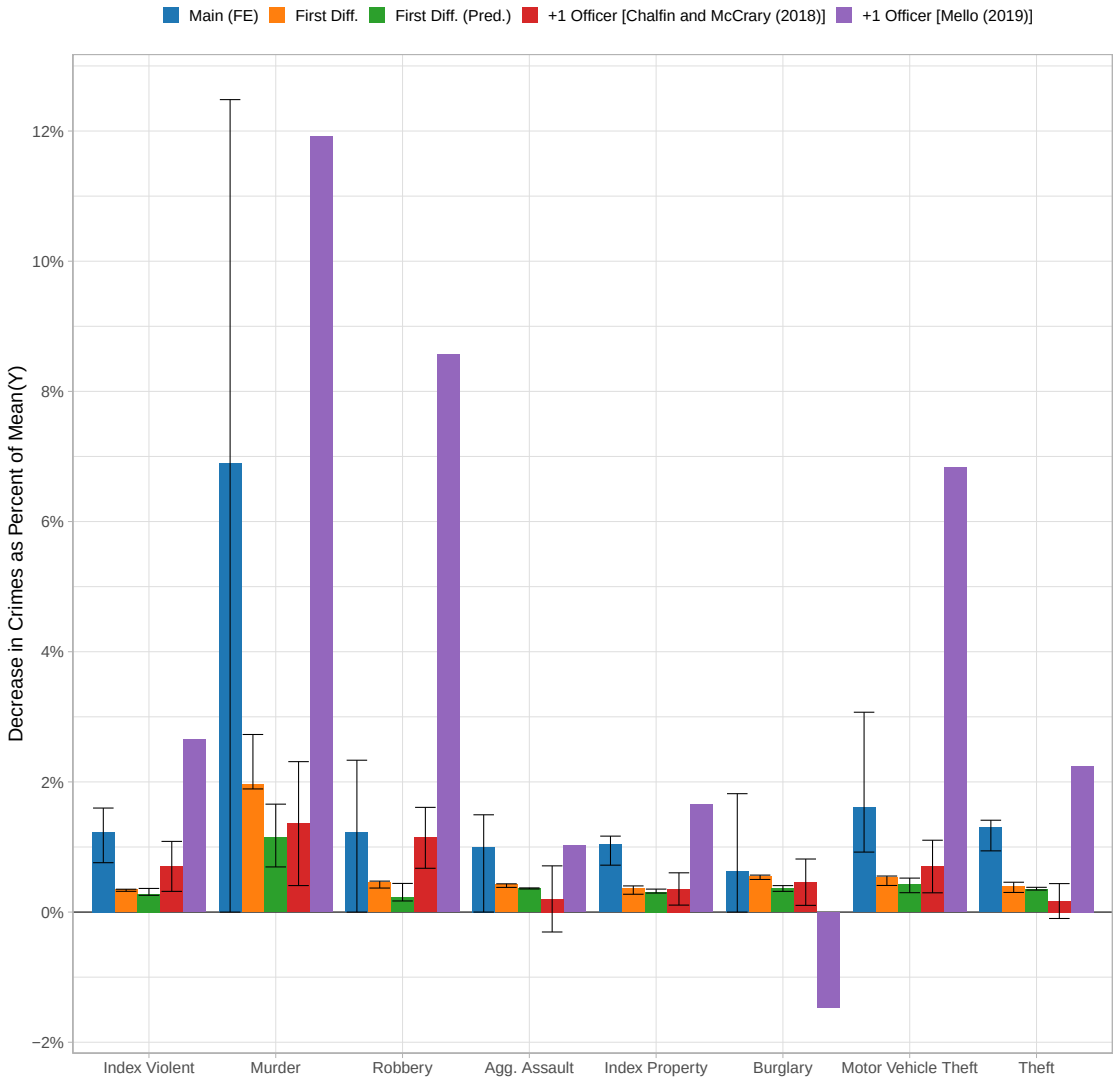
Note: Figure displays the mean outcome for officers moving between assignments (GABs), with outcomes demeaned by year, grouped by the quantile of the GAB they are transitioning to and from, in the seven shifts leading up to and after the switch (x-axis), with the shift right before the switch being 0. The quantile of an officer's assignment is computed at the outcome-year-level and leaving out the officer themselves, and quantiles of the leave-out values are weighted by observations within each year. Figure only displays transitions for officers in quantile 1 (lowest), with star shapes and solid lines, and quantile 4 (highest), with circle shapes and dotted lines, outcome GABs prior to their switch, while colors correspond to the quantile of the assignment the officer is switching to. Not all officer-assignment spells are six shifts or longer, so the set of officer-assignment spells contributing to the mean outcomes for x=1 is larger than the set of officer-assignment spells contributing to the mean outcomes for x=2, for example.

Figure 3: Magnitude of Officer and Assignment Effect Heterogeneity by Specification



Note: Figure displays the KSS bias-corrected standard deviations of officer and assignment effects (recovered from equation (1)) across different specifications scaled by the mean of the outcome, with each point denoting an estimate for a specific time period, the bar denoting the median across these estimates, and the confidence interval denoting the minimum and maximum estimates. Main (2 Year) refers to the main specification with 10 2-year rolling periods; 1 Year Period refers to each year of data analyzed separately; Geographic refers to the main specification but subsetting to geographic units only; 3-4 Year period refers to estimating the model using 3 distinct periods which cover 2009-2012, 2013-2015, 2016-2019; Long Unit-Watch (GAB) refers to selecting for spells in which the officer spends at least 20 (4) consecutive days in the unit-watch (GAB); Donut refers to removing the first and last day of each officer-GAB spell; High Outcome selects for GABs which have average outcomes (all main outcomes summed) of at least 0.25 within a year; No Partners removes observations in which the officer has another officer in their car; Predicted On includes observations where the officer was predicted to be working on that exact day, watch, and unit; Assign = Unit-Watch-Beat (Unit-Watch) replaces the assignment definition of GAB in the main specification, Unit-Watch-Beat Category, with Unit-Watch-Exact Beat (Unit-Watch); Week refers to aggregating the data to the weekly level; Non-Match Level computes leverages at the observation, rather than officer-assignment match, level.

Figure 4: Effect of Officer Quantity versus Quality



Note: Figure displays comparison between the effect of 1 additional officer added to a team using elasticities of crime with respect to number of officers from estimates from Mello (2019) and Chalfin and McCrary (2018) across index crime types ('quantity' effect) compared with the effect of reducing a single average officer's crime effects by 1 SD as recovered in this paper ('quality' effect). Elasticity estimates are converted to percents of the mean by multiplying the elasticity estimates by 1/49 (the percent increase in number of officers by adding 1 officer to the average team of 49 officers). Quality estimates are computed from the KSS bias corrected standard deviations of team effects (Ω), recovered by estimating equations (3) and (6), with each estimate being divided by 49 (the average team size) and divided by the mean of the outcome to provide an approximation of the effect of 1 SD decrease in a single officer's crime effect on an average team. Main specifications are at the district-hour level with controls for year-district-day of week-8 hour period interacted with either month or quarter or using first differences using either actual or predicted team composition. Each bar for the non-quantity estimates represents the median of estimates recovered from each of the three multi-year sample periods. Values of zero indicate that the bias correction produced a zero or negative variance for team effects. Note that if the covariance between officer effects on teams is relatively large in magnitude compared with the variance across officers, then this is a poor approximation. Confidence intervals on the Chalfin and McCrary estimates correspond to converted 95% CIs using standard errors reported in the paper, and confidence intervals on the quality effects denote maximum and minimum values across the samples/specifications.

Table 1: Summary Statistics of Officer Data

	Pooled	By Year	By 2-Year and LLC
	(1)	(2)	(3)
A: Demographics and Experiences			
White	0.484	0.484	0.484
Black	0.233	0.233	0.235
Hispanic	0.246	0.246	0.245
Asian/Native American	0.0367	0.0367	0.0367
Male	0.808	0.808	0.809
Female	0.192	0.192	0.191
Age	41.5	41.5	41.7
Tenure	12.6	12.6	12.8
Military	0.201	0.201	0.196
Military (Active)	0.0121	0.0121	0.0112
Spanish Speaking	0.148	0.148	0.149
Edu >= BA	0.307	0.307	0.306
Edu < BA	0.376	0.376	0.377
Edu Missing	0.317	0.317	0.317
B: By Officer			
Avg. Total GABs	15.4	4.66	6.32
Avg. Total ABs	46.3	10.8	16.1
Avg. Total Units	2.29	1.17	1.31
Avg. Share Most Common GAB	0.457	0.715	0.654
Avg. Share Most Common Unit	0.831	0.971	0.951
Avg. Obs. (Shifts)	1,547	182	351
C: Pooled			
Geographic Unit	0.785	0.785	0.791
Movers - GAB	0.999	0.92	0.963
Movers - Unit	0.844	0.406	0.522
Total Units	54	44.6	45
Total Beats	18,490	5,852	7,325
Total Assignments (AB)	34,898	9,000	11,509
Total Assignments (GABs)	2,879	1,314	1,251
Unique POs	13,400	7,916	8,125
In Car	0.801	0.801	0.805
N Obs.	13,928,210	1,273,145	2,449,965
D: Outcomes (per 100 Shifts)			
Arrest Index	1.59	1.59	1.57
Arrest NonIndex	11	11	10.8
Traffic Stop	12.5	12.5	11.3
Street Stop	34.9	34.9	35.3
Force	0.358	0.358	0.358
Complaints [2016]	0.243	0.243	0.231
Complaints [2019]	0.0963	0.0963	0.0942

Note: Table presents summary statistics for officer-day level observations for days officer are present. Column (1) displays all years pooled, Column (2) first summarizes at the year level, Column (3) first summarises over the 2 year pooled samples in the leave-one-out largest connected component (see Section 3). Age and tenure are in units of years. Geographic units refer to those which map directly to police districts. ABs refer to the interaction between beat, watch, and unit, while GAB refers to generalized assigned beats which uses the type of an assigned beat rather than exact assigned beats to interact with watch and unit (AB). The outcomes section takes the average across officer-day observations and multiplies them by 100 to get the mean per 100 officer-shifts. Panel B computes officer-level statistics then takes the observation-weighted average over officers.

Table 2: Summary Statistics of Teams Data

	2009-2019	2010-2019	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)	(5)
Panel A: Composition of Team (Active Officers)					
Share Black	0.231 (0.222)	0.228 (0.219)	0.251 (0.241)	0.232 (0.22)	0.206 (0.196)
Share Hispanic	0.252 (0.138)	0.255 (0.138)	0.231 (0.14)	0.24 (0.133)	0.286 (0.134)
Share Other Race	0.0419 (0.0437)	0.0424 (0.0436)	0.0391 (0.0442)	0.0415 (0.042)	0.0458 (0.0441)
Share Female	0.208 (0.0831)	0.207 (0.0824)	0.221 (0.0874)	0.205 (0.0791)	0.197 (0.079)
Share S. Age = [25,28]	0.356 (0.0801)	0.359 (0.0796)	0.343 (0.0794)	0.365 (0.0739)	0.367 (0.0819)
Share S. Age = [29,32]	0.233 (0.0676)	0.233 (0.0669)	0.226 (0.0679)	0.231 (0.0631)	0.241 (0.068)
Share S. Age = >33	0.235 (0.078)	0.233 (0.0774)	0.245 (0.078)	0.23 (0.0741)	0.226 (0.0784)
Share Military	0.209 (0.202)	0.223 (0.206)	0.0998 (0.0894)	0.185 (0.165)	0.353 (0.229)
Share Spanish Speaking	0.151 (0.1)	0.15 (0.0985)	0.154 (0.113)	0.148 (0.0959)	0.15 (0.0874)
Share Edu < BA	0.373 (0.105)	0.377 (0.103)	0.345 (0.108)	0.375 (0.0997)	0.406 (0.0918)
Share Edu >= BA	0.298 (0.0949)	0.302 (0.0941)	0.27 (0.0948)	0.299 (0.0888)	0.331 (0.088)
Share Tenure 5-10	0.203 (0.146)	0.196 (0.147)	0.273 (0.151)	0.232 (0.144)	0.105 (0.0846)
Share Tenure 10-15	0.233 (0.116)	0.232 (0.116)	0.269 (0.121)	0.256 (0.113)	0.184 (0.0936)
Share Tenure 15-20	0.217 (0.133)	0.224 (0.132)	0.201 (0.132)	0.246 (0.136)	0.226 (0.126)
Share Tenure >20	0.156 (0.147)	0.163 (0.15)	0.105 (0.11)	0.16 (0.145)	0.214 (0.164)
Panel B: Officer Counts					
Team Size (N. Active POs)	48.8 (18.1)	49.3 (18.2)	46.5 (17.7)	52.9 (19.1)	48.9 (17.4)
N. PO Total	53 (18.7)	53.5 (18.8)	50.5 (18.3)	56.8 (19.5)	53.5 (18.3)
N. PO Reg. Watch	37.1 (14.7)	37.3 (14.9)	37.7 (15.2)	39.9 (15.5)	35 (13.8)
N. PO Tact. Watch	9.75 (7.78)	9.84 (7.75)	8.82 (7.58)	10.5 (8.24)	10.2 (7.42)
N. PO Desk	3.98 (2.08)	3.94 (2.04)	3.65 (1.89)	3.76 (1.74)	4.31 (2.3)
N. PO Overtime - Other	0.934 (2.1)	1.03 (2.19)	0 (0)	0.879 (1.64)	2 (2.91)
N. PO Out of Unit	0.235 (0.705)	0.236 (0.673)	0.311 (0.742)	0.166 (0.47)	0.228 (0.735)
N. PO Overtime - Special	0.272 (1.52)	0.301 (1.6)	0 (0)	0.31 (1.57)	0.541 (2.14)
N. PO Overtime - VRI	0.789 (2.12)	0.873 (2.22)	0 (0)	1.32 (2.85)	1.25 (2.36)
N. Detectives	0.196 (0.508)	0.175 (0.486)	0.463 (0.723)	0.0897 (0.309)	0.00216 (0.0511)
N. Sgt	5.86 (2.27)	5.91 (2.28)	5.65 (2.26)	5.66 (2.05)	6.32 (2.39)
N. Other Non-PO	3.21 (2.02)	3.26 (2.04)	2.82 (1.88)	2.7 (1.73)	4.04 (2.14)
N. Prob/30+	4.28 (5.14)	4.48 (5.3)	1.83 (2.4)	2.32 (2.96)	8.29 (6.1)
Total Teams (District-Hours)	2032449 (0)	1836414 (0)	585304 (0)	539480 (0)	711630 (0)

Note: Table presents summary statistics of team-level data by sample period at the district-hour level, with means and standard deviations (in parentheses). Panel A displays the average shares of demographics of officers counted in the team production model (equation (3)). Panel B contains the set of additional controls including total team size (only counts officers on team) and the other officer counts which include non-team officers.

Table 3: Summary Statistics of Crimes by District-Day-Hour

	Aggregate			Mean by District-Hour			Mean per 100,000 Pop. by District-Hour		
	Total Crimes	Share of Total	Clearance Rate	Crimes	Arrest	Uncleared	Crimes	Arrest	Uncleared
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Index Violent									
Violent	287,382	0.092	0.167	0.141	0.024	0.118	0.135	0.023	0.112
Murder	5,686	0.002	0.377	0.003	0.001	0.002	0.003	0.001	0.002
Robbery	124,105	0.040	0.091	0.061	0.006	0.056	0.059	0.006	0.053
Agg. Assault	140,940	0.045	0.229	0.069	0.016	0.053	0.065	0.015	0.050
Index Property									
Property	994,335	0.317	0.096	0.489	0.047	0.442	0.499	0.047	0.452
Burglary	172,661	0.055	0.054	0.085	0.005	0.080	0.090	0.005	0.085
Motor Vehicle Theft	125,695	0.040	0.070	0.062	0.004	0.058	0.065	0.004	0.060
Theft	691,588	0.222	0.111	0.340	0.038	0.302	0.342	0.038	0.305
Nonindex									
Nonindex	1,856,748	0.592	0.372	0.914	0.340	0.574	0.886	0.311	0.575
Drug	299,120	0.096	0.996	0.147	0.147	0.001	0.129	0.129	0.001
Simple Violent	644,157	0.207	0.216	0.317	0.069	0.248	0.307	0.067	0.240
Weapon	44,574	0.014	0.775	0.022	0.017	0.005	0.020	0.016	0.005
All Other	868,897	0.279	0.251	0.428	0.107	0.320	0.429	0.099	0.330

Note: Table presents summary statistics for crime data between 2009 and 2019 split into three main headers. The aggregate section (Columns (1)-(3)) provides total counts for crimes by type, share of total, and clearance rate. The second header (Columns (4)-(6)), displays statistics at the district-day-hour level, with the mean number of crimes per district-day-hour, the mean number cleared by arrest, and the mean number uncleared. The third header (Columns (7)-(9)) displays the same as Columns (4)-(6) but with crime per 100,000 population.

Table 4: Decomposition of Officer-Level Results

			Share of Var(Y)				
	SD(PO) / Mean(Y)	SD(PO) per 100 Shifts	Total Var(Y) Exp.	Var(PO)	Var(Assign)	2 x Cov(PO, Assign)	Cor(PO, Assign)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Street Stops	0.875	32.9	0.16	0.0797	0.068	0.0122	0.0878
	[0.77, 1.17]	[14.5, 47.1]	[0.151, 0.184]	[0.068, 0.0974]	[0.0593, 0.0961]	[-0.00978, 0.0251]	[-0.0506, 0.18]
Traffic Stops	1.95	16.4	0.223	0.133	0.069	0.00852	0.0337
	[1.08, 2.82]	[12.3, 33.2]	[0.164, 0.263]	[0.109, 0.181]	[0.0249, 0.138]	[-0.0778, 0.0598]	[-0.273, 0.551]
Arrests - Nonindex	0.886	9.93	0.143	0.061	0.0503	0.0256	0.238
	[0.741, 0.952]	[6.65, 11.5]	[0.117, 0.156]	[0.0505, 0.0755]	[0.0364, 0.0678]	[0.0199, 0.035]	[0.155, 0.298]
Arrests - Index	0.775	1.22	0.0221	0.00765	0.0155	-0.00192	-0.0879
	[0.731, 1.12]	[1.01, 1.71]	[0.0189, 0.0253]	[0.00631, 0.0153]	[0.0115, 0.0182]	[-0.0101, 0.00148]	[-0.303, 0.084]
Uses of Force	1.16	0.421	0.00597	0.00462	0.00172	-0.0001	-0.0276
	[1.1, 1.3]	[0.308, 0.476]	[0.00527, 0.00721]	[0.00342, 0.00531]	[0.00116, 0.00233]	[-0.0007, 0.0003]	[-0.118, 0.0562]
Complaints [2016]	0.993	0.231	0.00289	0.00241	0.00176	-0.0009	-0.259
	[0.919, 1.06]	[0.0866, 0.288]	[0.00105, 0.00392]	[0.0009, 0.00262]	[0.00106, 0.00201]	[-0.00156, -0.0005]	[-0.456, -0.122]
Complaints [2019]	1.23	0.124	0.00163	0.00178	0.00117	-0.0006	-0.252
	[0.341, 2.73]	[0.0246, 0.199]	[0.0009, 0.00292]	[0.00008, 0.00446]	[0.0005, 0.0046]	[-0.00799, 0.0004]	[-0.883, 0.907]

Note: Table displays summary of various statistics (columns) by outcome variable (rows). Estimates are bias corrected using KSS. The first row for each statistic refers to the median estimate across rolling 2-year periods and the second row displays the minimum and maximum values. Columns (1) and (2) display the standard deviations of officer effects recovered from estimating equation (1), divided by the mean of the outcome and multiplied by 100 to refer to the effect of a 1 SD increase in officer effects per 100 shifts. Columns (3)-(6) display the components of equation (7) divided by the variance of the outcome, with Column (3) being the summation of Columns (4)-(6). Column (7) is the correlation between officer and assignment effects.

Table 5: Cross-Outcome Correlation in Officer Effects

	Traffic Stops	Arrests - Nonindex	Arrests - Index	Use of Force	Complaints [2016]	Complaints [2019]
	(1)	(2)	(3)	(4)	(5)	(6)
Street Stops	0.162	0.386	0.292	0.164	-0.0694	-0.066
	[-0.005, 0.54]	[0.29, 0.51]	[0.17, 0.4]	[-0.11, 0.3]	[-0.55, 0.057]	[-0.56, 0.62]
Traffic Stops		0.181	0.0909	0.0714	0.0101	0.0262
		[0.091, 0.39]	[-0.071, 0.39]	[-0.082, 0.19]	[-0.28, 0.11]	[-0.19, 0.82]
Arrests - Nonindex			0.412	0.373	0.234	0.18
			[0.22, 0.63]	[0.29, 0.6]	[-0.057, 0.44]	[-0.08, 0.8]
Arrests - Index				0.294	0.0495	0.072
				[0.15, 0.48]	[-0.093, 0.22]	[-0.23, 0.41]
Uses of Force					0.479	0.505
					[0.32, 0.58]	[0.17, 0.9]
Complaints [2016]						0.854
						[0.45, 0.95]

Note: Table presents cross-outcome correlations between officer effects for each outcome. Estimates are bias corrected using KSS. The first row for each outcome refers to the median estimate across rolling 2-year periods and the second row displays the minimum and maximum values.

Table 6: Implied Regression Results of High Quality Arrests on 1 SD in Other Behaviors

	(1)	(2)	(3)
Arrests - Nonindex	0.0104	0.0104	0.0106
	[0.00647, 0.0112]	[0.0073, 0.0121]	[0.00693, 0.0117]
Arrests - Drug	0.0113	0.0111	0.0109
	[0.00408, 0.0123]	[0.00366, 0.0125]	[0.00363, 0.0114]
Arrests - Violent Index	0.00371	0.00367	0.00282
	[0.00226, 0.0053]	[0.0025, 0.00541]	[0.00203, 0.0053]
Arrests - Property Index	0.00204	0.00207	0.00461
	[-0.00213, 0.00385]	[-0.00199, 0.00382]	[-0.00199, 0.00591]
Street Stops	-	-0.00102	-0.0008
	-	[-0.00209, -0.0005]	[-0.00156, 0.0006]
Traffic Stops	-	0.0003	0.0002
	-	[-0.0003, 0.001]	[-0.0003, 0.00115]
Uses of Force	-	-	-0.00146
	-	-	[-0.0064, 0.001]
Complaints [2016]	-	-	0.00252
	-	-	[-0.0007, 0.0069]
Change in R2	0.75	0.00271	0.00754
	[0.661, 0.796]	[0.0007, 0.00879]	[0.00178, 0.0364]
Mean Dep. Var	0.0192	0.0192	0.0192
	[0.0173, 0.0204]	[0.0173, 0.0204]	[0.0173, 0.0204]
SD(Resid PO)/Mean(Y)	0.595	0.586	0.576
	[0.556, 0.643]	[0.552, 0.642]	[0.542, 0.617]

Note: Table displays the implied regression results based on the variance-covariance relationship between officer propensities to make arrests which result in incarceration ('high quality' arrests) and their primary behavioral propensities, with the first row reflecting the median coefficient and the second row containing the minimum and maximum coefficient across 2-year period samples. All coefficients are scaled by 1 SD in the respective officer effect. Resid. PO refers to the residualized outcome (officer propensity to make arrests resulting in incarceration) after removing variation explained by the implied regression, and Mean(Y) refers to the mean of the outcome (average number of arrests resulting in incarceration per shift).

Table 7: Decomposition of Team Effects on Crime

	Mean(Y)	SD(Team)/Mean(Y)		Var(Team)/Var(Y)		Share Var(Y) Explained	
	FE	FE	FD	FE	FD	FE	FD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Violent							
Index Violent	0.136	0.603	0.163	0.0461	0.00169	0.0476	0.00169
	[0.12, 0.14]	[0.37, 0.78]	[0.13, 0.18]	[0.015, 0.077]	[0.0011, 0.0022]	[0.041, 0.066]	[0.0011, 0.0022]
Murder	0.0023	3.38	0.87	0.0335	0.000876	0.00506	0.000876
	[0.0021, 0.0032]	[0, 6.1]	[0.34, 1.3]	[-0.039, 0.081]	[0.00018, 0.0018]	[0.0047, 0.0059]	[0.00018, 0.0018]
Robbery	0.0532	0.6	0.198	0.0182	0.00117	0.0246	0.00117
	[0.053, 0.065]	[0, 1.1]	[0.084, 0.23]	[-0.016, 0.08]	[0.00019, 0.0015]	[0.019, 0.033]	[0.00019, 0.0015]
Agg. Assault	0.0642	0.49	0.184	0.0138	0.0011	0.0329	0.0011
	[0.057, 0.07]	[0, 0.73]	[0.17, 0.21]	[-0.0059, 0.032]	[0.00086, 0.0015]	[0.031, 0.044]	[0.00086, 0.0015]
Property							
Index Property	0.466	0.506	0.162	0.0967	0.00602	0.126	0.00602
	[0.45, 0.56]	[0.35, 0.57]	[0.13, 0.2]	[0.058, 0.13]	[0.0039, 0.011]	[0.12, 0.14]	[0.0039, 0.011]
Burglary	0.0799	0.305	0.222	0.00808	0.00206	0.0412	0.00206
	[0.063, 0.12]	[0, 0.89]	[0.16, 0.28]	[-0.005, 0.085]	[0.00075, 0.0041]	[0.023, 0.078]	[0.00075, 0.0041]
Motor Vehicle Theft	0.0545	0.786	0.231	0.0402	0.00173	0.0229	0.00173
	[0.053, 0.085]	[0.45, 1.5]	[0.15, 0.27]	[0.016, 0.12]	[0.0006, 0.0027]	[0.019, 0.033]	[0.0006, 0.0027]
Theft	0.331	0.635	0.177	0.108	0.00521	0.122	0.00521
	[0.33, 0.36]	[0.46, 0.69]	[0.15, 0.22]	[0.064, 0.13]	[0.0035, 0.0087]	[0.11, 0.17]	[0.0035, 0.0087]
Nonindex							
Drug	0.14	1.23	0.432	0.151	0.00979	0.184	0.00979
	[0.069, 0.17]	[1.2, 1.7]	[0.25, 0.64]	[0.12, 0.19]	[0.0051, 0.017]	[0.14, 0.22]	[0.0051, 0.017]
Simple Violent	0.295	0.483	0.129	0.0653	0.00258	0.0774	0.00258
	[0.29, 0.32]	[0.41, 0.58]	[0.11, 0.14]	[0.044, 0.09]	[0.0017, 0.003]	[0.062, 0.1]	[0.0017, 0.003]
Weapon	0.0176	1.26	0.308	0.0334	0.000954	0.0177	0.000954
	[0.016, 0.026]	[0, 2.1]	[0.25, 0.36]	[-0.046, 0.076]	[0.00055, 0.0015]	[0.015, 0.03]	[0.00055, 0.0015]
All Other	0.415	0.437	0.234	0.0664	0.0107	0.0803	0.0107
	[0.41, 0.44]	[0.27, 0.49]	[0.22, 0.25]	[0.027, 0.084]	[0.0091, 0.012]	[0.064, 0.091]	[0.0091, 0.012]

Note: Table displays summary of various statistics (columns) by outcome variable (rows) which are the number of crimes per district-hour. Estimates are recovered from estimating equations (3) for FE under two different district-time fixed effects (district-year-quarter(month)-day of week -8 hour window) and (6) for FD using actual and predicted team membership, on three samples (2010-2012, 2013-2015, 2016-2019). The first row for each outcome is the median estimate across sample groups while the second row contains the minimum and maximum across sample groups. Zero or negative values in Columns reflects plug-in estimates being entirely attributable to bias.

Table 8: Effect of Team Average Officer Effects on Crimes (1 SD / Mean Y)

	Violent Crime				Property Crime				Drug Crime			
	OLS		IV		OLS		IV		OLS		IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A: 1 SD(Team Avg. Shrunken Officer FEs)												
Street Stops	-0.001	0.001	-0.014	-0.013	-0.006**	-0.005	-0.005	-0.004	0.003	-0.004	0.012	0.011
	(0.006)	(0.007)	(0.013)	(0.017)	(0.003)	(0.003)	(0.007)	(0.009)	(0.006)	(0.008)	(0.015)	(0.019)
Traffic Stops	-0.003	-0.002	0.056**	0.016	-0.002	-0.002	-0.02	-0.016	0.012	0.008	0.03	0.014
	(0.008)	(0.007)	(0.026)	(0.023)	(0.004)	(0.004)	(0.012)	(0.011)	(0.007)	(0.006)	(0.021)	(0.019)
Arrests - Drug	-0.008	0.001	0.006	0.018	-0.005*	-0.005	-0.003	-0.004	0.222***	0.338***	0.257***	0.452***
	(0.006)	(0.009)	(0.015)	(0.025)	(0.003)	(0.004)	(0.007)	(0.011)	(0.009)	(0.013)	(0.022)	(0.036)
Arrests - Non-Drug Nonindex	-0.004	-0.015*	-0.042**	-0.072**	-0.004	-0.01**	0.012	0.012	-0.029***	-0.053***	-0.064***	-0.088***
	(0.007)	(0.009)	(0.018)	(0.028)	(0.003)	(0.005)	(0.009)	(0.014)	(0.008)	(0.011)	(0.02)	(0.031)
Arrests - Violent Index	0.004	0.007	0	0.002	0.002	0.005	0	0.007	-0.024***	-0.05***	-0.065***	-0.084***
	(0.006)	(0.006)	(0.015)	(0.016)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)	(0.008)	(0.017)	(0.019)
Arrests - Property Index	-0.001	-0.007	0.003	0.003	-0.001	-0.003	0.008	0.003	-0.001	-0.007	-0.01	-0.063**
	(0.006)	(0.009)	(0.019)	(0.028)	(0.004)	(0.006)	(0.01)	(0.016)	(0.006)	(0.009)	(0.017)	(0.027)
Uses of Force	-0.003	0.001	-0.002	-0.008	-0.001	-0.004	-0.017**	-0.016	0.029***	0.04***	0.051***	0.083***
	(0.007)	(0.009)	(0.019)	(0.023)	(0.003)	(0.004)	(0.008)	(0.01)	(0.007)	(0.01)	(0.019)	(0.024)
Complaints	0.004	0.003	-0.011	0.005	-0.002	-0.004	0.029***	0.012	0.014*	-0.003	0.037	-0.021
	(0.008)	(0.009)	(0.023)	(0.024)	(0.004)	(0.004)	(0.011)	(0.012)	(0.009)	(0.011)	(0.024)	(0.028)
Mean DV	0.132	0.132	0.132	0.132	0.491	0.491	0.492	0.492	0.123	0.123	0.123	0.123
N. Obs	1836414	1836414	1795250	1795250	1836414	1836414	1795250	1795250	1836414	1836414	1795250	1795250
FE Type	Raw	Shrunken	Raw	Shrunken	Raw	Shrunken	Raw	Shrunken	Raw	Shrunken	Raw	Shrunken

Note: Table displays OLS and IV estimates recovered from estimating equation (10) by regressing the number of reported crimes at the district-hour level on team composition as measured by standard deviations of team averages of officer effects on activities. The sample used is the pooled 2010-2019 period. District-Time controls used are District-Year-Month-Day of Week-Hour. Estimates are divided by the mean of the outcome variable. The first stage (equation (11)) for the IV uses predicted team compositions based on the operations calendar rotation. All specifications control for polynomials of officer counts and team sizes, and control for exact-date fixed effects. Differences in sample size between OLS and IV are due to missing predicted teams for some observations. Standard errors in parentheses are clustered at the District-Year-Month-Day of Week-Hour level. *** p<0.01, ** p<0.05, * p<0.1. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Elasticities of Crime with Respect to Policing

Officer Activity	Index Violent		Index Property	
	Endo. Only	+ Pred. Controls	Endo. Only	+ Pred. Controls
	(1)	(2)	(3)	(4)
N. Arrests	-0.25*	-0.2	0.025	0.07
	[-0.52, 0.03]	[-0.48, 0.079]	[-0.099, 0.15]	[-0.06, 0.2]
N. Nonindex Arrests	-0.16	-0.14	0.00069	0.017
	[-0.37, 0.052]	[-0.32, 0.048]	[-0.092, 0.093]	[-0.071, 0.1]
N. QOL Arrests	-0.12*	-0.067	0.0028	-0.0063
	[-0.24, 0.012]	[-0.18, 0.046]	[-0.055, 0.06]	[-0.06, 0.048]
N. Street Stops	-0.04	-0.022	-0.027	-0.021
	[-0.15, 0.066]	[-0.13, 0.089]	[-0.079, 0.025]	[-0.076, 0.033]
N. Traffic Stops	0.017	0.043	-0.033	-0.036
	[-0.087, 0.12]	[-0.064, 0.15]	[-0.082, 0.016]	[-0.086, 0.015]

Note: Table displays the elasticities of index violent and property crime with respect to different policing activities, with 95% confidence intervals displayed in brackets below, over the pooled sample (2010-2019). Elasticity estimates and confidence intervals are computed by estimating equation (12) regressing the number of crimes which occurred in a specific district-hour on the number of policing activities (e.g., street stops) and instrumenting for the policing activity using the predicted average team propensity for that activity based on the rotation calendar (equation (13)), then are divided by the mean of the outcome and multiplying by the mean of the activity. Even columns control for other predicted team average propensities in the first and second stage. Predicted officer effects used to compute team average predicted effects are shrunken and computed separately over 2010-2012, 2013-2015, and 2016-2019. Standard errors used to compute confidence intervals are clustered at the year-month-day of week-district-hour level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 10: Projecting Officer Effects onto Officer Observables (scaled by Mean(Y))

	$\hat{\theta}_i^k / Mean(Y^k)$						
	Street Stops	Traffic Stops	Arrests - Nonindex	Arrests - Index	Uses of Force	Complaints [2016]	Complaints [2019]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Demographics							
Black	-0.12***	-0.05**	-0.08***	-0.06***	-0.16***	-0.1**	-0.07
	[-0.15***, -0.02]	[-0.09***, 0.04]	[-0.11***, -0.05***]	[-0.08**, -0.02]	[-0.29***, -0.07*]	[-0.14***, -0.002]	[-0.24***, 0.14*]
Hispanic	-0.03***	0.005	-0.02***	-0.02	-0.08***	-0.06**	-0.006
	[-0.06***, -0.01]	[-0.04, 0.06***]	[-0.04***, -0.003]	[-0.05***, 0.0009]	[-0.12***, -0.05**]	[-0.08***, -0.003]	[-0.1**, 0.08]
Other Race	-0.02**	0.001	-0.02***	-0.008	-0.04***	-0.02	-0.03
	[-0.03***, -0.008]	[-0.02***, 0.06**]	[-0.04***, -0.004]	[-0.02, 0.03**]	[-0.08***, -0.02]	[-0.04**, 0.02]	[-0.04*, 0.06]
Female	-0.07***	-0.07***	-0.1***	-0.05***	-0.19***	-0.13***	-0.16***
	[-0.08***, -0.04***]	[-0.09***, -0.02*]	[-0.16***, -0.07***]	[-0.12***, -0.009]	[-0.22***, -0.14***]	[-0.15***, -0.1***]	[-0.2***, -0.04]
S. Age [25,28]	-0.009	0.07***	-0.02*	-0.02	-0.06***	-0.008	0.007
	[-0.02, 0.01]	[-0.002, 0.15***]	[-0.04***, -0.002]	[-0.04*, 0.008]	[-0.12*, 0.006]	[-0.02, 0.02]	[-0.04, 0.04]
S. Age [29,32]	-0.02**	0.09***	-0.02***	-0.04*	-0.11***	-0.05**	-0.04
	[-0.05***, 0.01]	[0.005, 0.21***]	[-0.06***, -0.01]	[-0.06***, 0.03]	[-0.14***, -0.01]	[-0.06***, -0.02]	[-0.07**, 0.1]
S. Age >33	-0.03***	0.1***	-0.07***	-0.06***	-0.18***	-0.09***	-0.11***
	[-0.05***, 0.02**]	[-0.009, 0.17***]	[-0.1***, -0.04***]	[-0.09***, -0.04**]	[-0.22***, -0.1***]	[-0.11***, -0.09***]	[-0.17***, 0.05]
Experiences							
Spanish	-0.003	-0.02	-0.004	0.007	0.009	0.03	0.004
	[-0.009, 0.03***]	[-0.04, 0.03]	[-0.02*, 0.002]	[-0.01, 0.04**]	[-0.005, 0.05**]	[-0.007, 0.09***]	[-0.08**, 0.12]
Military	0.06***	0.08***	0.11***	0.09***	0.08***	-0.007	-0.04
	[0.03, 0.11***]	[-0.05***, 0.12***]	[0.05***, 0.19***]	[0.05***, 0.19***]	[0.02, 0.16***]	[-0.04, 0.04]	[-0.15*, 0.16*]
Military (Active)	0.006	0.03	0.02**	0.03**	0.002	-0.02	-0.008
	[0.0007, 0.03***]	[-0.02, 0.06]	[0.002, 0.06***]	[0.02, 0.06*]	[-0.03**, 0.17***]	[-0.04***, 0.02]	[-0.05*, 0.15**]
Edu >= BA	0.04***	0.02	-0.004	0.02	-0.02	-0.03	-0.13
	[0.02*, 0.07***]	[-0.01, 0.08***]	[-0.02***, 0.01*]	[-0.01, 0.06***]	[-0.05**, 0.04]	[-0.08***, 0.003]	[-0.15***, -0.04]
Edu < BA	0.002	0.003	-0.008	0.02	-0.04	-0.02	-0.11
	[-0.009, 0.05***]	[-0.04, 0.04]	[-0.03***, 0.01]	[0.01, 0.04**]	[-0.07***, 0.03]	[-0.08***, 0.03]	[-0.17***, 0.05]
Tenure							
5-10	-0.09***	-0.13***	-0.04***	-0.03	-0.18***	-0.04	-0.06
	[-0.14***, 0.007]	[-0.17***, -0.01]	[-0.06***, 0.04**]	[-0.17, 0.1***]	[-0.22***, 0.16]	[-0.08, -0.01]	[-0.38, 0.01]
10-15	-0.15***	-0.24***	-0.11***	-0.09***	-0.26***	-0.17***	-0.21
	[-0.21***, -0.04*]	[-0.28***, -0.06]	[-0.2***, -0.02]	[-0.27*, 0.07]	[-0.39***, 0.05]	[-0.19***, 0.01]	[-0.39, -0.12]
15-20	-0.2***	-0.3***	-0.15***	-0.14***	-0.38***	-0.18**	-0.27
	[-0.23***, -0.08***]	[-0.36***, -0.12**]	[-0.22***, -0.08***]	[-0.35**, 0.00005]	[-0.47***, -0.003]	[-0.25***, -0.1]	[-0.55***, -0.18]
>20	-0.21***	-0.37***	-0.18***	-0.19***	-0.38***	-0.22***	-0.26*
	[-0.28***, -0.16***]	[-0.46***, -0.13**]	[-0.26***, -0.1***]	[-0.39**, -0.1**]	[-0.48***, 0.001]	[-0.23***, -0.08]	[-0.45***, -0.06]
R2	0.08	0.03	0.11	0.04	0.08	0.02	0.01
	[0.06, 0.12]	[0.006, 0.23]	[0.08, 0.13]	[0.02, 0.07]	[0.02, 0.1]	[0.02, 0.03]	[0.005, 0.02]

Note: Table presents results from estimating equation (14) on officer effects scaled by the mean of the relevant outcome. Rows reflect the coefficient on each officer-specific variable, with the first row reflecting the median coefficient and the second row containing the minimum and maximum coefficient across 2-year period samples. S. Age refers to the officer's starting age in the CPD, Edu refers to their education level. The reference group is white male officers starting younger than 25, who do not speak Spanish, were not in the military, with missing educational information, and less than 5 years of tenure. Regressions exclude probationary officers with tenure less than 1.5 years. Standard errors are heteroskedastic robust adjusted using KSS and are used to compute p-values. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Effect of Changing 1 Officer over 200 Shifts

	Officer Activity							Index Crimes	
Period	Street Stops	Traffic Stops	A. Nonindex	A. Index	Force	Cmpl. [2016]	Cmpl. [2019]	Violent	Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black									
2010-2012	-25***	-0.88	-3.9***	-0.44***	-0.18***	-0.11***	-0.052**	0.67	-0.97
2013-2015	-31***	-0.16	-4.9***	-0.4***	-0.29***	-0.053*	-0.0021	1.1	-0.67
2016-2019	-2.4**	5.1***	-2.1***	-0.43***	-0.21***	-	-0.01	-0.61	0.39
Hispanic									
2010-2012	-7.7***	0.79	-1**	-0.14	-0.13***	-0.023	-0.0083	1.4	-1.9
2013-2015	-7.9***	-0.13	-2.5***	-0.076	-0.12***	-0.058***	-0.011	0.26	2.9
2016-2019	-0.84	4.3***	-0.55*	-0.16**	-0.092***	-	-0.00085	0.23	1.5
Female									
2010-2012	-8.3***	-3***	-6.6***	-0.39***	-0.35***	-0.13***	-0.096***	-1.3	3
2013-2015	-15***	-0.47	-5.3***	-0.41***	-0.38***	-0.16***	-0.079***	-1.1	0.041
2016-2019	-3.8***	-4.8***	-2.7***	-0.3***	-0.26***	-	-0.039***	-0.38	3.8**
Tenure 5-10									
2010-2012	-13***	-3.9***	-4.7***	-0.32***	-0.3***	-0.046	0.01	0.15	1.1
2013-2015	-30***	-3.7***	0.54	0.4	-0.32***	-0.028	-0.078	-1.5	4.9
2016-2019	-6.5***	-16***	-3.5***	-0.32*	-0.23***	-	-0.0098	-1.3	2.8
Tenure 10-15									
2010-2012	-20***	-7.4***	-8.2***	-0.73***	-0.54***	-0.18***	-0.095***	-2.3	11***
2013-2015	-50***	-5.2***	-2.7***	-0.16	-0.48***	-0.071	-0.098*	-3.5	6.1
2016-2019	-8.7***	-24***	-5.1***	-0.6***	-0.29***	-	-0.026	1.3	2.7
Tenure 15-20									
2010-2012	-28***	-11***	-12***	-1.1***	-0.7***	-0.22***	-0.15***	-2	12***
2013-2015	-55***	-6***	-6***	-0.6**	-0.71***	-0.15***	-0.16**	-3.5	10**
2016-2019	-13***	-31***	-6.6***	-0.98***	-0.38***	-	-0.04*	-1.1	0.97
Tenure >20									
2010-2012	-41***	-15***	-17***	-1.8***	-0.79***	-0.3***	-0.17***	-4**	9.4**
2013-2015	-66***	-9.2***	-7.7***	-1.3***	-0.82***	-0.2***	-0.19***	-4	10**
2016-2019	-14***	-30***	-8.4***	-1.5***	-0.42***	-	-0.013	-1.9	2.1

Note: Table displays the estimated change in officer activities and index crimes of due to officer-level characteristics based on the results of projecting officer effects on activities and crimes (equation (14) and (15)) over 200 9-hour shifts, or around 1 full year, assuming each officer works on a team with the average size during that sample period. Officer characteristics displayed are a subset of the full ones used in the regressions, see Table 11 for all characteristics. The reference group is white male officers starting younger than 25, who do not speak Spanish, were not in the military, with missing educational information, and less than 5 years of tenure. P-values, represented by stars, are computed using KSS adjusted heteroskedasticity robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

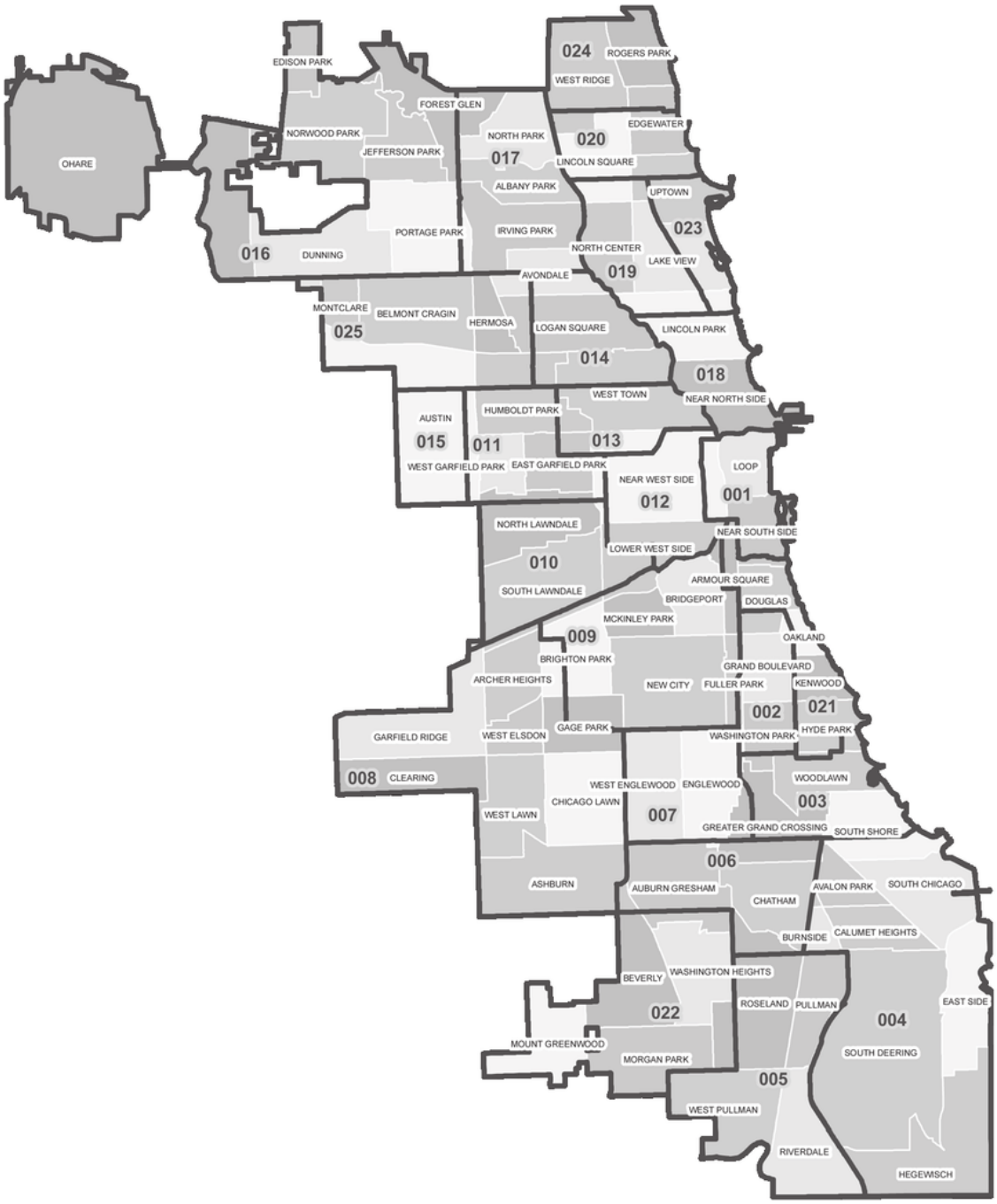
Table 12: Effect of Team Demographic Composition on Crimes (1 SD / Mean Y)

	Violent		Property		Drug		Simple Violent		Weapon		Other Nonindex	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Demographics												
Share Black	0.019	0.024	-0.007	-0.024	-0.049***	-0.119***	0.017**	-0.002	-0.069**	0.159	0.01	0.04
	(0.013)	(0.044)	(0.006)	(0.02)	(0.013)	(0.041)	(0.008)	(0.028)	(0.031)	(0.136)	(0.007)	(0.025)
Share Hispanic	0.004	-0.012	-0.001	-0.018	0.023***	-0.009	0.003	0.018	-0.004	0.015	-0.001	0
	(0.008)	(0.028)	(0.004)	(0.013)	(0.008)	(0.026)	(0.005)	(0.017)	(0.02)	(0.076)	(0.005)	(0.016)
Share Other Race	0.002	-0.014	-0.004*	-0.007	0.01***	0.017*	0.001	-0.005	-0.014	-0.015	0.001	0.011*
	(0.004)	(0.012)	(0.002)	(0.005)	(0.004)	(0.009)	(0.003)	(0.007)	(0.009)	(0.028)	(0.002)	(0.006)
Share Female	0.005	0.013	0.007***	0.009	-0.012***	-0.007	0.003	0.005	-0.013	-0.007	0.002	0.006
	(0.004)	(0.015)	(0.002)	(0.007)	(0.004)	(0.013)	(0.003)	(0.009)	(0.01)	(0.041)	(0.002)	(0.008)
Share S. Age = [25,28]	0.001	-0.007	0	-0.007	-0.013***	-0.024	0.002	0.001	0.004	0.033	0.001	0.003
	(0.005)	(0.016)	(0.002)	(0.008)	(0.005)	(0.015)	(0.003)	(0.01)	(0.012)	(0.046)	(0.003)	(0.009)
Share S. Age = [29,32]	0.003	-0.008	-0.001	-0.004	-0.01**	-0.023*	0.001	0.003	0.013	0.059	0	0.002
	(0.004)	(0.013)	(0.002)	(0.006)	(0.004)	(0.012)	(0.003)	(0.009)	(0.01)	(0.037)	(0.002)	(0.008)
Share S. Age = >33	0.003	-0.007	0.001	0	-0.019***	-0.029*	0.001	-0.011	-0.01	0.03	0.001	0.003
	(0.005)	(0.016)	(0.003)	(0.007)	(0.005)	(0.015)	(0.003)	(0.01)	(0.012)	(0.045)	(0.003)	(0.009)
Experiences												
Share Military	-0.008	-0.04	0.003	0.017	-0.06***	-0.075*	0.003	0.012	0.018	0.109	0.01	-0.029
	(0.015)	(0.049)	(0.007)	(0.022)	(0.016)	(0.045)	(0.01)	(0.031)	(0.038)	(0.146)	(0.008)	(0.027)
Share Spanish Speaking	0	0.006	-0.003	0.018*	-0.007	-0.062***	0	0.011	-0.037**	-0.031	-0.002	-0.006
	(0.007)	(0.022)	(0.003)	(0.011)	(0.007)	(0.021)	(0.004)	(0.013)	(0.016)	(0.061)	(0.004)	(0.012)
Share Edu < BA	-0.006	-0.019	0.002	0.002	-0.018***	-0.031**	0.005	-0.002	-0.01	0.038	-0.01***	-0.014
	(0.006)	(0.017)	(0.003)	(0.008)	(0.006)	(0.016)	(0.004)	(0.011)	(0.014)	(0.052)	(0.003)	(0.01)
Share Edu >= BA	0	-0.014	0.007**	0.007	-0.022***	-0.041***	0.006*	0.019	-0.029**	-0.034	0.001	0.005
	(0.005)	(0.017)	(0.003)	(0.009)	(0.005)	(0.016)	(0.004)	(0.011)	(0.013)	(0.051)	(0.003)	(0.01)
Tenure												
Share Tenure 5-10	-0.003	0.02	0.001	0.018	-0.022*	-0.067*	-0.023***	-0.024	-0.018	0.118	0.01*	-0.001
	(0.011)	(0.038)	(0.005)	(0.017)	(0.011)	(0.034)	(0.007)	(0.023)	(0.029)	(0.111)	(0.006)	(0.021)
Share Tenure 10-15	0.006	0.018	0.005	0.032*	-0.067***	-0.064*	-0.008	-0.004	-0.013	0.143	-0.004	-0.038*
	(0.01)	(0.036)	(0.005)	(0.017)	(0.01)	(0.033)	(0.006)	(0.022)	(0.024)	(0.112)	(0.005)	(0.021)
Share Tenure 15-20	0.013	0.029	0.016***	0.035*	-0.077***	-0.073*	-0.006	0.014	-0.021	0.122	0.008	-0.02
	(0.012)	(0.042)	(0.006)	(0.019)	(0.012)	(0.038)	(0.008)	(0.026)	(0.029)	(0.124)	(0.006)	(0.024)
Share Tenure >20	0.02	0.01	0.028***	0.049*	-0.104***	-0.11**	0.012	0.02	-0.015	0.135	-0.004	-0.081**
	(0.015)	(0.055)	(0.008)	(0.026)	(0.015)	(0.05)	(0.01)	(0.036)	(0.037)	(0.168)	(0.008)	(0.032)
Mean DV	13.22	13.23	49.08	49.21	12.29	12.32	30.24	30.21	2.03	2.02	41.88	41.9
N. Obs	1836414	1796191	1836414	1796191	1836414	1796191	1836414	1796191	1836414	1796191	1836414	1796191

Note: Table displays OLS and IV estimates recovered from estimating equation (16) by regressing the number of reported crimes at the district-hour level on team composition as measured by standard deviations of team shares of officer demographics. The sample used is the pooled 2010-2019 period. District-Time controls used are District-Year-Month-Day of Week-Hour. Estimates are divided by the mean of the outcome variable. The first stage (equation (17)) for the IV uses predicted team compositions based on the operations calendar rotation. All specifications control for polynomials of officer counts and team sizes, and control for exact-date fixed effects. Differences in sample size between OLS and IV are due to missing predicted teams for some observations. Standard errors in parentheses are clustered at the District-Year-Month-Day of Week-Hour level. *** p<0.01, ** p<0.05, * p<0.1.

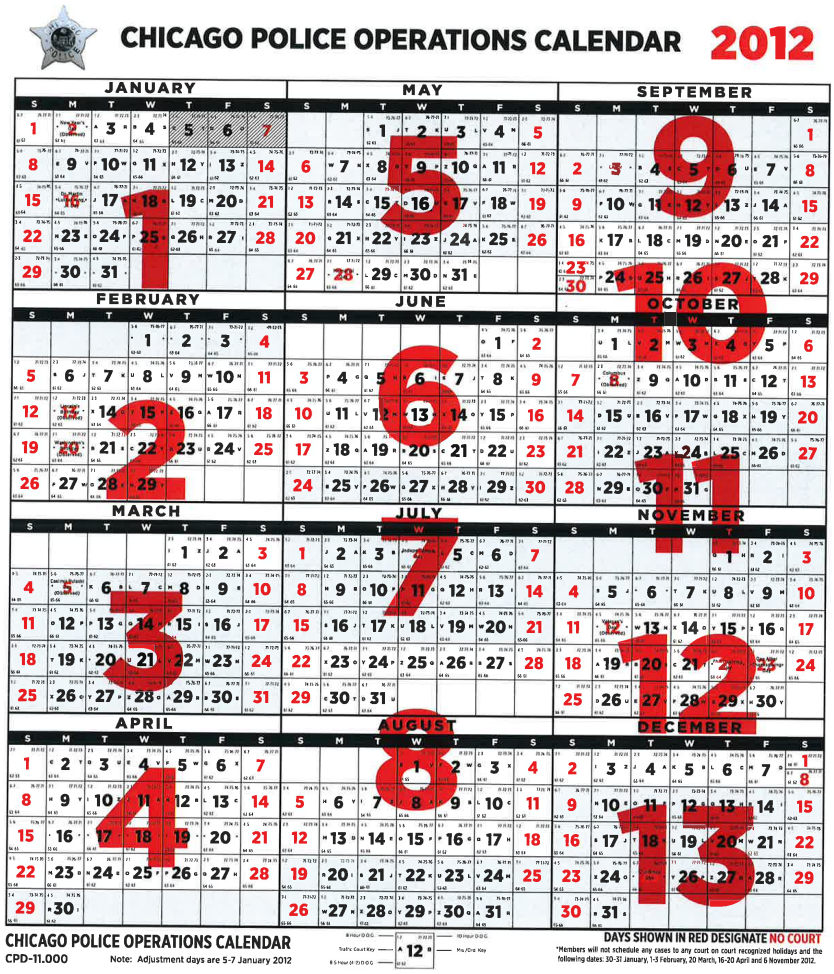
A Appendix A: Background and Data Figures and Tables

Figure A.1: Map of CPD Districts (2009) and Neighborhoods from Markovic (2009)



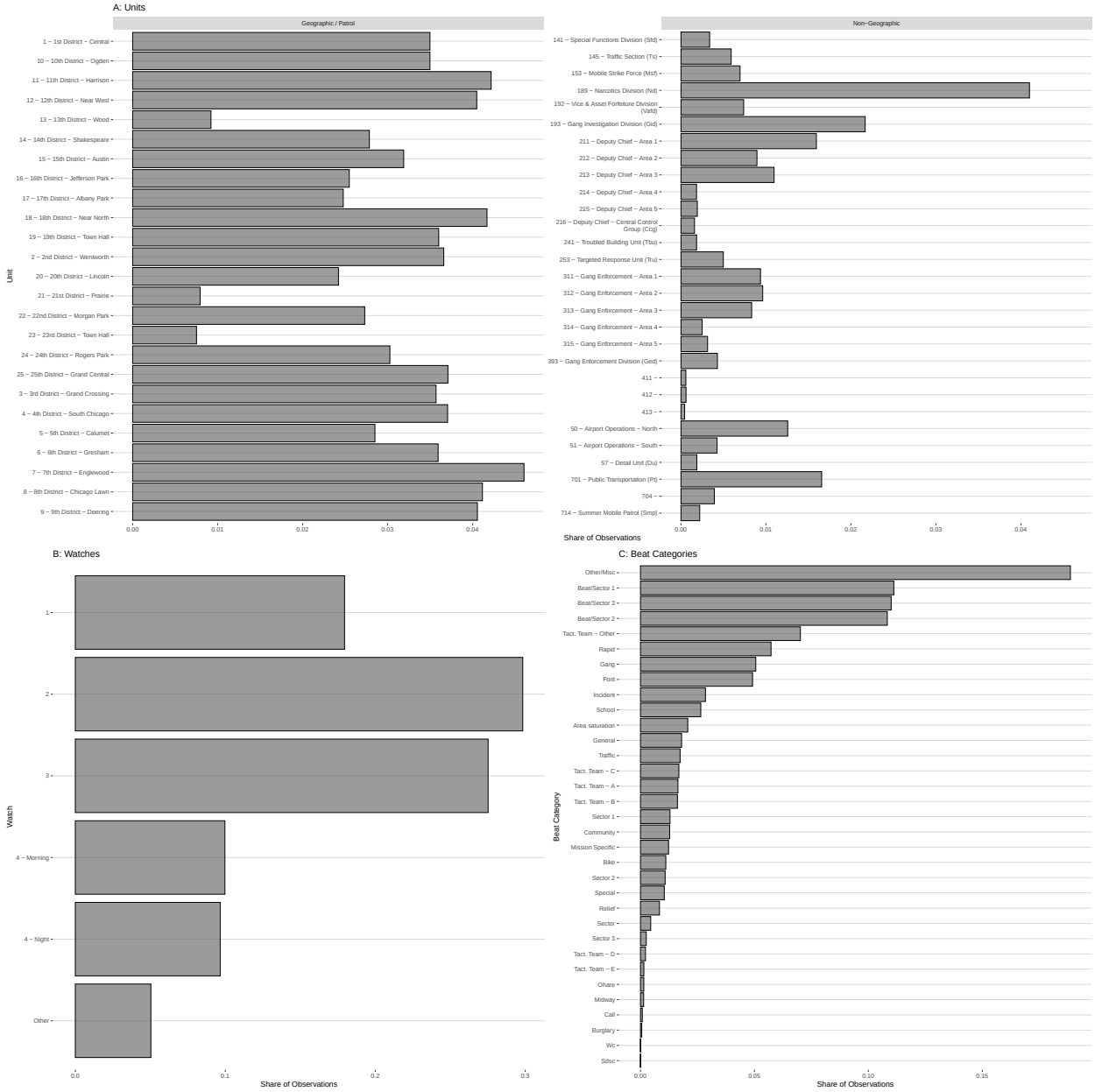
Note: Figure displays the Chicago Police Department districts numbered 1-25 from 2009 with labels for Chicago neighborhoods, sourced from Markovic (2009). In 2012, districts 13, 21, and 23 were absorbed by other districts mainly 12, 2, and 19, in the post-2012 district map.

Figure A.2: CPD Operations Calendar (2012)



Note: Figure displays an example of the CPD operations calendar for the year 2012.

Figure A.3: Share of Observations by Unit, Watch, and Beat Category



Note: Figure displays the share of observations in the analysis sample that corresponds to each unit (Panel A) separated by patrol units which map directly to police districts and non-geographic units which do not map to specific districts, each watch (Panel B) with watch 4 broken into morning and night corresponding to start times and all other watches beyond 1-4 pooled into "Other", and each beat category (Panel C). The interaction of unit, watch, and beat category produces a unique generalized assigned beat (GAB).

B Appendix B: Detailed Discussion of Internal Validity

B.1 Switching and Connectivity

For officer-level outcomes, the identifying variation comes from officers working in different assignments across time. As Abowd et al. (2002) shows, officer and assignment effects can be identified for officers and assignments within a connected network, which depends on the structure of the bipartite graph produced by the connections between firms and workers. So, identification of officer and assignment effects requires that we observe officers moving between assignments with sufficient variety and frequency. Even within the connected set, if mobility (the degree of switching between officers and assignments) is low, then variance and covariance measurements from these identified effects may be biased due to weak connectivity or "limited mobility bias" (Andrews et al. (2008), Jochmans & Weidner (2019), Bonhomme et al. (2022)).[23]

As shown in Table 1, officers work in a relatively large number of assignments (GABs) within a given year (over 4 GABs) but generally only work in one unit (1.17 per year). Across unit switches are more rare, with the average share of the most common GAB being 72% of an officer's observations within a year while the most common unit is 97% of observations in a given year.

Naturally, the best metric for the degree of connectivity is the size of the largest connected component (connected set) constructed from the network between officers and assignments. I follow the current literature pruning the network to find the largest connected component that is robust to the removal of any single officer, as in Kline et al. (2020) and Bonhomme et al. (2022). I will refer to this leave-one-out largest connected component as the "LLC" for brevity.

The frequency of switches between assignments results in almost all observations ($> 95\%$)

[23]Simple OLS regressions with officer and assignments fixed effects do not generally estimate the model. Effectively, because the matrix of officer and assignment fixed effects ($Z = [\theta', \psi']$) is sparse, in practices $Z'Z$ will often not have full rank.

belong to the LLC, as displayed in Table B.1. Similarly, the sample of only geographic units (1-25) and regular watches (1-4) accounts for around 76% of observations, while its LLC is still over 96% of the remaining sample. One concern may be that the high connectivity is only due to very short spells in different assignments and unit-watches which produce a connected network but the results are dependent upon officers switching due to short-term shocks. Table B.1 also shows the size of the sample and LLC share after filtering for observations with assignment spells at least 4 days and unit-watch spells at least 20 days, which correspond to $> 71\%$ and $> 85\%$ of the full sample, and their LLC's maintain over 95 of the resulting subsamples suggesting short-switches are not driving the connectivity. All LLC's are over 90% of their respective sample, with the exception of using unit-watch-beat assignment fixed effects, which results in only around 40% of the sample being kept in the LLC. On the other hand, the total samples for predicted on, high-outcome, and no partner samples are relatively small, generally around 35-50% of the complete sample being included — note that small samples for these in 2009 are due to a lack of stops data and a lack of predicted assignments for that year (they begin in 2010).

Table B.1: Samples and Connected Sets by 2 Year Periods

	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total Counts										
N. Shifts	2,733,785	2,701,257	2,698,313	2,643,135	2,593,014	2,590,857	2,518,356	2,369,998	2,246,574	2,240,870
N. Unique POs	8,715	8,787	8,638	8,363	8,518	8,441	8,374	8,294	8,263	8,677
N. Unique GABs	1,544	1,469	1,433	1,364	1,591	1,644	1,700	1,647	1,586	1,609
Main (2 Year)										
Share of Total	1	1	1	1	1	1	1	1	1	1
LLC Share of Sample	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.96	0.96	0.96
Regular Units/Watches										
Share of Total	0.75	0.76	0.78	0.78	0.77	0.77	0.77	0.78	0.76	0.75
LLC Share of Sample	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97
Long Unit-Watch										
Share of Total	0.86	0.86	0.87	0.87	0.87	0.88	0.88	0.88	0.88	0.87
LLC Share of Sample	0.97	0.97	0.96	0.97	0.97	0.95	0.96	0.96	0.96	0.96
Long GAB										
Share of Total	0.76	0.74	0.72	0.73	0.73	0.73	0.72	0.73	0.74	0.73
LLC Share of Sample	0.98	0.98	0.98	0.97	0.98	0.96	0.95	0.97	0.96	0.96
Donut										
Share of Total	0.7	0.68	0.67	0.68	0.68	0.68	0.67	0.68	0.69	0.68
LLC Share of Sample	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.97	0.97	0.97
High Outcome										
Share of Total	0.35	0.56	0.58	0.64	0.66	0.63	0.47	0.35	0.41	0.46
LLC Share of Sample	1	1	0.99	1	0.99	0.98	0.99	1	0.99	1
No Partners										
Share of Total	0.39	0.38	0.38	0.38	0.38	0.39	0.38	0.38	0.39	0.39
LLC Share of Sample	0.96	0.96	0.96	0.96	0.96	0.95	0.94	0.95	0.95	0.95
Predicted On										
Share of Total	0.29	0.66	0.74	0.73	0.72	0.71	0.7	0.7	0.68	0.67
LLC Share of Sample	0.97	0.98	0.98	0.97	0.98	0.97	0.97	0.97	0.96	0.96
Assign = Unit-Watch-Beat										
Share of Total	1	1	1	1	1	1	1	1	1	1
LLC Share of Sample	0.38	0.43	0.38	0.37	0.43	0.44	0.36	0.38	0.42	0.43
Assign = Unit-Watch										
Share of Total	1	1	1	1	1	1	1	1	1	1
LLC Share of Sample	1	1	1	1	1	1	0.99	1	1	1
Match Effects										
Share of Total	1	1	1	1	1	1	1	1	1	1
LLC Share of Sample	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.96	0.96	0.96

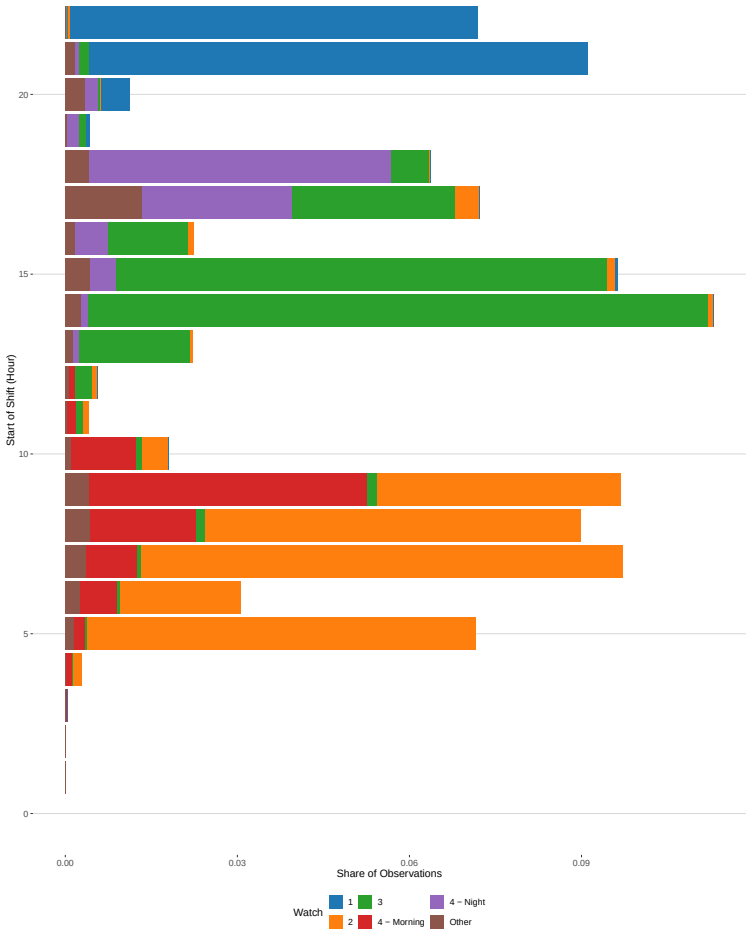
Note: Table presents summary of observations, total unique officers, and total unique assignments (GABs) in each 2-year sample period, and the relative sizes of subsamples used along with the relative sizes of the largest leave-one-out connected component in terms of share of sample observations. Main (2 Year) refers to the main specification with 10 2-year rolling periods; 1 Year Period refers to each year of data analyzed separately; Geographic refers to the main specification but subsetting to geographic units only; 3-4 Year period refers to estimating the model using 3 distinct periods which cover 2009-2012, 2013-2015, 2016-2019; Long Unit-Watch (GAB) refers to selecting for spells in which the officer spends at least 20 (4) consecutive days in the unit-watch (GAB); Donut refers to removing the first and last day of each officer-GAB spell; High Outcome selects for GABs which have average outcomes (all main outcomes summed) of at least 0.25 within a year; No Partners removes observations in which the officer has another officer in their car; Predicted On includes observations where the officer was predicted to be working on that exact day, watch, and unit based on the rotation as discussed in Section 4; Assign = Unit-Watch-Beat (Unit-Watch) replaces the assignment definition of GAB in the main specification, Unit-Watch-Beat Category, with Unit-Watch-Exact Beat (Unit-Watch); Week refers to aggregating the data to the weekly level; Match Effects interacts officer and beat types for officer effects and unit and watch effects for assignment effects. Match effects uses the Main (2 Year) sample and LLCs.

For team-level outcomes, the variation comes from officers working in overlapping groups due to start and end times for the same watch not all being the same and ensuring teams

overlap. Figure B.1 displays the distribution of observed start times by watch number, which shows that each watch tends to have a 2-3 hour period in which almost all officers start their shifts, meaning there is overlap with officers working on the prior shift, producing variation in team composition. Additionally team size varies over the day. Figure B.3 displays the average team size by hour of day for teams classified into quantile based on their (within unit-year-quarter). Teams tend to be smallest in the early morning (around 5 am) and the differences between the groups is compressed. Teams tend to increase in size over the day and dispersion between groups grows, though there are small increases relative to the trends around 6 am and 2 pm, with teams tending to peak in size around 3pm and 11 pm — Figure B.4 displays the distributions of team sizes over each hour by team size group.[24]

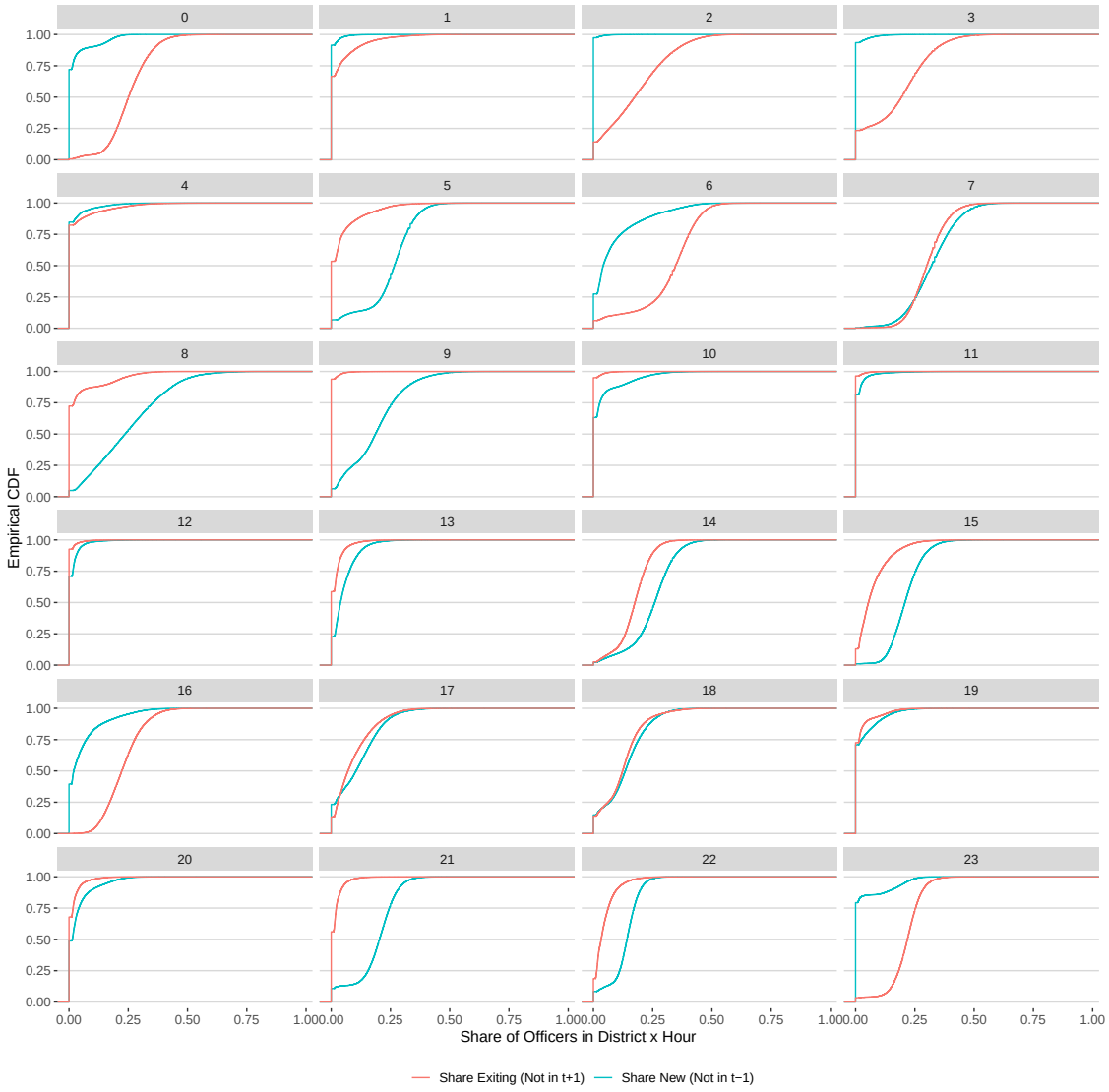
[24]Figure B.2 shows the empirical CDFs of the share of officers working in a district-hour that are new (were not working in hour $h-1$) and exiting (will not be working in hour $h+1$). In some times of day, there is not a significant change in the teams, such as 4-5am and 8 and 10 pm. However, some times tend to have many new entries, such as 7 am, 2pm, and 9 pm, while others have many exiting officers such as 1-3 am and 3pm; some times have both many new entries and exits, such as 6 am, 5pm, and 11pm.

Figure B.1: Frequency of Shift Start Hours by Watch



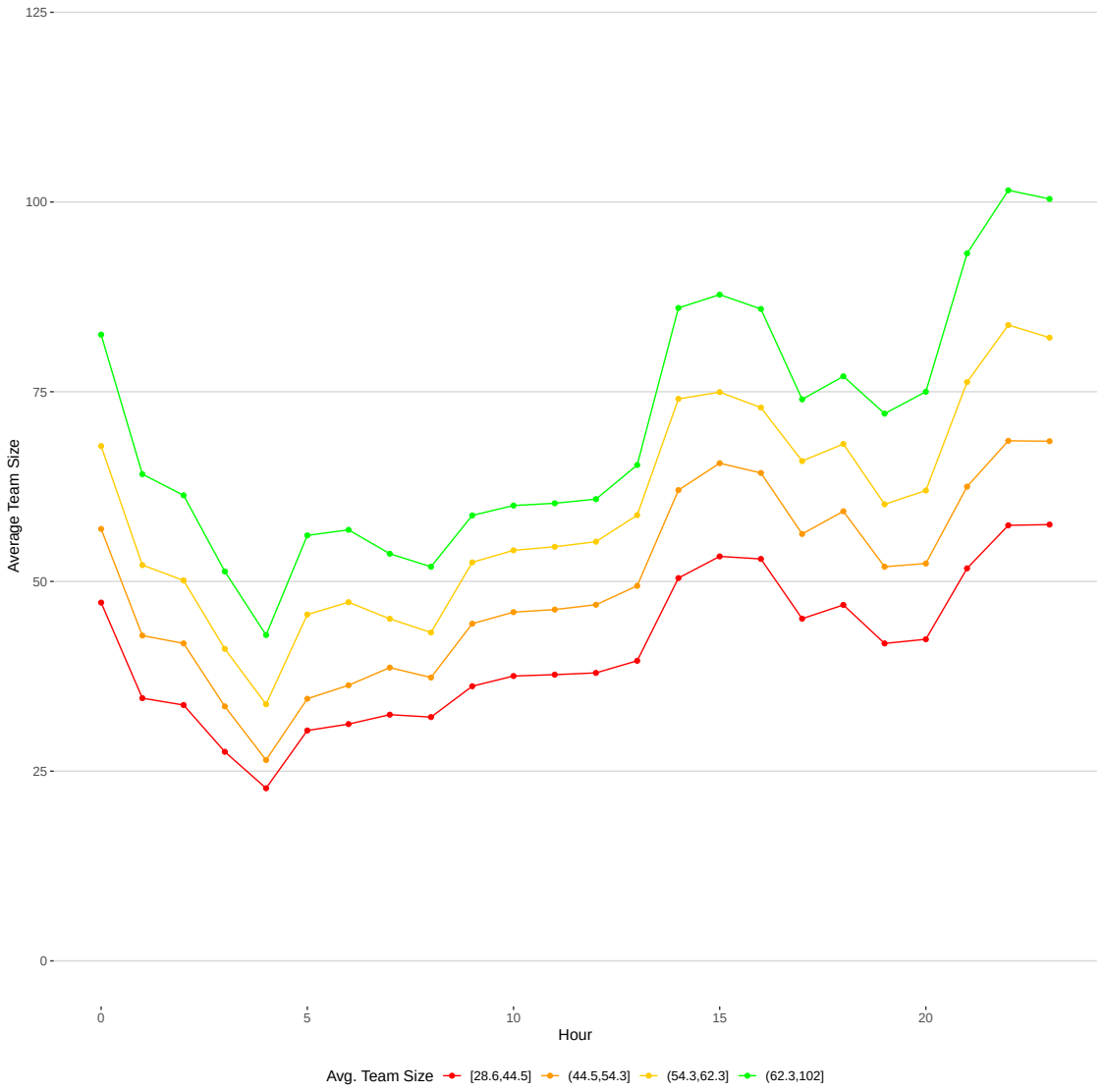
Note: Figure displays the share of observations (officer-shift) in the analysis sample by shift start hour, with colors indicating different watches, with watch 4 broken into morning and night corresponding to start times and all other watches beyond 1-4 pooled into "Other".

Figure B.2: ECDF of Share Entering and Exiting Officers by Team (District-Hour)



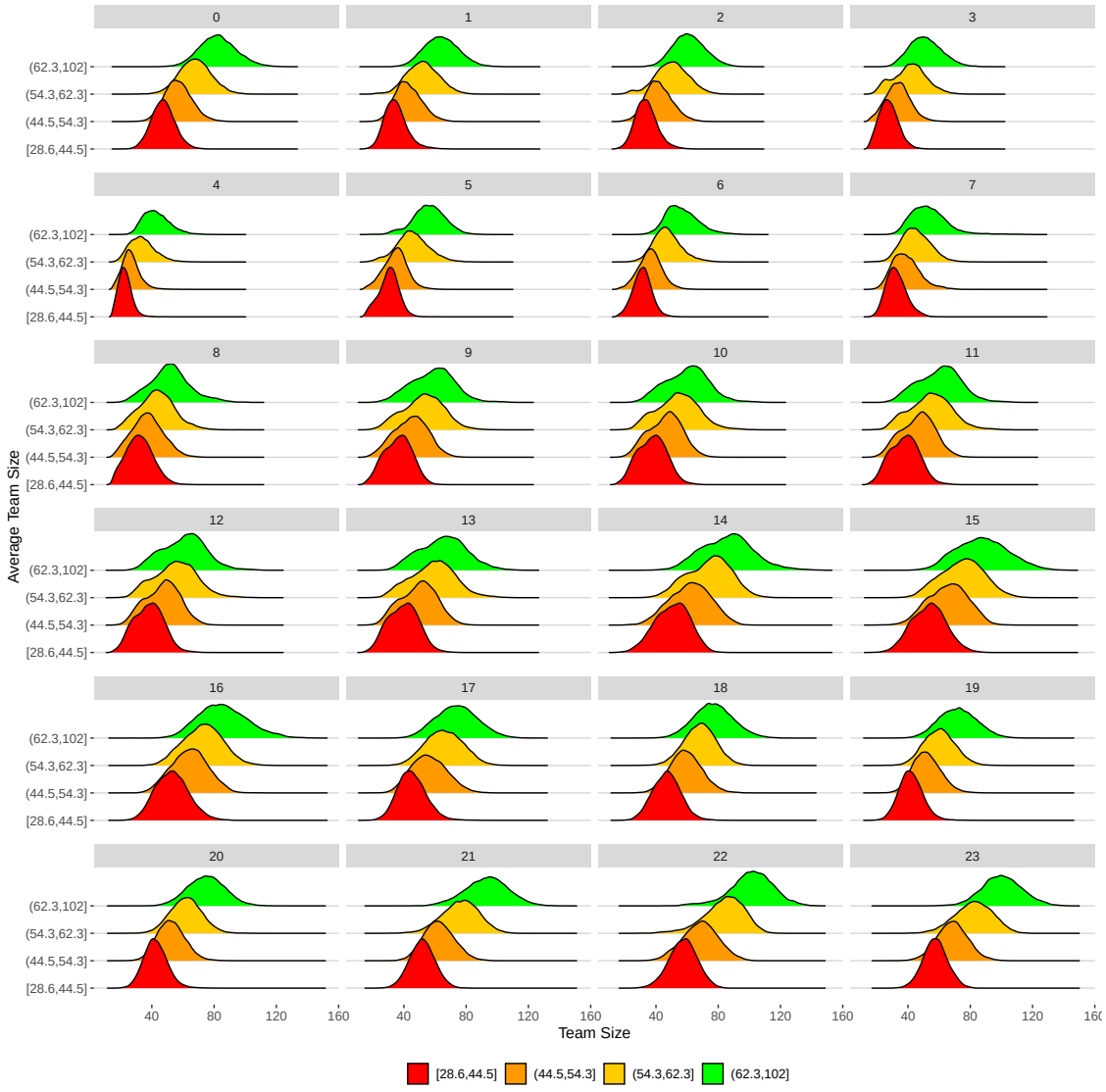
Note: Figure displays the empirical CDF of the shares of officers which are entering a team (started their shift in the hour) and share of officers which are exiting the team (ending their shift in the hour, meaning they will not be on the team in the next hour), with the denominator being the total number of officers on the team (district-hour) in the hour.

Figure B.3: Team Size by Hour in Day



Note: Figure displays the average team size (officers working in a district-hour) by hour of day with teams grouped into quantiles by their average size computed at the district-year-quarter level.

Figure B.4: Distribution of Team Sizes by Hour in Day



Note: Figure displays the distributions of team size (officers working in a district-hour) by hour of day with teams grouped into quantiles by their average size computed at the district-year-quarter level.

Similar to the officer-level model, identification of the officer effects in the team production model requires that the teams matrix (B) (along with the other fixed effects and controls) is non-singular. In the officer-assignment case the (largest) set of identified officer effects is in the largest connected component, in the team production model finding the largest identified set is more complicated. Bonhomme (2021) provides a pruning algorithm to iteratively remove indicators (officers) for whom effects cannot be identified, then removing any team

in which they are on, until convergence. To ensure non-singularity for all fixed effects, I prune the matrix containing officer indicators and district-time fixed effects to ensure both ω and γ are identified. The pruning process removes a small share of observations and officer/district-time indicators (see Table B.2), with at least 98% of observations remaining in the largest identified set, indicating that the degree of variation in team composition is sufficiently large.

Table B.2: Pruning and KSS Statistics for Crime Effects

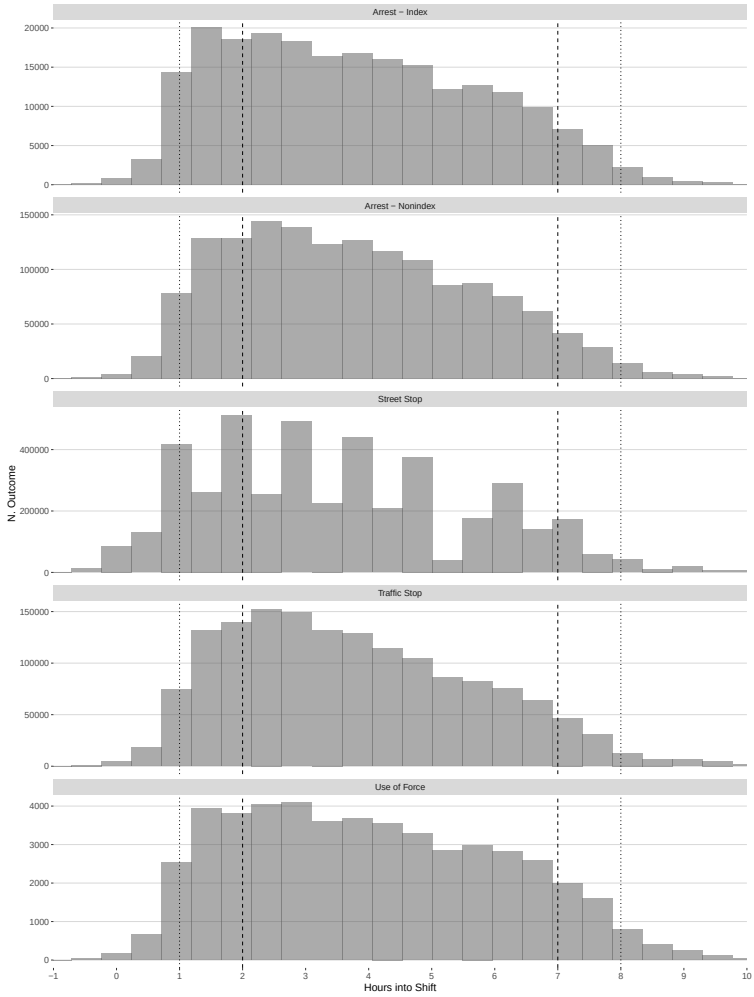
Spec Type	Date Controls	Sample Period	Initial Obs	Final Obs	Share Obs Remaining	Max PII
(1)	(2)	(3)	(4)	(5)	(6)	(7)
First Differences		2010-2012	573156	573152	1.0000	0.9141
Pred. Team First Diff		2010-2012	548530	288448	0.5259	0.9853
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2010-2012	585304	584725	0.9990	0.9900
Officer + Sgt	Year x Quarter x District x Watch x DoW	2010-2012	585304	585193	0.9998	0.2582
Pooled Officer	Year x Month x District x Watch x DoW	2010-2012	585304	585286	1.0000	0.2736
Pooled Officer	Year x Quarter x District x Watch x DoW	2010-2012	585304	585286	1.0000	0.2414
First Differences		2013-2015	516865	516832	0.9999	0.6022
Pred. Team First Diff		2013-2015	513332	236320	0.4604	0.9900
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2013-2015	539480	538719	0.9986	0.9900
Officer + Sgt	Year x Quarter x District x Watch x DoW	2013-2015	539480	539316	0.9997	0.4264
Pooled Officer	Year x Month x District x Watch x DoW	2013-2015	539480	539316	0.9997	0.4256
Pooled Officer	Year x Quarter x District x Watch x DoW	2013-2015	539480	539316	0.9997	0.4271
First Differences		2016-2019	685627	685609	1.0000	0.8918
Pred. Team First Diff		2016-2019	672284	449616	0.6688	0.9900
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2016-2019	711630	711357	0.9996	0.9900
Officer + Sgt	Year x Quarter x District x Watch x DoW	2016-2019	711630	711519	0.9998	0.6271
Pooled Officer	Year x Month x District x Watch x DoW	2016-2019	711630	711519	0.9998	0.6387
Pooled Officer	Year x Quarter x District x Watch x DoW	2016-2019	711630	711519	0.9998	0.6209

Note: Table displays the results of the Bonhomme (2021) pruning algorithm on the crime data using the fixed effect specification with different date controls (year x district x day of week-8 hour time period (watch) x month or quarter) with observations are at the district-day-hour level or the first difference specification using actual or predicted (Pred.) team membership and observations at the first difference (within districts across hours) level. Max PII refers to the maximum leverage of any observation within the sample.

The identifying variation for team effects comes from officers entering and exiting their shifts at different times compared with other members of their current team. One concern is that officers may not be deployed at the beginning and end of their shifts meaning they are not an active component of their 'team' — a related concern is that officers strategically wait

until the ends of their shifts to make more arrests to earn overtime. I do not find this to be the case, Figure B.5 displays the distribution of activity metrics (stops, arrests, and use of force) over the duration into an officers' shift. It shows that officers rapidly increase their activity within 1 hour of their shift starting with activity levels peaking around 3 hours in, then slowly declining without any clear drop off right before their shift ends. Overall, this is inconsistent with officers being inactive at the starts or ends of their shifts and with officers strategically waiting for the end of their shifts to make arrests.

Figure B.5: Distribution of Outcomes by Number of Hours into Shift



Note: Figure displays the number of each outcomes which have time-stamp information (not complaints) grouped by the hour into the officer's shift they occurred – stops time stamps are at the hour-level pre-2016 producing more bunching at hours rather than half-hours. Most shifts are 9 hours long. Vertical lines denote 1 hour and 8 hours into the shift (dotted) and 2 hours and 7 hours into the shift (dashed).

Finally, while the canonical worker-firm model focuses on logged wages, making the worker and firm effects identifying changes rather than levels, I use this model to study differences in levels, which vary widely. Part of this is due to the large number of zeros and the number of assignments and officers which never experience a specific outcome within a given time frame. I attempt to overcome this limitations in two ways. First, I estimate the model using the subset of assignments which frequently have a subset of outcomes occurring in at least 25% of shifts — this upper bound is to remove very uncommon assignments with very high means. In Table B.1, this high-outcome sample is between 35% to 65% of the data, due to changes in stopping frequency, and the LLCs are almost the entirety of the subsample always over 99%. Second, I estimate the model using a Poisson specification, which removes observations belonging to assignments which have no variation in the outcome variable.

B.2 Selecting Days-On

Given that there is sufficient variation to recover estimates for officer-level and team effects, the most direct threat to the identification of the parameters of interest in the following sections is that officers differentially select which days they work based on unobserved shocks to local crime rates or other conditions that allow for more enforcement, misconduct, or force. This would likely result in biased estimates of θ_i and ω_i because officers are choosing when to work based on the value of unobserved shocks ϵ. To overcome this, the exogenous variation for identification in the empirical sections will leverage the predetermined assignment of individual officers to work on specific days due to their DOGs being determined in the prior year through the bidding process.

The existence and importance of the operations calendar, in addition to CPD officers being governed by a strong union, makes such violations unlikely. As a first note, even after filtering uncommon beat types and allowing for occasional weekend rules and alternative rotations within some units (though still fixed based on rules), the 4-2 rotation is still, by far, the most common observed 'spell' type (meaning 4 days are observed worked in a row

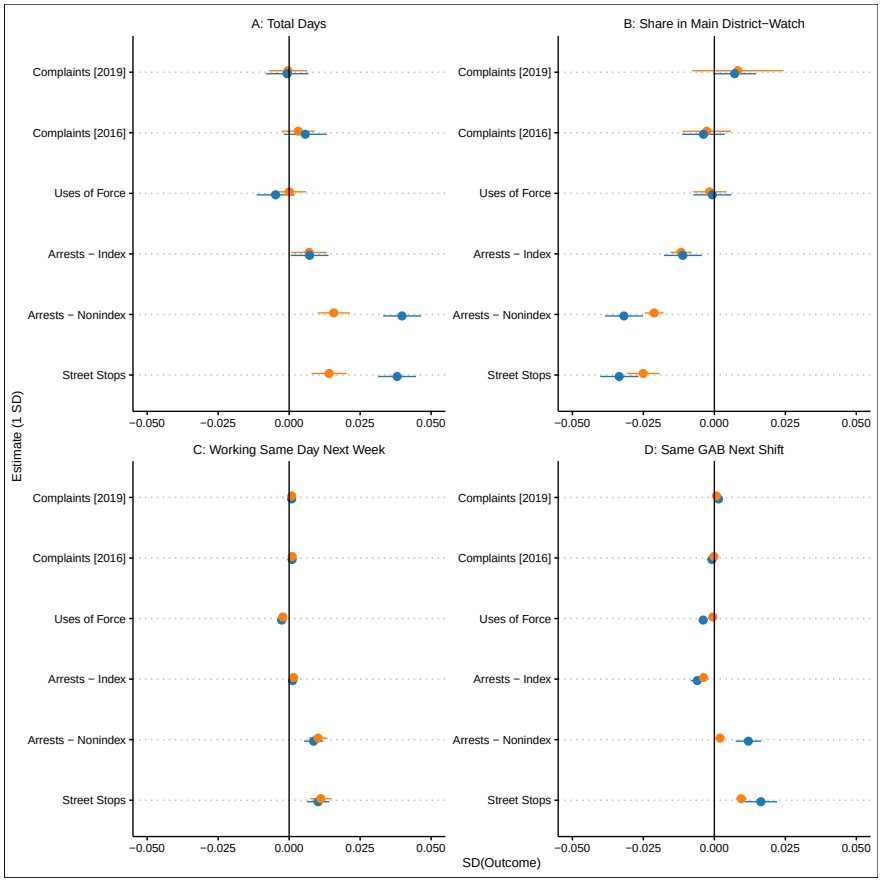
followed by 2 days not worked) at 43% of observations followed by 5-2 at 13%. All other spell types are significantly less common (the next most common is 4-3 at 5%).

Beyond this, we can directly test the strength of the calendar following Gudgeon et al. (2023). Each year, I construct a predicted rotation for each officer based on the first set of observations, which follow a 4-2 (or, in a minority of cases, 5-2) rotation. Then I project this predicted rotation backward (if the first 4-2 is observed in the middle of January, for example) and forward throughout the rest of the year. This prediction is when the officer 'should' be working and 'should' be off based on a crude usage of the rotation system, ignoring all vacation days, injuries, training, potential for switching units, or changing DOGs. Nevertheless, as shown in Figure 1, which displays the share of correctly predicted days in each week of the year, this crude prediction is quite accurate at around 85% correctly predicted early on which decreases to 70% by the end of the year. For reference, if officers were randomly working 4 out of 6 days, the share of correct predictions would be less than 50%. This consistency in predictive power of even a crude rotation is evidence of the binding nature of the calendar and the difficulty in officers deviating, even if they could observe shocks to local crime conditions on a day-to-day basis.

To test for systematic violations of the calendar, we perform the following tests. First, if officers cannot easily manipulate the calendar system, their outcomes should not relate to calendar-related statistics, such as the number of days they work in the year, the share of days they work in their main assignment, or their likelihood of working on the same date in the following week. As a note, around 90% of observations involve officers working in their most common (permanent) unit-watch in any given year, with the other 10% including in-year unit transfers and any deviations due to departmental demands. If, for example, high arrest officers preferred working more then we'd expect average arrest rates to relate to the number of days they work or the amount of time they spend in other units. Similarly, if officers can manipulate their schedules then it is possible that outcomes on a specific date would be predictive of deviations in the future or the probability of working the same date

next week, both of which should be unaffected if the calendar binds. As shown in Panels A-C in Figure B.6, no outcome has an economically significant relationship with the share of days an officer spends in their main unit-watch, the total number of days they work, or their likelihood of working the same day next week with or without assignment (unit-year-watch) fixed effects — meaning a 1 SD increase in any activity has no more than a 0.05 SD increase or decrease in the outcome. Other work on the CPD have done alternative tests to verify the calendar system is obeyed in practice with positive results (e.g., Ba, Knox, et al. (2021), Ba, Bayer, et al. (2021), Gudgeon et al. (2023))— notably Gudgeon et al. (2023) finds that officers (the focus of this paper) do obey the calendar system and even lieutenants, to a lesser extent, who have much more discretion are still strongly influenced by it.

Figure B.6: Deviations from Permanent Unit, Calendar, and Assignment



Note: Figure displays the deviations from the operations calendar and steady assignment (GAB) following the occurrence of a specific officer-level outcome (estimates scaled by 1 SD), and outcomes in the regressions are standardized. Points denote estimates with and without year-unit-watch fixed effects.

B.3 Endogenous Mobility

The next threat to identification is that errors are not conditionally independent or strictly exogenous with respect to the observables due to "endogenous mobility" (Card et al. (2013)). In this environment, officers are permitted to deliberately sort into specific assignments for any static reason. For example, they move from one assignment to another because of better working hours, lower crime rates, they will have more opportunities to commit misconduct, or because the commute from their home is shorter. These factors are fixed for the assignment, and thus are captured in ψ_J, and any match-specific component (e.g., commuting distance) can be assumed to be mean zero, and thus unobserved shocks to outcomes are not an issue.

However, Card et al. (2013) outlines three sources of endogenous mobility which can also arise in the policing environment, and 'event study' figures are traditionally used to assess the prevalence of such issues. Figure 2 displays such event studies for all main outcomes, with officers working in assignments classified by the frequency of the respective outcome within that year, then the assignments are binned into 4 groups based on their quantile of their average outcome (Q1-4, where 4 is the highest). Figure 2 displays the event study plots for each core outcome, showing the average outcome by switch type (e.g., Q4 to Q1, Q1 to Q3, etc.) for the six shifts before and after an officer changes their assignment (note that many shifts occur after less than 6 days, but these contribute to the means as well).

First, officers may be sorting into assignments due to match-specific effects, meaning that $\epsilon^k_{i,t}$ contains a component an officer-assignment component which officers are using to determine where to work. If this match component is not mean zero, this is a violation of strict exogeneity; then when officers move between assignments we would expect that the officers who move from assignments with lower outcomes to ones with higher outcomes would experience out-sized gains in their individual frequency of that outcome compared with officers moving from higher to lower. For example, if match effects exist for arrests, then officers who are moving from relatively low arrest assignments to higher arrest ones should see a much larger gains in their individual arrests compared with officers moving from

high arrest assignments to low arrest ones. Figure 2 shows that this is not the case. Officers who move from Q1 assignments to higher quantiles do not experience out-sized increases in outcomes compared to the declines of those moving from Q4 assignments to lower ones.

Second, there may be a drift component in $\epsilon_{i,t}^{k}$ if officers learn during their assignments and this is predictive of future movements. While we expect officers to learn as they work, this would be an issue if officers who are experiencing upward (or downward) trajectories in their outcomes systematically switch to higher (or lower) outcome assignments because this drift component would then produce upward bias in the variation attributed to assignment effects because higher outcome assignments would receive officers who are on (unobservably) upward trends and the opposite for lower outcome assignments. As in Figure 2 there tend to be downward trends for Q4 officers prior to switches, however this is true for switches to all quantiles from Q4, meaning that the downward trend is not indicative of a downward move. Furthermore, the decline for Q4 and increase for Q1 (when present) is driven by the day right before the move and there is a symmetric pattern after they move.

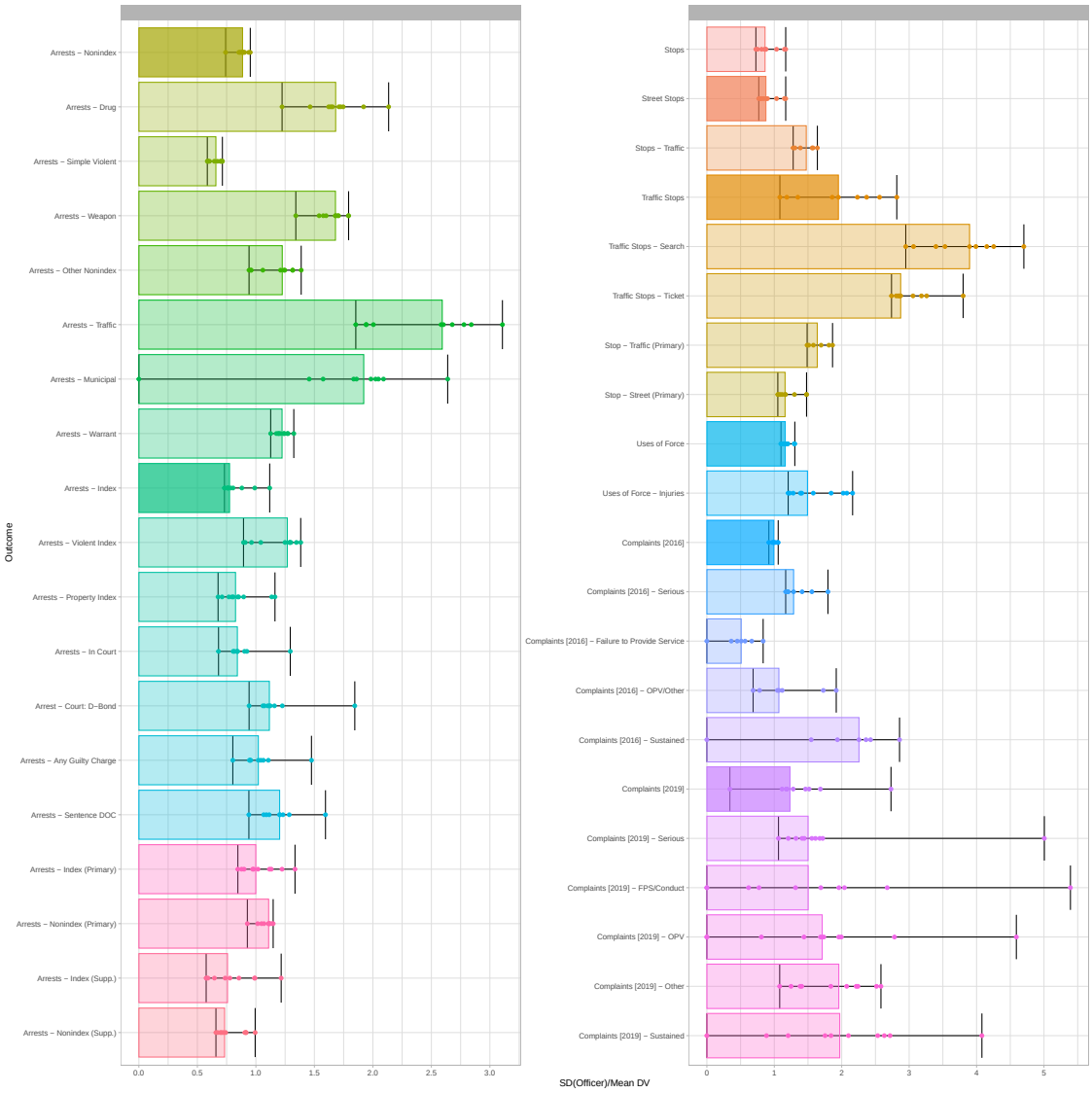
Third, if officers switch assignments based on transitory shocks to their outcomes — for example, if an officer who receives a complaint due to a random shock is then moved to a higher complaint assignment the next day this would produce biased estimates. However, shown in the event study plots, there is no sudden jump or drop in average outcomes among switchers, and this is bolstered by the results in Panel D in Figure B.6, which displays the relationship between officer outcomes and the likelihood of working in the same GAB in the following shift. As shown, a specific outcome on a given day has virtually zero relationship both with and without unit-year-watch fixed effects.

While these patterns are generally consistent with the exogenous mobility assumptions, I probe the robustness of my main results using alternative samples which should alleviate any remaining concerns. I use alternative samples for the main analysis focusing on long spells within GABs (at least 4 shifts), long spells within unit-watches (at least 20 shifts), a 'donut' approach where I remove the first and last observation within an officer's spell for

their consecutive shifts in a GAB, and I aggregate data at the week level and the officer's GAB within a week is their modal GAB. These alternative specifications will avoid over-weighting short and sporadic spells which may be more susceptible to endogenous mobility concerns and counting the dips/rises in outcomes immediately before and after a switch if present. Additionally, if there are concerns over over- or under- specificity in definition of an assignment, I redefine it as either Unit-Watch or Unit-Watch-Beat as well. Finally, Appendix D.3 discusses how much variation is explained when using a full match-effects specification.

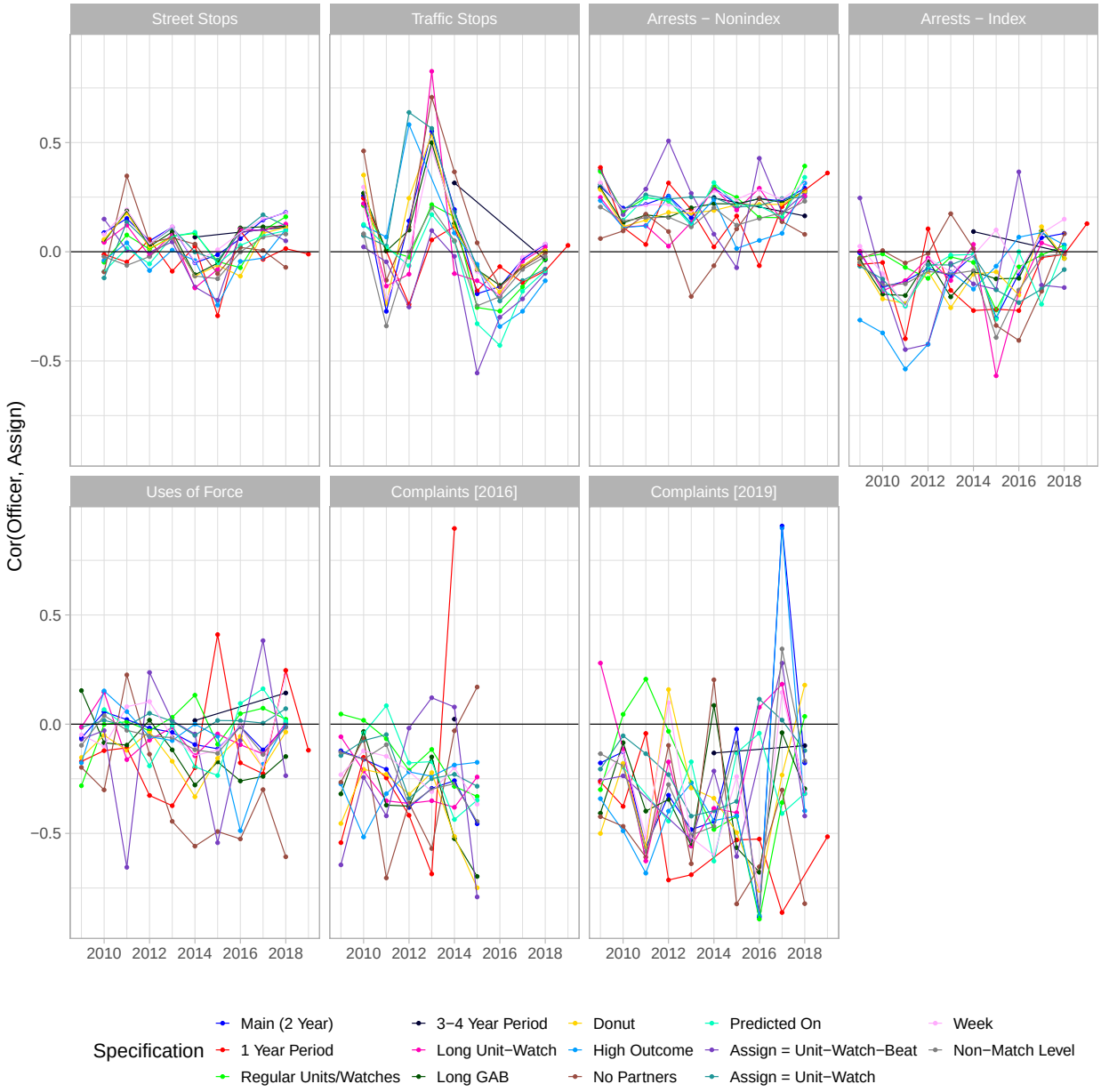
C Appendix C - Additional Figures and Tables for Results

Figure C.1: SD(Officer) for Additional Outcomes



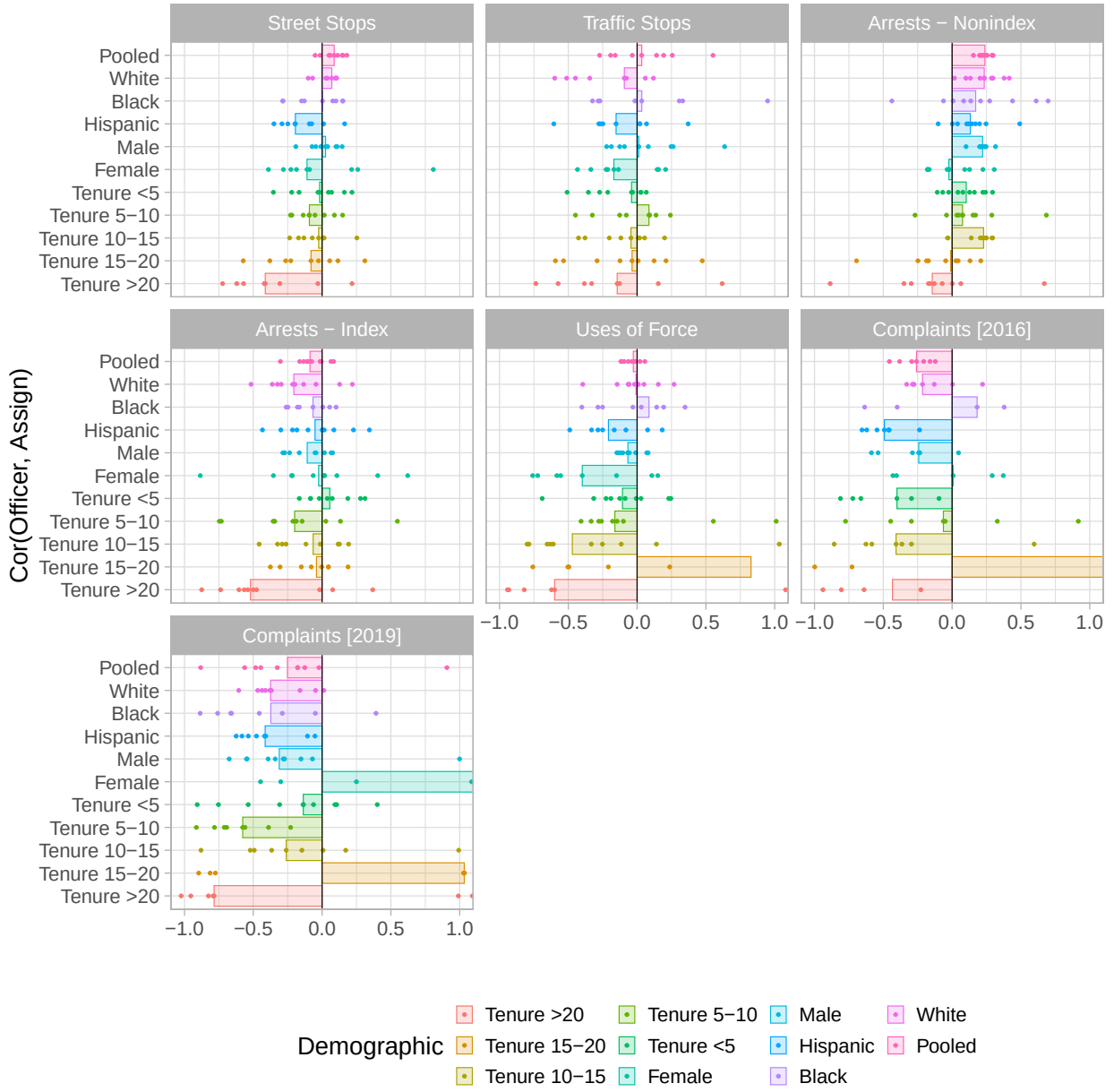
Note: Figure displays the KSS bias-corrected standard deviations of officer effects (recovered from equation (1)) for additional outcomes scaled by the mean of the outcome, with each point denoting an estimate for a specific time period, the bar denoting the median across these estimates, and the confidence interval denoting the minimum and maximum estimates. Main (2 Year) refers to the main specification with 10 2-year rolling periods. Main outcomes are shaded while non-main outcomes are more transparent.

Figure C.2: Degree of Officer-Assignment Sorting over time by Specification



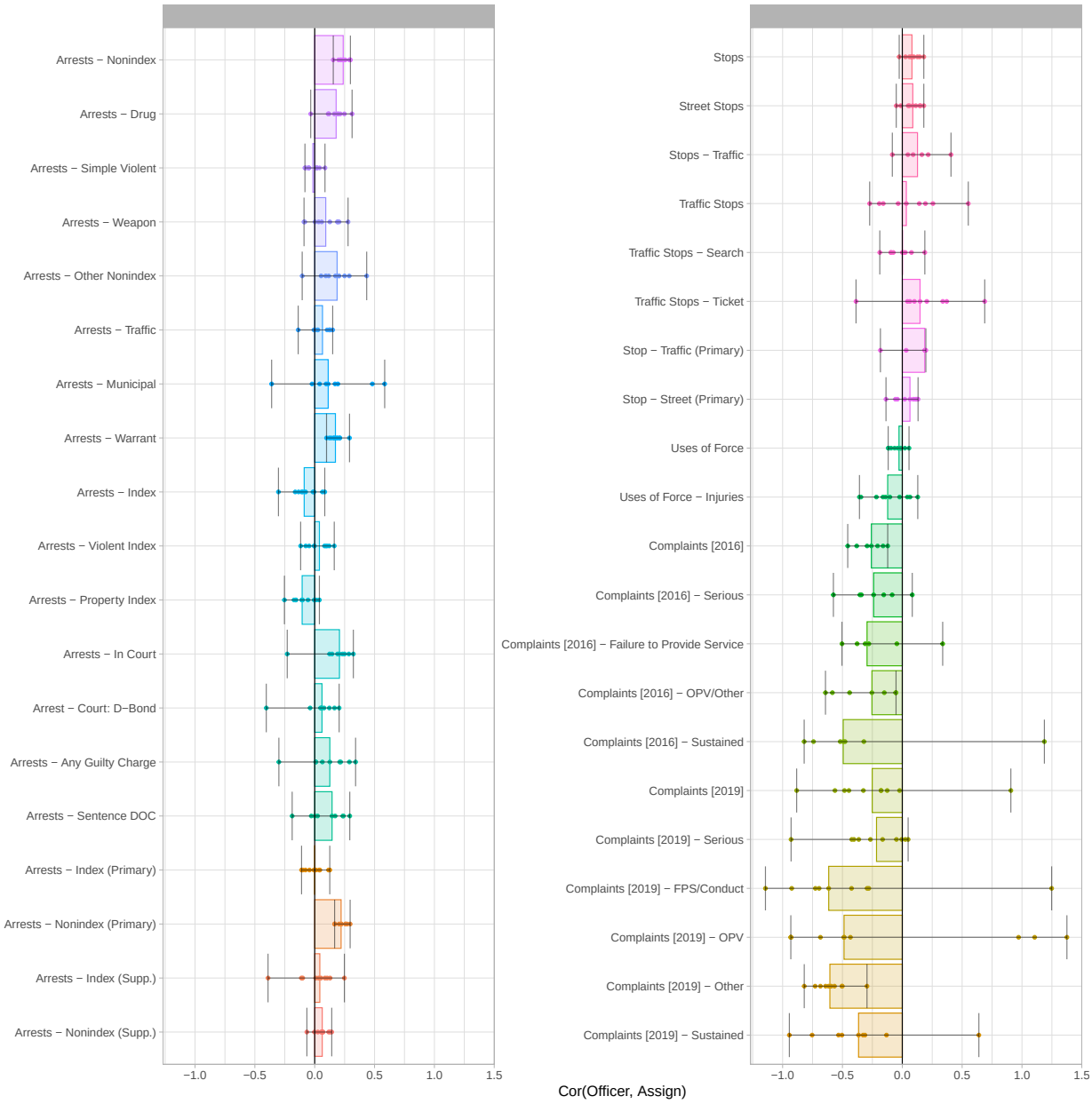
Note: Figure displays the KSS bias corrected correlations between officer and assignment effects, recovered from estimating the different specification and equation (1), with the x-axis denoting the time period of the analysis sample used.Main (2 Year) refers to the main specification with 10 2-year rolling periods; 1 Year Period refers to each year of data analyzed separately; Geographic refers to the main specification but subsetting to geographic units only; 3-4 Year period refers to estimating the model using 3 distinct periods which cover 2009-2012, 2013-2015, 2016-2019; Long Unit-Watch (GAB) refers to selecting for spells in which the officer spends at least 20 (4) consecutive days in the unit-watch (GAB); Donut refers to removing the first and last day of each officer-GAB spell; High Outcome selects for GABs which have average outcomes (all main outcomes summed) of at least 0.25 within a year; No Partners removes observations in which the officer has another officer in their car; Predicted On includes observations where the officer was predicted to be working on that exact day, watch, and unit; Assign = Unit-Watch-Beat (Unit-Watch) replaces the assignment definition of GAB in the main specification, Unit-Watch-Beat Category, with Unit-Watch-Exact Beat (Unit-Watch); Week refers to aggregating the data to the weekly level; Non-Match Level computes leverages at the observation, rather than officer-assignment match, level.

Figure C.3: Degree of Officer-Assignment Sorting by Demographic Group



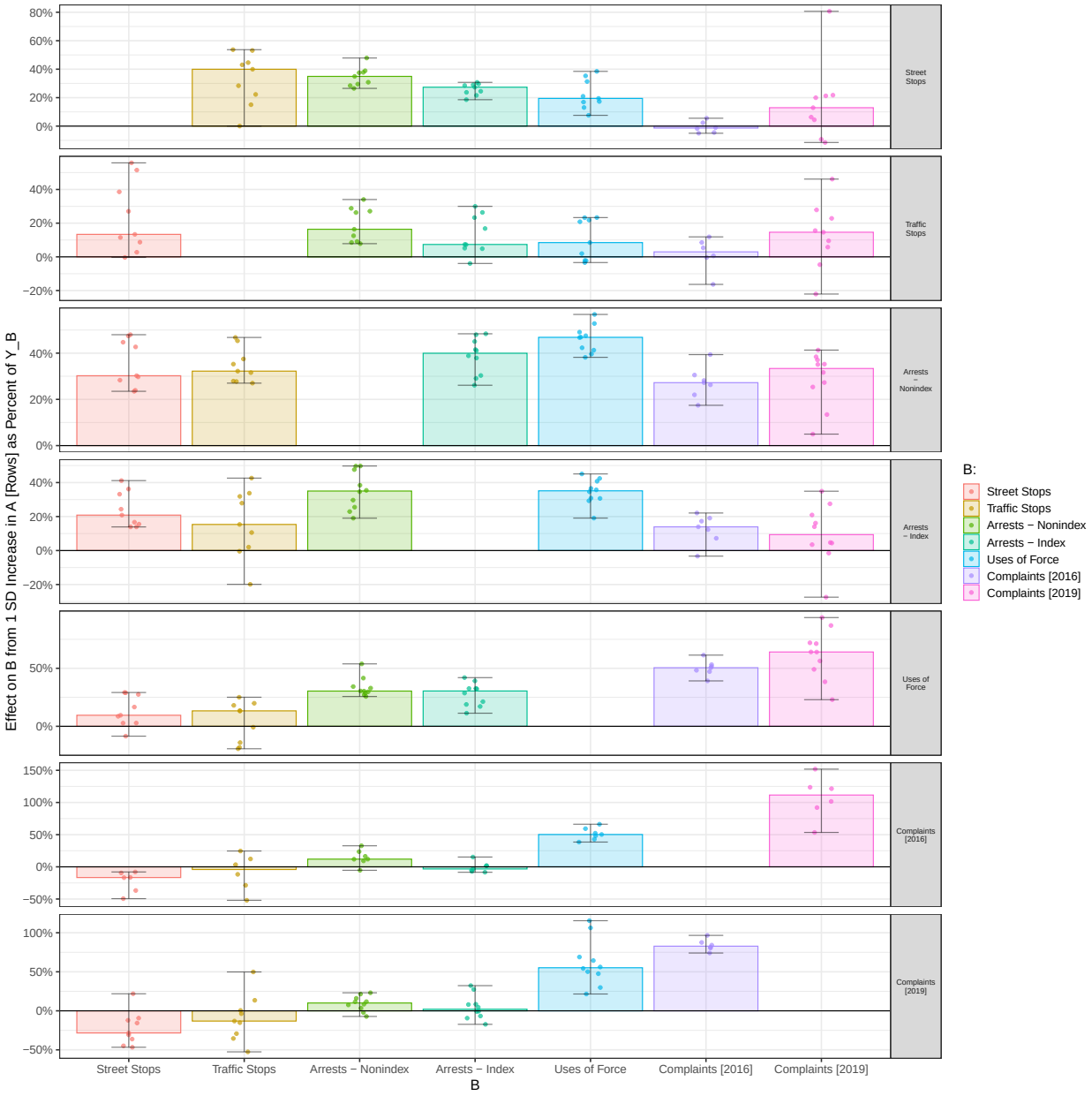
Note: Figure displays the KSS bias corrected correlations between officer and assignment effects, recovered from estimating equation (1) using subsamples of different demographic groups (y-axis), across 10 2-year periods with each point denoting an estimate for a specific time period and the bar denoting the median across these estimates.

Figure C.4: Degree of Officer-Assignment Sorting for Additional Outcomes



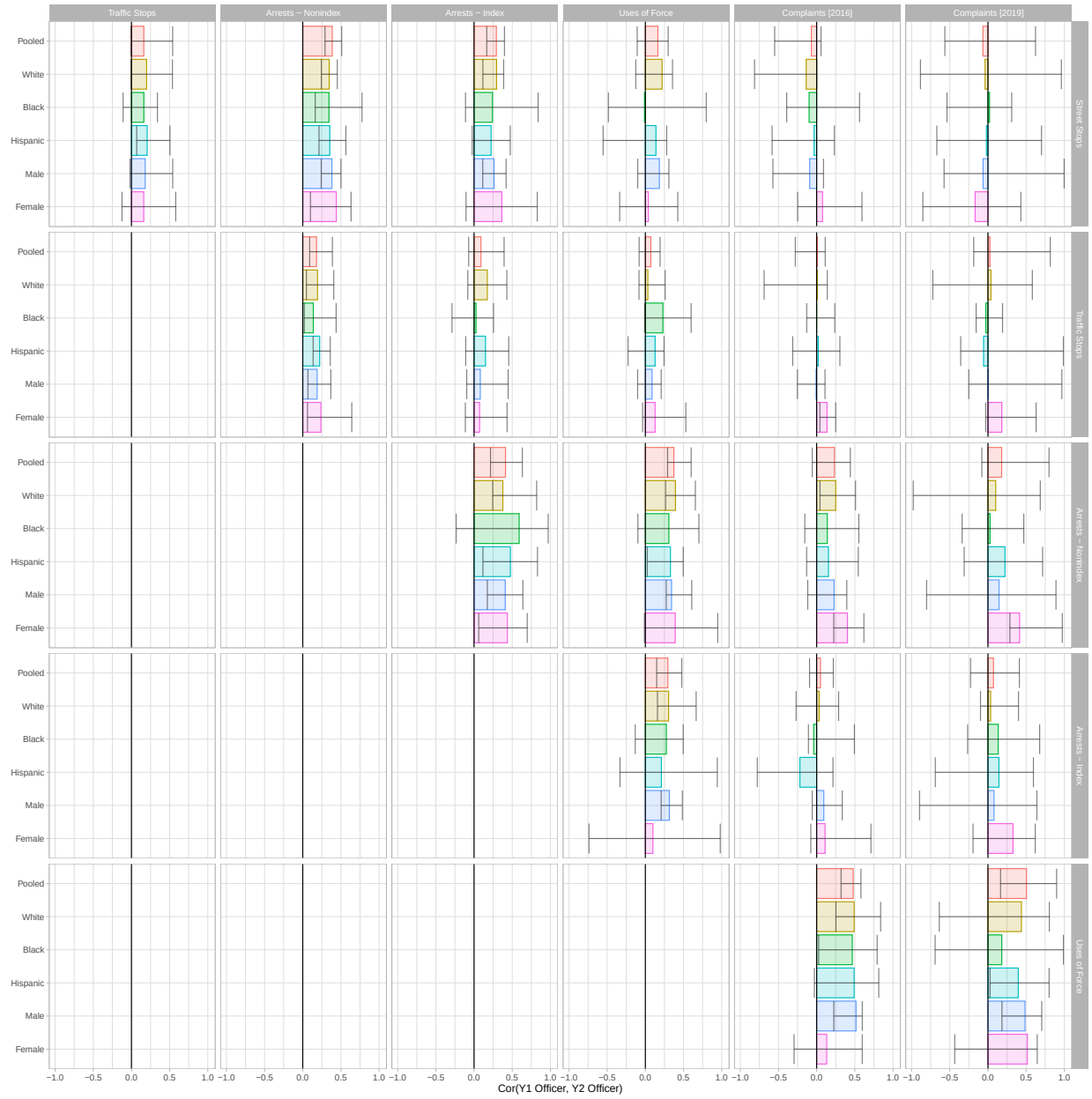
Note: Figure displays the KSS bias corrected correlations between officer and assignment effects, recovered from estimating equation (1) for additional outcomes, across 10 2-year periods with each point denoting an estimate for a specific time period and the bar denoting the median across these estimates.

Figure C.5: Effect of 1 SD Change Across Officer Activities



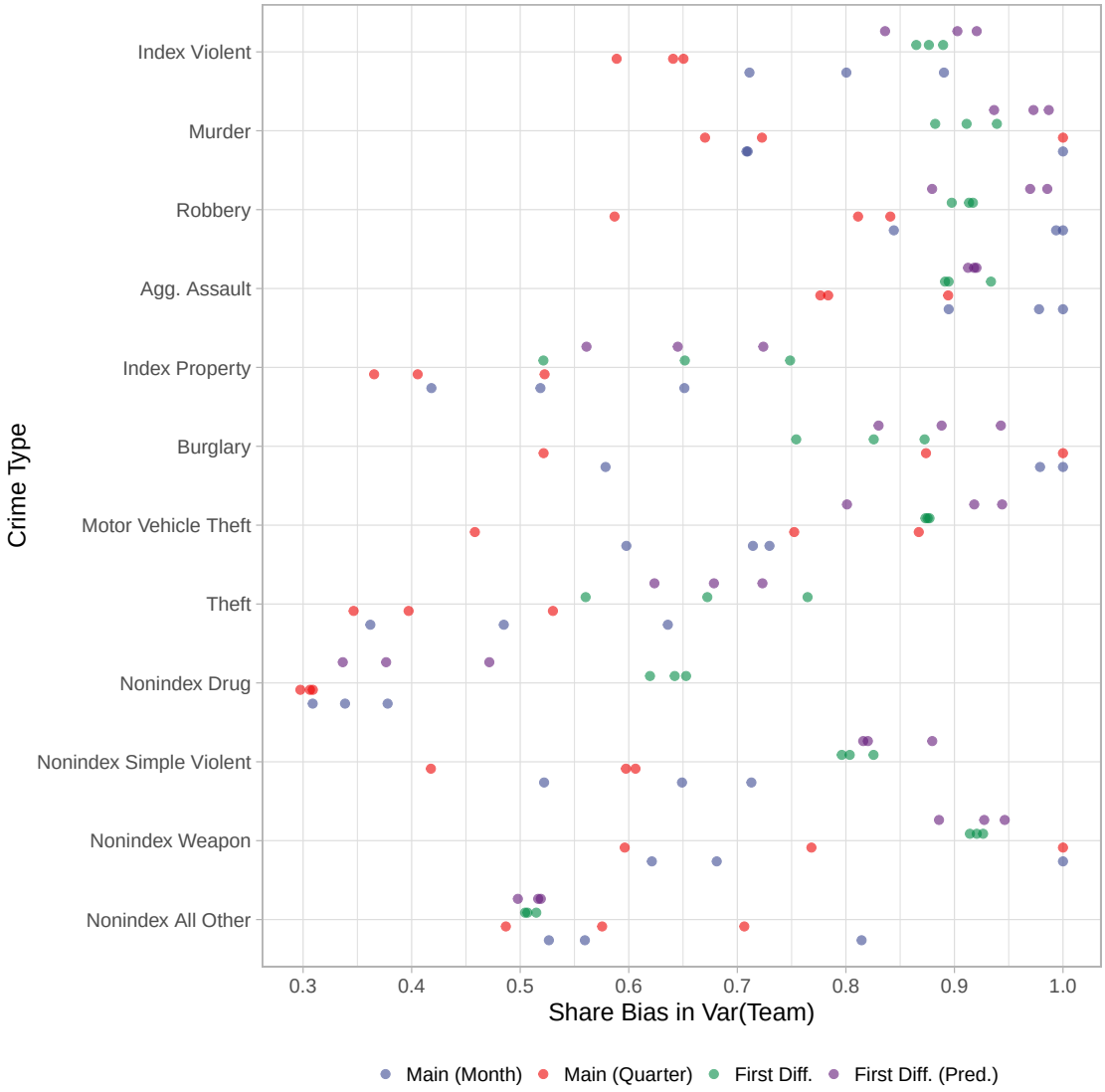
Note: Figure displays the effect of a 1 SD increase in one dimension of officer effects [A, rows] on another dimension of officer effects [B, columns] as a percent of the mean of the raw outcome B. Officer effects are recovered from estimating equation (1), across 10 2-year periods with each point denoting an estimate for a specific time period and the bar denoting the median across these estimates, and the error bars denoting the minimum and maximum across samples.

Figure C.6: Cross-Outcome Correlation in Officer Effects by Demographic Group



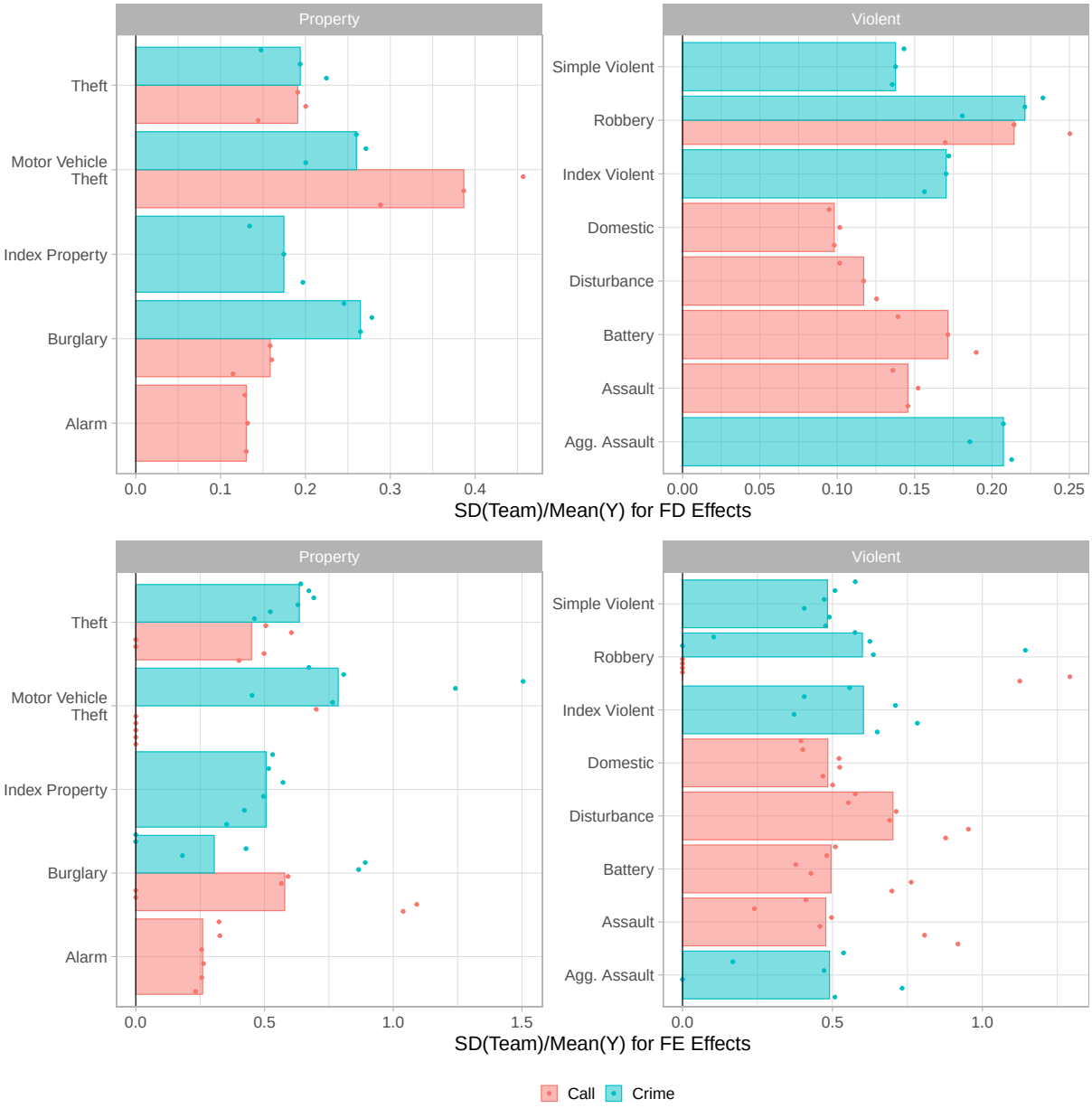
Note: Figure displays the KSS bias corrected correlations between officer effects for two different outcomes for different demographic groups, with officer effects recovered from estimating equation (1) using subsamples of different demographic groups, across 10 2-year periods with each point denoting a correlation for a specific time period and the bar denoting the median across these correlations.

Figure C.7: Share of Plug-In Var(Team) Attributable to Bias



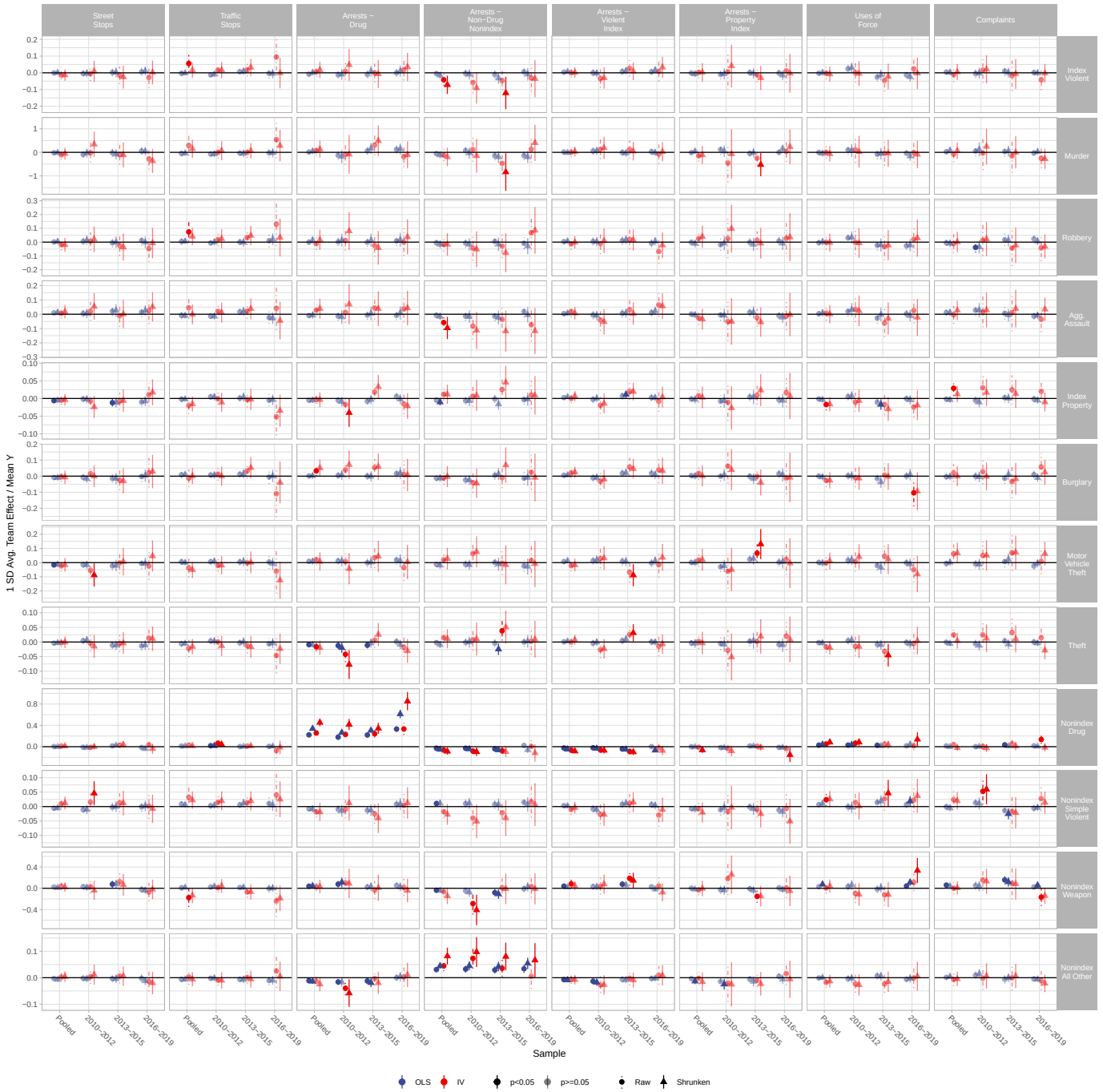
Note: Figure displays the share of plug-in (non-bias corrected) $Var(\Omega)$ across crime types and specifications. Main (Month/Quarter) refers to the fixed effect specification using district-hour level with controls for year-district-day of week-8 hour period interacted with either month or quarter, and First Diff. refers to the first difference specification with either actual or predicted (Pred.) team membership.

Figure C.8: Influence of Teams on 911 Calls



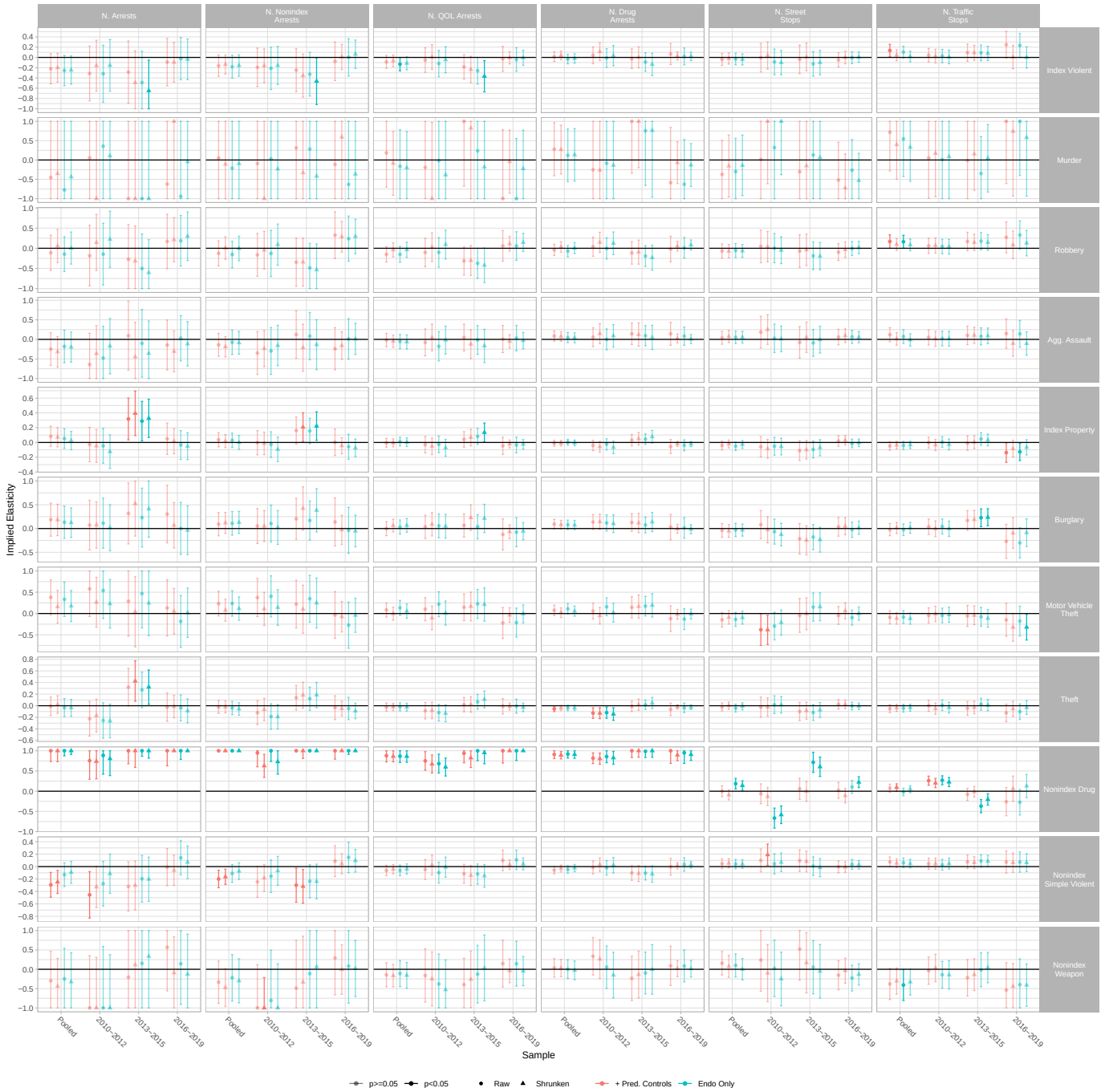
Note: Figure displays the KSS bias corrected standard deviations (left panel) and variance (right panel) of team effects (Ω) for 911 call outcomes (y-axis), scaled by the mean of the outcome and by the overall variance of the outcome, respectively. Team effects were estimated using the fixed effect, equation (3), and first difference specification, equation (6), and each point corresponds to one of three multi-year samples under different fixed effects or using actual or predicted team composition, while the bars denote the median of these estimates, and the colors correspond to whether or not the calls included in the outcome included reporting calls. Zero or negative values reflects plug-in estimates being entirely attributable to bias.

Figure C.9: Effect of Team Average Predicted Officer Effects on Crime



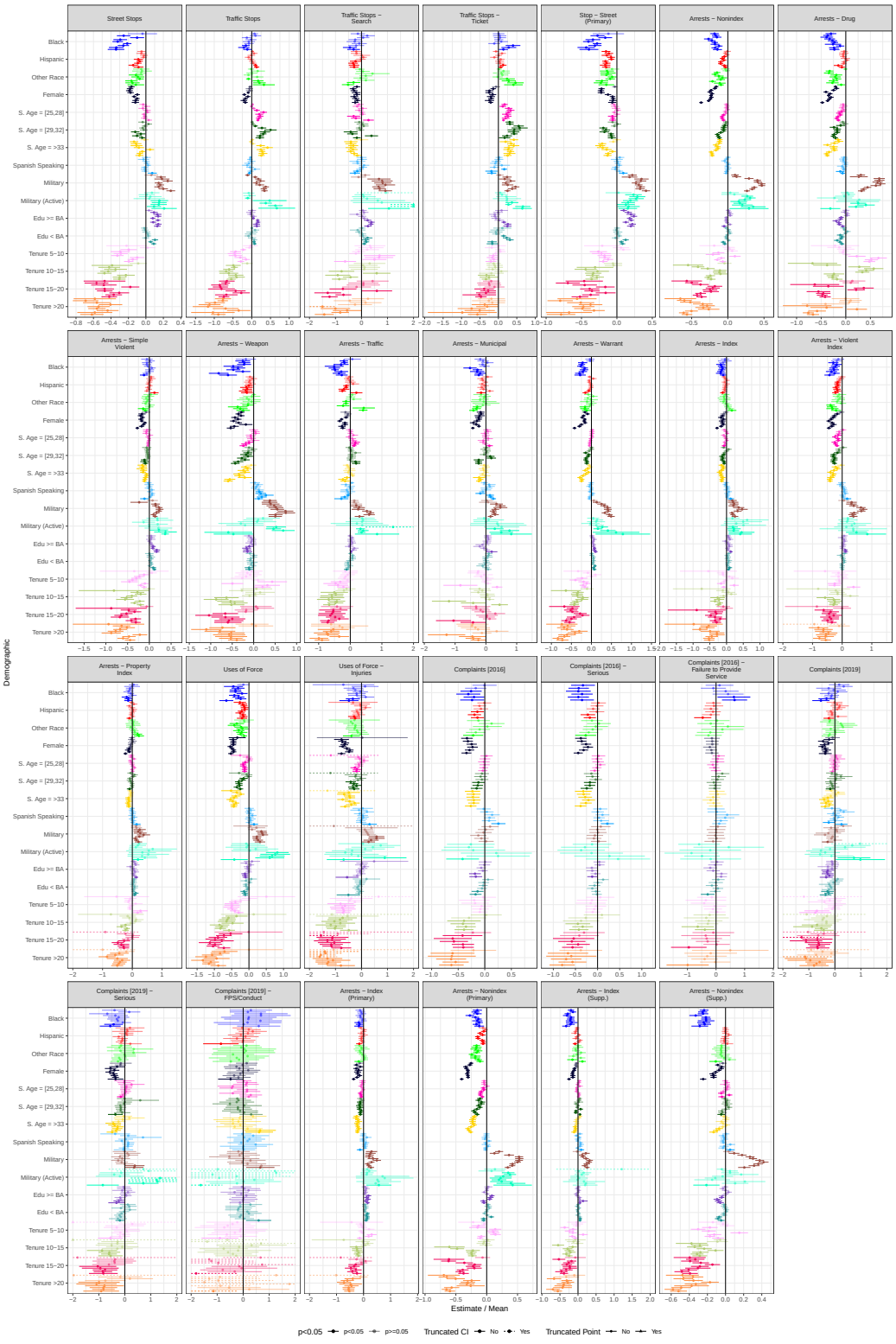
Note: Figure displays coefficients and associated 95% confidence intervals for the effect of 1 standard deviation increase in team average officer propensity for each officer-level behavior (columns) on different crime types (rows), from estimating equation (10) for OLS / reduced form and equation (11) is the first stage, over the pooled sample (2010-2019) or one of the three sample periods. All regressions use year-month-district-day of week-1 hour fixed effects and additional controls. Officer effects are either not shrunken (raw) or shrunken. Standard errors are clustered at the year-month-district-day of week-1 hour level.

Figure C.10: Elasticities of crime with respect to policing activities



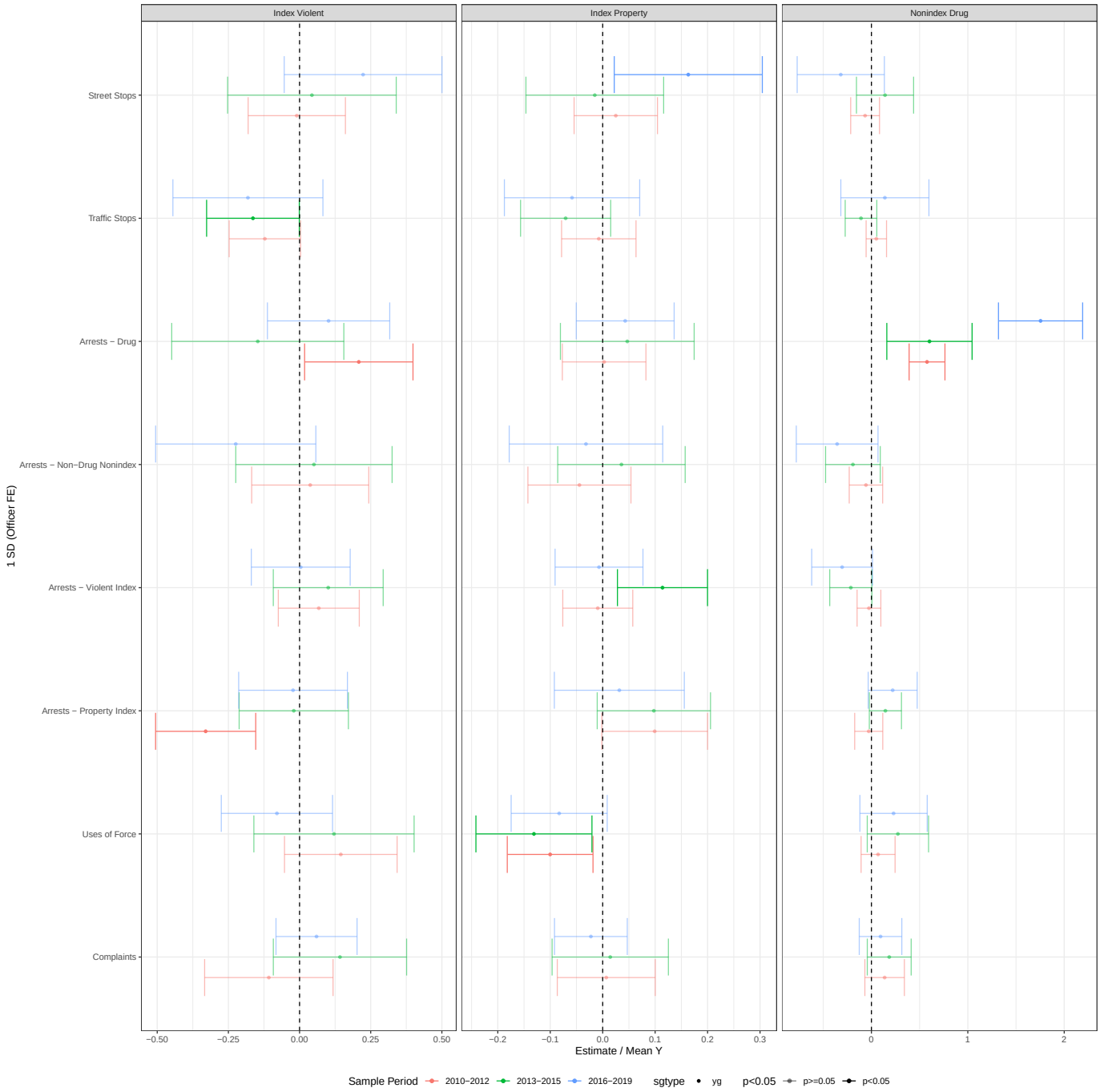
Note: Figure displays coefficients and associated 95% confidence intervals for elasticities of crimes (rows) with respect to policing activities (columns) recovered by estimating equation (12) and instrumenting for the number of activities in a district-hour (e.g., street stops) using predicted team average propensities for that activity as in equation (13), then dividing the coefficient by the mean of the crime type and multiplying it by the mean of the activity, across sample periods or a pooled sample. Officer propensities are either raw (not shrunken) or shrunken, and the specification is either "Endo. Only" meaning it only includes instruments for the endogenous activity or "+ Pred. Controls" meaning the first and second stage control for other predicted team average propensities. Standard errors used to compute confidence intervals are clustered at the year-month-day of week-district-hour level. Figures are truncated to ensure readability.

Figure C.11: Projecting Officer Activity Effects onto Officer Characteristics



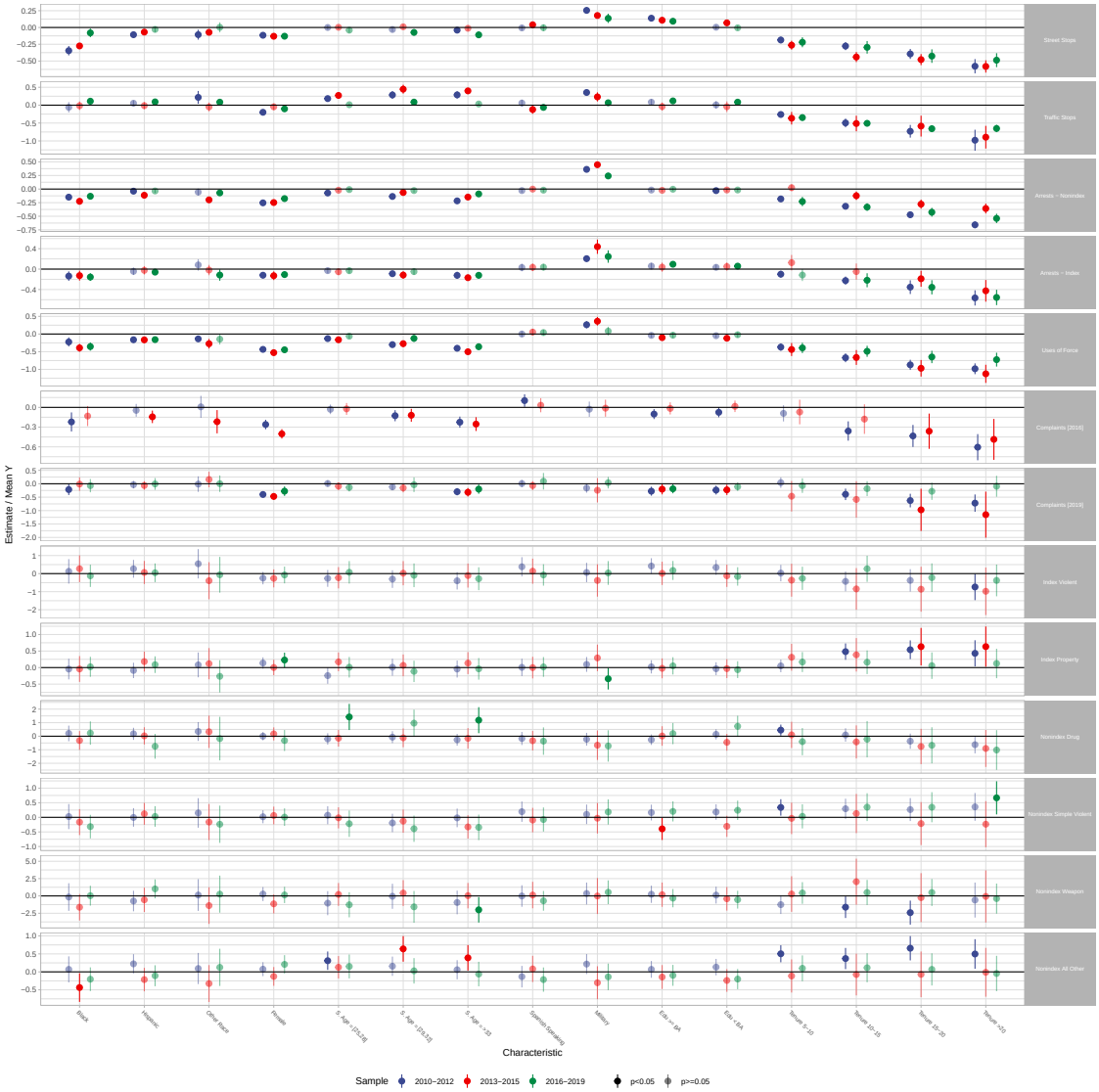
Note: Figure displays the results of regressing officer effects for each outcome (equation (1)) on officer characteristics (equation (9)), with coefficients scaled by the mean of the outcome variable. Each point denotes a point estimate for a single 2-year sample, and intervals denote 95% confidence intervals computed from KSS adjusted heteroskedasticity robust standard errors. Figures are truncated at -2 to 2 (200% decrease or increase compared with the mean), with triangle points denoting a truncated point estimate and dashed intervals denoting a truncated confidence interval. Confidence intervals and points are bolded if they are statistically significant at $p < 0.05$.

Figure C.12: Projecting Officer Crime Effects onto Officer Characteristics



Note: Figure displays coefficients and associated 95% confidence intervals (using KSS adjusted heteroskedasticity robust standard errors) from projecting officer crime effects onto shrunken officer activity fixed effects across three sample periods for three crime types. Coefficients and confidence intervals are scaled to be the effect of a 1 SD increase in the shrunken officer activity effect relative to the mean of the raw outcome. Officer crime effects are initially recovered by estimating equation (3) over three sample periods using year-month-district-day of week-8 hour period fixed effects and additional controls.

Figure C.13: Projecting Officer Activity and Crime Effects onto Officer Characteristics



Note: Figure displays the results of estimating equations (14) and (15) by projecting officer effects on activities and crime (rows) onto officer observables (columns) over the three sample periods, with officer effects recovered from estimating equation (1) and equation (3). Standard errors are heteroskedastic robust adjusted using KSS and are used to compute p-values.

Table C.1: Summary of Team Compositions and Related Variables by District-Hour

	Pooled	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)
A: Demographics and Experiences				
Share Black	0.228 (0.219)	0.251 (0.241)	0.232 (0.22)	0.206 (0.196)
Share Hispanic	0.255 (0.138)	0.231 (0.14)	0.24 (0.133)	0.286 (0.134)
Share Other Race	0.042 (0.044)	0.039 (0.044)	0.041 (0.042)	0.046 (0.044)
Share Female	0.207 (0.082)	0.221 (0.087)	0.205 (0.079)	0.197 (0.079)
Share S. Age = [25,28]	0.359 (0.08)	0.343 (0.079)	0.365 (0.074)	0.367 (0.082)
Share S. Age = [29,32]	0.233 (0.067)	0.226 (0.068)	0.231 (0.063)	0.241 (0.068)
Share S. Age = >33	0.233 (0.077)	0.245 (0.078)	0.23 (0.074)	0.226 (0.078)
Share Military	0.223 (0.206)	0.1 (0.089)	0.185 (0.165)	0.353 (0.229)
Share Spanish Speaking	0.15 (0.099)	0.154 (0.113)	0.148 (0.096)	0.15 (0.087)
Share Edu < BA	0.377 (0.103)	0.345 (0.108)	0.375 (0.1)	0.406 (0.092)
Share Edu >= BA	0.302 (0.094)	0.27 (0.095)	0.299 (0.089)	0.331 (0.088)
Share Tenure 5-10	0.196 (0.147)	0.273 (0.151)	0.232 (0.144)	0.105 (0.085)
Share Tenure 10-15	0.232 (0.116)	0.269 (0.121)	0.256 (0.113)	0.184 (0.094)
Share Tenure 15-20	0.224 (0.132)	0.201 (0.132)	0.246 (0.136)	0.226 (0.126)
Share Tenure >20	0.163 (0.15)	0.105 (0.11)	0.16 (0.145)	0.214 (0.164)
B: Raw Officer Activity FEs				
Street Stops	0 (0.011)	0 (0.011)	0 (0.016)	0 (0.0064)
Traffic Stops	0 (0.01)	0 (0.0045)	0 (0.003)	0 (0.015)
Arrests - Any	0 (0.0043)	0 (0.0053)	0 (0.0043)	0 (0.0032)
Arrests - Drug	0 (0.0014)	0 (0.0017)	0 (0.0016)	0 (0.00075)
Arrests - Non-Drug Nonindex	0 (0.0025)	0 (0.0028)	0 (0.0024)	0 (0.0021)
Arrests - Quality of Life	0 (0.0022)	0 (0.0028)	0 (0.0024)	0 (0.0014)
Arrests - Violent Index	0 (0.00019)	0 (0.00019)	0 (0.00022)	0 (0.00015)
Arrests - Property Index	0 (0.0005)	0 (0.0006)	0 (0.00052)	0 (0.00036)
Uses of Force	0 (0.00017)	0 (0.0002)	0 (0.00017)	0 (0.00012)
Complaints	0 (0.0001)	0 (0.00013)	0 (0.00013)	0 (0.000043)
Complaints - Failure to Provide Service	0 (0.000038)	0 (0.00005)	0 (0.000039)	0 (0.000024)
C: Shrunken Officer Activity FEs				
Share Missing Officer FEs	0.04 (0.031)	0.033 (0.028)	0.037 (0.029)	0.048 (0.034)
Street Stops	0.0012 (0.016)	0.0025 (0.016)	0.0016 (0.022)	-0.00017 (0.0082)
Traffic Stops	-0.0008 (0.01)	0.00076 (0.0064)	-0.000078 (0.0045)	-0.0026 (0.015)
Arrests - Any	0.00088 (0.0068)	0.0018 (0.0083)	0.00084 (0.0068)	0.00016 (0.0052)
Arrests - Drug	-0.000038 (0.0022)	0.00013 (0.0028)	-0.00014 (0.0024)	-0.000094 (0.0015)
Arrests - Non-Drug Nonindex	0.00043 (0.0038)	0.00099 (0.0043)	0.00042 (0.0037)	-0.000028 (0.0033)
Arrests - Quality of Life	0.00016 (0.0037)	0.00046 (0.0046)	0.000061 (0.0039)	-0.0000091 (0.0026)
Arrests - Violent Index	0.000026 (0.00015)	0.000026 (0.00015)	0.000035 (0.00018)	0.000019 (0.00012)
Arrests - Property Index	0.000091 (0.00073)	0.000089 (0.00076)	0.00013 (0.00085)	0.000068 (0.00058)
Uses of Force	0.0000024 (0.00016)	0.0000093 (0.00019)	-0.0000027 (0.00015)	0.00000059 (0.00013)
D: Active Officer Counts				
Complaints	-0.0000018 (0.000061)	-0.0000027 (0.000094)	-0.0000024 (0.000053)	-0.00000068 (0.000014)
Complaints - Failure to Provide Service	0.00000091 (0.0000046)	0.0000015 (0.000007)	0.0000012 (0.0000042)	0.00000017 (0.00000073)
Team Size (N. Active POs)	49.328 (18.194)	46.492 (17.73)	52.905 (19.068)	48.949 (17.423)
N. PO Total	53.516 (18.847)	50.468 (18.347)	56.837 (19.544)	53.504 (18.282)
N. PO Reg. Watch	37.297 (14.874)	37.688 (15.166)	39.947 (15.46)	34.968 (13.771)
N. PO Tact. Watch	9.836 (7.752)	8.815 (7.585)	10.454 (8.237)	10.206 (7.418)
N. PO Desk	3.938 (2.043)	3.654 (1.89)	3.756 (1.742)	4.309 (2.304)
N. PO Overtime - Other	1.034 (2.188)	0 (0)	0.879 (1.635)	2.001 (2.915)
N. PO Out of Unit	0.236 (0.673)	0.311 (0.742)	0.166 (0.47)	0.228 (0.735)
N. PO Overtime - Special	0.301 (1.596)	0 (0)	0.31 (1.567)	0.541 (2.139)
N. PO Overtime - VRI	0.873 (2.215)	0 (0)	1.324 (2.85)	1.25 (2.363)
N. Detectives	0.175 (0.486)	0.463 (0.723)	0.09 (0.309)	0.0022 (0.051)
N. Sgt	5.914 (2.276)	5.654 (2.26)	5.657 (2.052)	6.322 (2.387)
E: Stops and Arrests				
N. Other Non-PO	3.256 (2.04)	2.817 (1.879)	2.702 (1.731)	4.035 (2.136)
N. Prob/30+	4.477 (5.3)	1.827 (2.401)	2.317 (2.964)	8.294 (6.1)
N. Arrests	0.507 (0.953)	0.599 (1.065)	0.554 (1.009)	0.396 (0.786)
N. Nonindex Arrests	0.44 (0.89)	0.531 (1.004)	0.481 (0.944)	0.333 (0.721)
N. QOL Arrests	0.25 (0.695)	0.368 (0.854)	0.271 (0.725)	0.136 (0.473)
N. Street Stops	1.358 (2.464)	1.43 (2.256)	2.427 (3.372)	0.488 (1.114)
N. Traffic Stops	1.182 (2.307)	0.644 (1.139)	0.526 (1.075)	2.122 (3.216)
Total Observations	1836414 (0)	585304 (0)	539480 (0)	711630 (0)

Note: Table presents summary statistics of team-level data by sample period at the district-hour level, with means and standard deviations (in parentheses). Panel A displays the average shares of demographics of officers, Panels B and C contains teams' average raw and shrunken officer effects (mean normalized to near zero) estimated using equation (1) separately in each sample period, Panel D contains the set of additional controls including total team size (only counts officers on team) and the other officer counts which include non-team officers used as controls, and Panel E contains stops and arrests occurring in that district-hour.

Table C.2: First Stage Joint F-Statistics and P-Values

Endogenous Variable	PO FE Type	Pooled	2010-2012	2013-2015	2016-2019
		(1)	(2)	(3)	(4)
A: Team Avg. PO FE					
Street Stops	Raw	6118.5 (p<0.001)	2450.6 (p<0.001)	4922.3 (p<0.001)	2377 (p<0.001)
Street Stops	Shrunken	4633.2 (p<0.001)	3611.2 (p<0.001)	3345.5 (p<0.001)	2344.4 (p<0.001)
Traffic Stops	Raw	1410.2 (p<0.001)	3217.3 (p<0.001)	3322.2 (p<0.001)	973.2 (p<0.001)
Traffic Stops	Shrunken	1049.4 (p<0.001)	2370.4 (p<0.001)	3382.9 (p<0.001)	971.5 (p<0.001)
Arrests - Drug	Raw	4235.2 (p<0.001)	3155.4 (p<0.001)	1874 (p<0.001)	1765.1 (p<0.001)
Arrests - Drug	Shrunken	3402.1 (p<0.001)	2035.1 (p<0.001)	1212.3 (p<0.001)	1897 (p<0.001)
Arrests - Non-Drug Nonindex	Raw	4523.9 (p<0.001)	2902.1 (p<0.001)	3252.9 (p<0.001)	1246.6 (p<0.001)
Arrests - Non-Drug Nonindex	Shrunken	4875.6 (p<0.001)	2349.1 (p<0.001)	2104 (p<0.001)	1366 (p<0.001)
Arrests - Violent Index	Raw	4462.7 (p<0.001)	2076.1 (p<0.001)	3304.4 (p<0.001)	1166.4 (p<0.001)
Arrests - Violent Index	Shrunken	4422.1 (p<0.001)	1760.3 (p<0.001)	3179 (p<0.001)	1180.6 (p<0.001)
Arrests - Property Index	Raw	3867.9 (p<0.001)	1053.1 (p<0.001)	3059.5 (p<0.001)	1230.6 (p<0.001)
Arrests - Property Index	Shrunken	2969.8 (p<0.001)	737.5 (p<0.001)	2621.8 (p<0.001)	998.1 (p<0.001)
Uses of Force	Raw	2886.6 (p<0.001)	1875.3 (p<0.001)	1335 (p<0.001)	767.3 (p<0.001)
Uses of Force	Shrunken	4102.3 (p<0.001)	1487.2 (p<0.001)	2909.2 (p<0.001)	954.3 (p<0.001)
Complaints	Raw	1762.4 (p<0.001)	725.5 (p<0.001)	1567 (p<0.001)	956.9 (p<0.001)
Complaints	Shrunken	2027.9 (p<0.001)	1142.3 (p<0.001)	1856.7 (p<0.001)	1041.3 (p<0.001)
B: Endogenous Variable Only					
N. Arrests	Raw	113.2 (p<0.001)	37.6 (p<0.001)	37.9 (p<0.001)	48.4 (p<0.001)
N. Arrests	Shrunken	134.3 (p<0.001)	45.9 (p<0.001)	43.2 (p<0.001)	50.9 (p<0.001)
N. Nonindex Arrests	Raw	77.8 (p<0.001)	27.1 (p<0.001)	30.5 (p<0.001)	24.6 (p<0.001)
N. Nonindex Arrests	Shrunken	93.6 (p<0.001)	35.6 (p<0.001)	29.4 (p<0.001)	35.9 (p<0.001)
N. QOL Arrests	Raw	177 (p<0.001)	73.1 (p<0.001)	73.8 (p<0.001)	38.1 (p<0.001)
N. QOL Arrests	Shrunken	202.7 (p<0.001)	81.9 (p<0.001)	59.2 (p<0.001)	80.2 (p<0.001)
N. Drug Arrests	Raw	143.7 (p<0.001)	60.3 (p<0.001)	63.1 (p<0.001)	30.2 (p<0.001)
N. Drug Arrests	Shrunken	154.6 (p<0.001)	50.6 (p<0.001)	48.9 (p<0.001)	92.2 (p<0.001)
N. Street Stops	Raw	79.4 (p<0.001)	32.3 (p<0.001)	53.5 (p<0.001)	43.5 (p<0.001)
N. Street Stops	Shrunken	92.8 (p<0.001)	37.1 (p<0.001)	54.5 (p<0.001)	62.7 (p<0.001)
N. Traffic Stops	Raw	78.7 (p<0.001)	54.3 (p<0.001)	36.4 (p<0.001)	54.6 (p<0.001)
N. Traffic Stops	Shrunken	93.9 (p<0.001)	52.9 (p<0.001)	36.8 (p<0.001)	65.5 (p<0.001)
C: + Pred. Team Avg. Controls					
N. Arrests	Raw	23.8 (p<0.001)	8.2 (p<0.001)	8.4 (p<0.001)	10.2 (p<0.001)
N. Arrests	Shrunken	29.1 (p<0.001)	9.8 (p<0.001)	9.7 (p<0.001)	13.9 (p<0.001)
N. Nonindex Arrests	Raw	23.4 (p<0.001)	8.2 (p<0.001)	10.4 (p<0.001)	7.2 (p<0.001)
N. Nonindex Arrests	Shrunken	32.5 (p<0.001)	11.1 (p<0.001)	11.3 (p<0.001)	14.4 (p<0.001)
N. QOL Arrests	Raw	31.7 (p<0.001)	13.9 (p<0.001)	14.4 (p<0.001)	8.3 (p<0.001)
N. QOL Arrests	Shrunken	39.4 (p<0.001)	15.3 (p<0.001)	12.6 (p<0.001)	17.6 (p<0.001)
N. Drug Arrests	Raw	21.6 (p<0.001)	11.1 (p<0.001)	11.8 (p<0.001)	5.5 (p<0.001)
N. Drug Arrests	Shrunken	28.8 (p<0.001)	9.5 (p<0.001)	11.2 (p<0.001)	16 (p<0.001)
N. Street Stops	Raw	79.4 (p<0.001)	32.3 (p<0.001)	53.5 (p<0.001)	43.5 (p<0.001)
N. Street Stops	Shrunken	92.8 (p<0.001)	37.1 (p<0.001)	54.5 (p<0.001)	62.7 (p<0.001)
N. Traffic Stops	Raw	78.7 (p<0.001)	54.3 (p<0.001)	36.4 (p<0.001)	54.6 (p<0.001)
N. Traffic Stops	Shrunken	93.9 (p<0.001)	52.9 (p<0.001)	36.8 (p<0.001)	65.5 (p<0.001)
D: Team Avg. Characteristics					
Share Black	-	762.7 (p<0.001)	1259.6 (p<0.001)	956.7 (p<0.001)	218.9 (p<0.001)
Share Hispanic	-	1609.4 (p<0.001)	596.1 (p<0.001)	2711.4 (p<0.001)	954.7 (p<0.001)
Share Other Race	-	2653 (p<0.001)	2138.5 (p<0.001)	1256.1 (p<0.001)	1006.3 (p<0.001)
Share Female	-	1211.8 (p<0.001)	786.5 (p<0.001)	984.2 (p<0.001)	403.2 (p<0.001)
Share S. Age = [25,28]	-	1671.2 (p<0.001)	1319.2 (p<0.001)	1664.3 (p<0.001)	585.5 (p<0.001)
Share S. Age = [29,32]	-	1585.2 (p<0.001)	1208.8 (p<0.001)	1213.7 (p<0.001)	626 (p<0.001)
Share S. Age = >33	-	1866.7 (p<0.001)	2472.9 (p<0.001)	1497.4 (p<0.001)	529.2 (p<0.001)
Share Military	-	992.8 (p<0.001)	1411 (p<0.001)	793.6 (p<0.001)	446.6 (p<0.001)
Share Spanish Speaking	-	2407.7 (p<0.001)	1380.8 (p<0.001)	2890.7 (p<0.001)	944.6 (p<0.001)
Share Edu < BA	-	1391.3 (p<0.001)	761.8 (p<0.001)	978 (p<0.001)	685.6 (p<0.001)
Share Edu >= BA	-	1298.1 (p<0.001)	1068.5 (p<0.001)	1029.4 (p<0.001)	516.1 (p<0.001)
Share Tenure 5-10	-	1314.3 (p<0.001)	1285.4 (p<0.001)	1314 (p<0.001)	414.5 (p<0.001)
Share Tenure 10-15	-	1466.2 (p<0.001)	1720.1 (p<0.001)	1297.1 (p<0.001)	492.3 (p<0.001)
Share Tenure 15-20	-	1620.1 (p<0.001)	1170.3 (p<0.001)	1470.4 (p<0.001)	858.9 (p<0.001)
Share Tenure >20	-	1432.4 (p<0.001)	2287.5 (p<0.001)	649 (p<0.001)	628.7 (p<0.001)

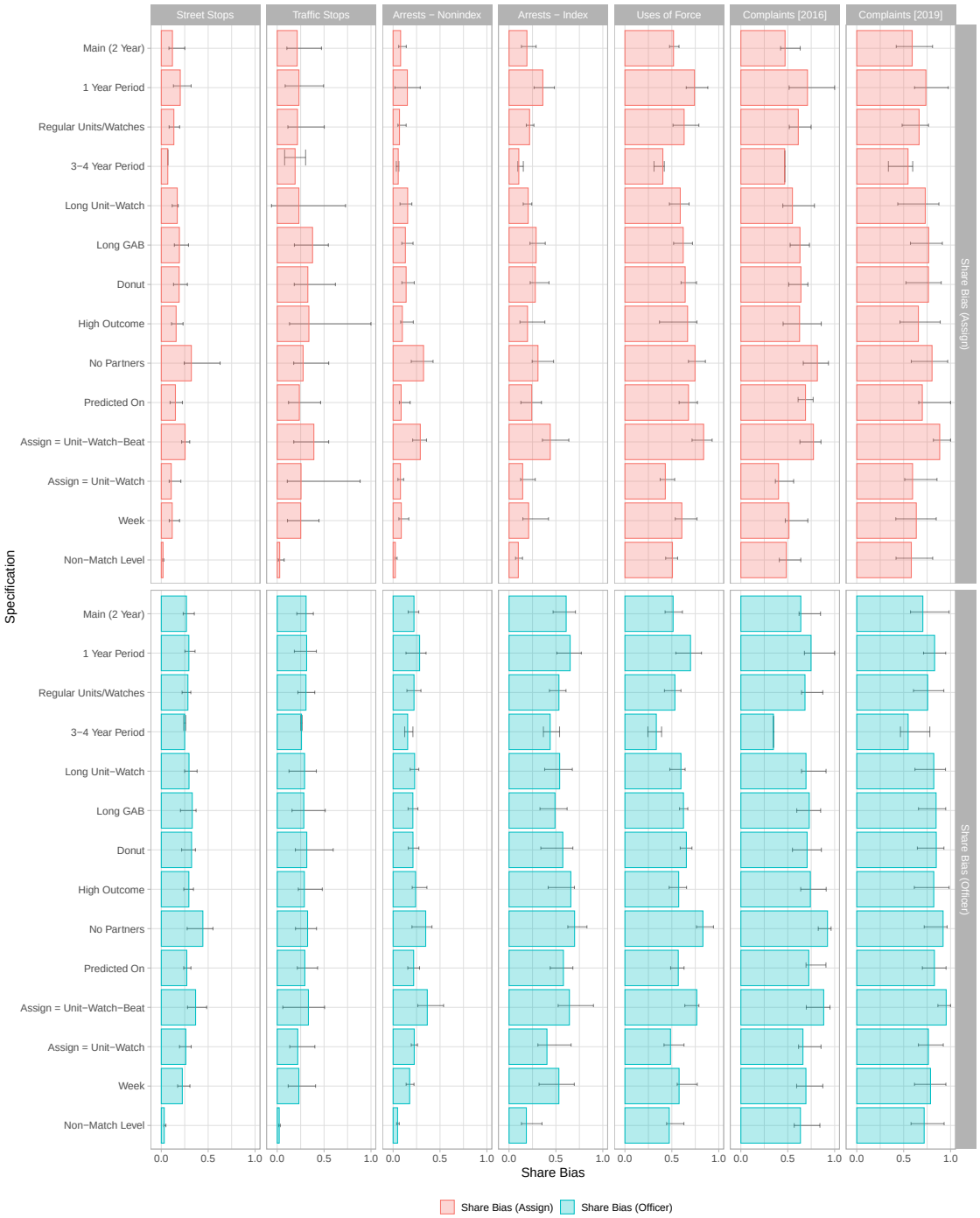
Note: Table displays joint F-test, among excluded instruments only, and associated p-values in parentheses for the first stage regressions, across endogenous variables and specifications.

D Appendix D - Additional Analyses

D.1 Bias in Plug-In Estimates

The importance of bias-correction is highlighted in Figure D.1 which displays the share of estimated bias relative to the non-bias corrected ('plug-in') variance component. For high-frequency outcomes, such as street stops, traffic stops, and nonindex arrests, generally less than 20% of the plug-in variance of officer effects is due to bias. For low-frequency outcomes' variances are significantly upwardly biased due to measurement error, particularly for complaints and use of force where the plug-in estimates are generally between 35-50% bias. Notably, bias correction is more impactful for officer effects, across all outcomes, compared with assignment effects, which is consistent with fewer observations per officer than per assignment.

Figure D.1: Share of Plug-In Variance of Officer and Assignment Effects Attributable to Bias



Note: Figure displays the share of the plug-in (non-bias corrected) variance in officer and assignment effects which are attributed to bias based on the KSS bias correction across specifications. The bars denote the median share of bias and the error bars denote the minimum and maximum share of bias across samples within the specification.

D.2 Comparison with Assignment Effects

Column (2) in Table 4 displays the bias corrected SDs of assignment effects. Compared with Column (1), we see that heterogeneity across officer effects is comparable to if not more important than heterogeneity across assignments. Over 100 shifts, 1 SD increase in assignment effects for each respective outcome would increase street stops by 28.4 (33.5 pre-2016 and 15.7 post-2016), traffic stops by 11.9, nonindex arrests by 8.55, index arrests by 1.78, uses of force by 0.249, and accusations of misconduct by 0.202 and 0.111.

For index arrests, assignment effects are generally more important with 1 SD officer effects being around 75% of 1 SD in assignment effects, consistent with index arrests being less discretionary than nonindex arrests, where 1 SD officer effect is 116% larger than 1 SD assignment effects. Additionally, officer and assignment effects explain a sizable portion of the variation in these enforcement outcomes, around 16% for stops, 14% for nonindex arrests, and 2% for index arrests, which are relatively infrequent.

D.3 Match Effects

To determine the quality of the fit using the main additive specification in equation (1), I estimate a model using officer-assignment (GAB) match effects as a comparison. Table D.1 shows that while around 10,000 parameters were estimated in the main specification (around 8,000 officer effects and 1,200 assignment effects), in the match effects model between 47,000 and 53,000 parameters were recovered. Figure D.2 compares the shares of $Var(\tilde{y}^k)$ that are explained by the officer and assignment parameters in the main additive model ($Exp^k = Var(\tilde{y}^k) - Var(\epsilon^k)$,x-axis) compared with the match effects model, either in terms of absolute increase in variation explained ($Exp^k_{Match} - Exp^k_{Add}$, top panel) and percent increase in variation explained relative to the additive model ($\frac{Exp^k_{Match} - Exp^k_{Add}}{Exp^k_{Add}}$) (y-axis).

As shown in Panel A of Figure D.2, estimating around 30,000 additional parameters results in an increase in share of variation explained of between 0 and 0.08, though this varies by outcome with the rarer and lower variation explained outcomes having smaller increases

(complaints, force, index arrests) between 0 and 0.015, nonindex arrests increasing by around 0.02, and stops and traffic stops increasing by around 0.04 and 0.07 after KSS bias correction (left figure). Note that bias correction proves more important for the low frequency outcomes as plug in variation explained (right figure) increases by around 0.02 for force, complaints and index arrests. However, these absolute changes are not proportional to the initial amount of variation explained in the additive model, which is shown in Panel B. While plug in variation explained increases massively in proportion to initial variation explained for rare outcomes (right panel), KSS bias correction shows that in most samples, complaints, use of force, and index arrests have their variation explained by around 50%, and nonindex arrests, street stops, and traffic stops have theirs increased by around 15%, 25%, and 35%.

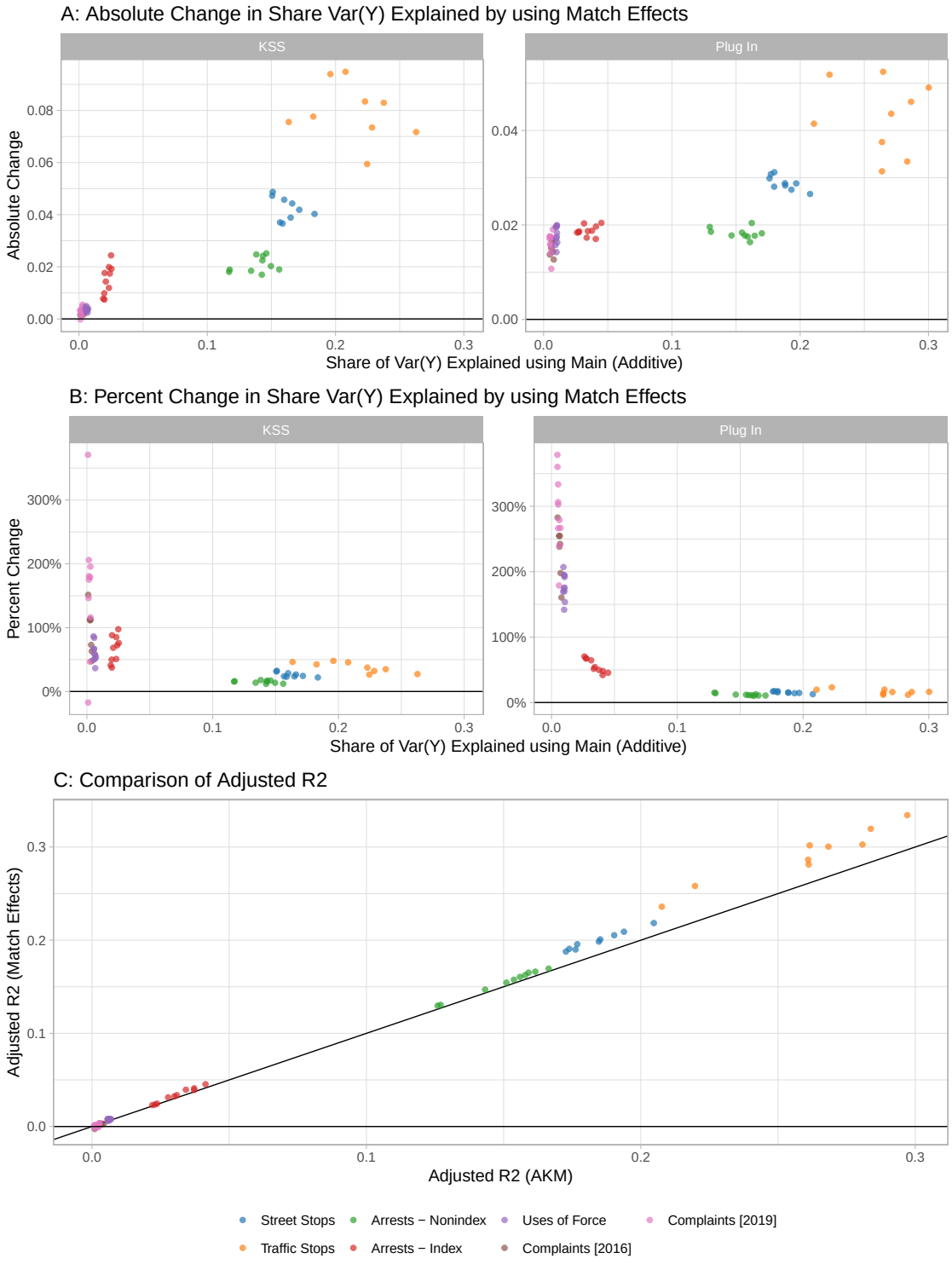
Panel C shows, after taking into account the number of parameters using adjusted R^2, the match effects model does not significantly increase adjusted R^2 for most outcomes. While these results do indicate that match effects between officers and their assignments are empirically important and worth examining, it is clear that the additive model is a useful and parsimonious approximation.

Table D.1: Comparison of Number of Parameters Estimated in Main (Additive) and Match Effects Models

Sample Period	Officer Effects	Assignment (GAB) Effects	Match Effects
	(1)	(2)	(3)
2009 - 2010	8,328	1,247	46,671
2010 - 2011	8,432	1,176	49,773
2011 - 2012	8,232	1,162	52,761
2012 - 2013	8,028	1,124	49,444
2013 - 2014	8,166	1,320	52,575
2014 - 2015	8,028	1,305	50,510
2015 - 2016	7,944	1,353	50,895
2016 - 2017	7,895	1,320	50,731
2017 - 2018	7,879	1,256	47,289
2018 - 2019	8,243	1,271	48,623

Note: Table displays the number of parameters of interest (officer effects and assignment effects) estimated in each sample period in the main specification (Columns (1)-(2)) while Column (3) show the number of officer-assignment match effects estimated when interacting officer and GAB effects.

Figure D.2: Comparison of Variation Explained in Additive (AKM) and Match Effects Model



Note: Figure summarizes comparisons in the amount of variation explained using the additive AKM model (equation (1)) compared with using a match effects model (officer-assignment effects). Panel A displays the share of variation in the outcome explained by the AKM model (x-axis) compared with the absolute change in share of variation explained when using the match effects model (match effects − AKM); Panel B has the same x-axis, but the y-axis is the percent change in share of variation explained (i.e., y-axis in Panel A divided by the x-axis). In Panels A and B, the first figure ("KSS") shows the KSS bias corrected results and the second figure ("Plug In") shows the non-bias corrected results. Panel C compares the adjusted R^2 for the AKM model (x-axis) to that of the match effects model (y-axis) and the solid black line denotes the 45-degree line. Each point represents the result from a specific outcome (color) and 2-year sample period.

D.4 Robustness and Extensions

To test the robustness of the main results to alternative specifications, I also estimate: the main specification for single year periods only; the main specification across three multi-year periods (2009-2012, 2013-2015, and 2016-2019); the main sample using more granular assignment effects (unit × watch × exact beat) and less granular assignment effects (unit × watch); aggregate data at the weekly level;[25] in addition to the subsamples on assignments with high average outcomes, officer-assignment (GAB) spells of at least 4 days, officer-unit-watch spells of at least 20 days, and dropping the first and last shifts for an officer-assignment match, samples including shifts where the officer is predicted to be working based on the calendar, samples excluding shifts where officers are working with partners (discussed in Appendix B.3).

D.4.1 Comparison across Samples

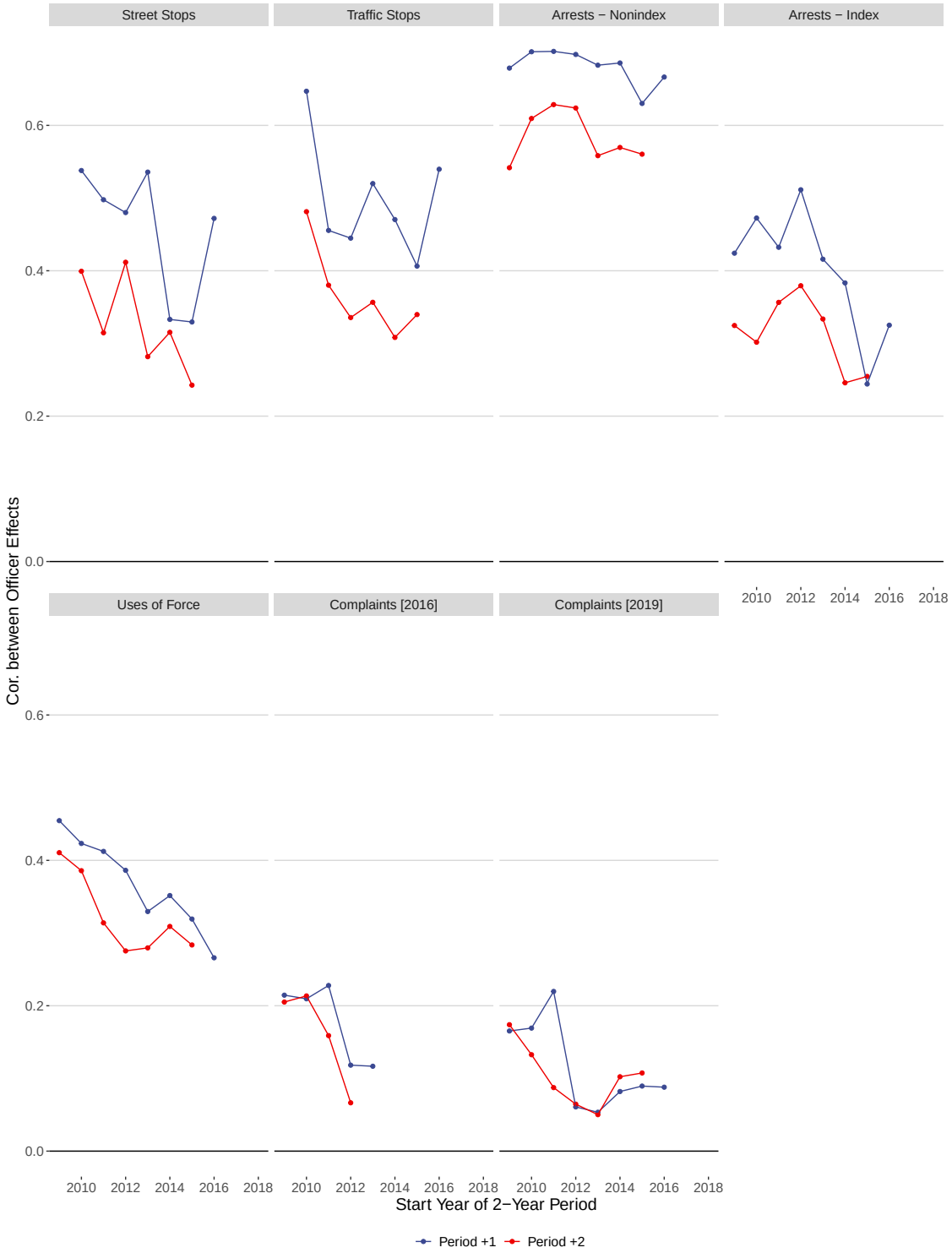
The prior results rely on recovering consistent estimates of officer effects, and the stability of these effects across time and specification is an important check to determine if they are actually measuring underlying officer behavior. To quantify the stability of the results, I first measure the correlation (not bias-corrected) between officer effects across periods of data (e.g., between the effects recovered from 2009-2010 and those from 2011-2012). These correlations are biased toward zero as the variance components are biased upwards and covariances are biased downwards. Nevertheless, as shown in Figure D.3, officer effects in one period are strongly correlated with officer effects in future periods: officer effects in period p (e.g., 2010-2011) are correlated with those in period $p+1$ (e.g., 2012-2013) above 0.6 for nonindex arrests, between 0.3-0.5 for stops, around 0.5 for traffic stops, 0.3-0.5 for index arrests, 0.3-0.4 for use of force, and between 0.05 and 0.2 for complaints. Looking 2 periods ahead (e.g., 2013-2014), correlations decline but still produce robust correlations with fixed

[25] With $y_{i,t}$ being the average outcomes per shift of officer i in week t and $J(i,t)$ being the modal assignment of officer i in week t, and $x_{i,t}$ includes both week fixed effects and the number of days an officer works in that week.

effects in the present.

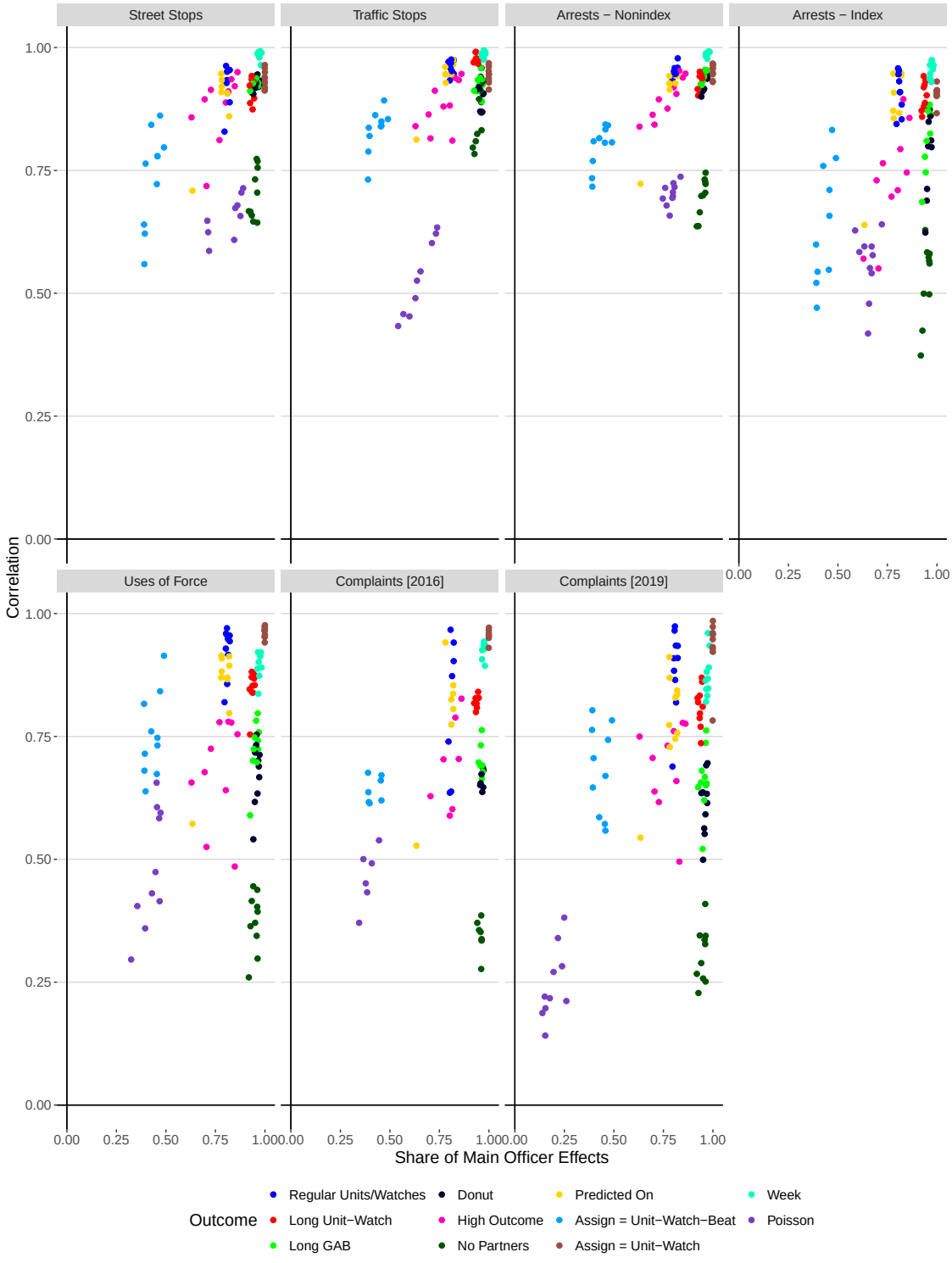
Second, the robustness tests have employed on alternative samples, and I test for the stability of officer effects across these different samples by taking the correlation between officer effects from the main sample and each alternative sample in Figure D.4. Again, even though these non-bias corrected correlations are biased toward zero, the cross-sample correlations are very strong, generally between 0.75-1 for most outcomes, with the lowest being complaints and use of force, though these are correlated with the main sample around 0.5-0.9 excluding a few outliers. The sample with the lowest correlation with the main effects are those with the most observations dropped from the total sample, No Partners and Assign = Unit-Watch-Beat, as shown in Table B.1.

Figure D.3: Correlation between Officer Effects Across Sample Periods



Note: Figure displays the (non-bias corrected) correlation (y-axis) between officer effects (main specification) in one sample period, with the x-axis denoting the first year in the two year sample period. Colors denote the correlations with one period or two periods into the future; for example, the correlation for start year 2009 and period +1 is the correlation between officer effects recovered from 2009 - 2010 and 2011 - 2012, while period +2 is the correlation between officer effects recovered from 2009 - 2010 and 2012 - 2013.

Figure D.4: Correlation between Officer Effects Across Specifications



Note: Figure displays the (non-bias corrected) correlation (y-axis) between officer effects recovered from the main specification with officer effects recovered in other specifications over the same two year sample period. The x-axis contains the share of officer effects recovered in the main specification that were also recovered in the alternative specification.

D.4.2 Comparison with Poisson

The main results utilized a linear model of officer and assignment effects with an unbounded outcome variable. However, all outcomes are count data making the prior model misspecified. Additionally, the traditional AKM model uses log-wages as the dependent variable, which makes worker and firm effects interpretable as their effects on changes rather than levels, as is the case with the specification I employ. The outcomes of this study have significant mass at zero, making a log-transformed specification infeasible, though a Poisson specification is more appropriate than using a $log(y+1)$ approach (J. Chen & Roth (2023)). Here, I employ a poisson regression to recover officer and assignment fixed effects and perform comparisons with the results of the linear model to determine if takeaways are significantly different. I estimate:

$$E[y_{it}|x_{it}, \theta_i, \psi_{J(i,t)}] = exp(\theta_i + \psi_{J(i,t)} + \beta x_{it}) \tag{18}$$

and recover estimates for officer and assignment fixed effects as with the linear model. I recover Poisson fixed effects on the rolling 2 year samples in the main specification and three multi-year periods (2009-2012, 2013-2015, and 2016-2019).

There are two main implications for estimation that differ from the linear model: 1. all observations for which there is no variation in the groups (e.g., assignments in which no arrests are made worked officers who never make arrests) are dropped and no fixed effects can be recovered, where as they are kept in the linear mode; and 2. the point estimates of fixed effects are biased due to the incidental parameters problem. The first point means that for rare outcomes, such as complaints, only a subset of effects will be able to be recovered.

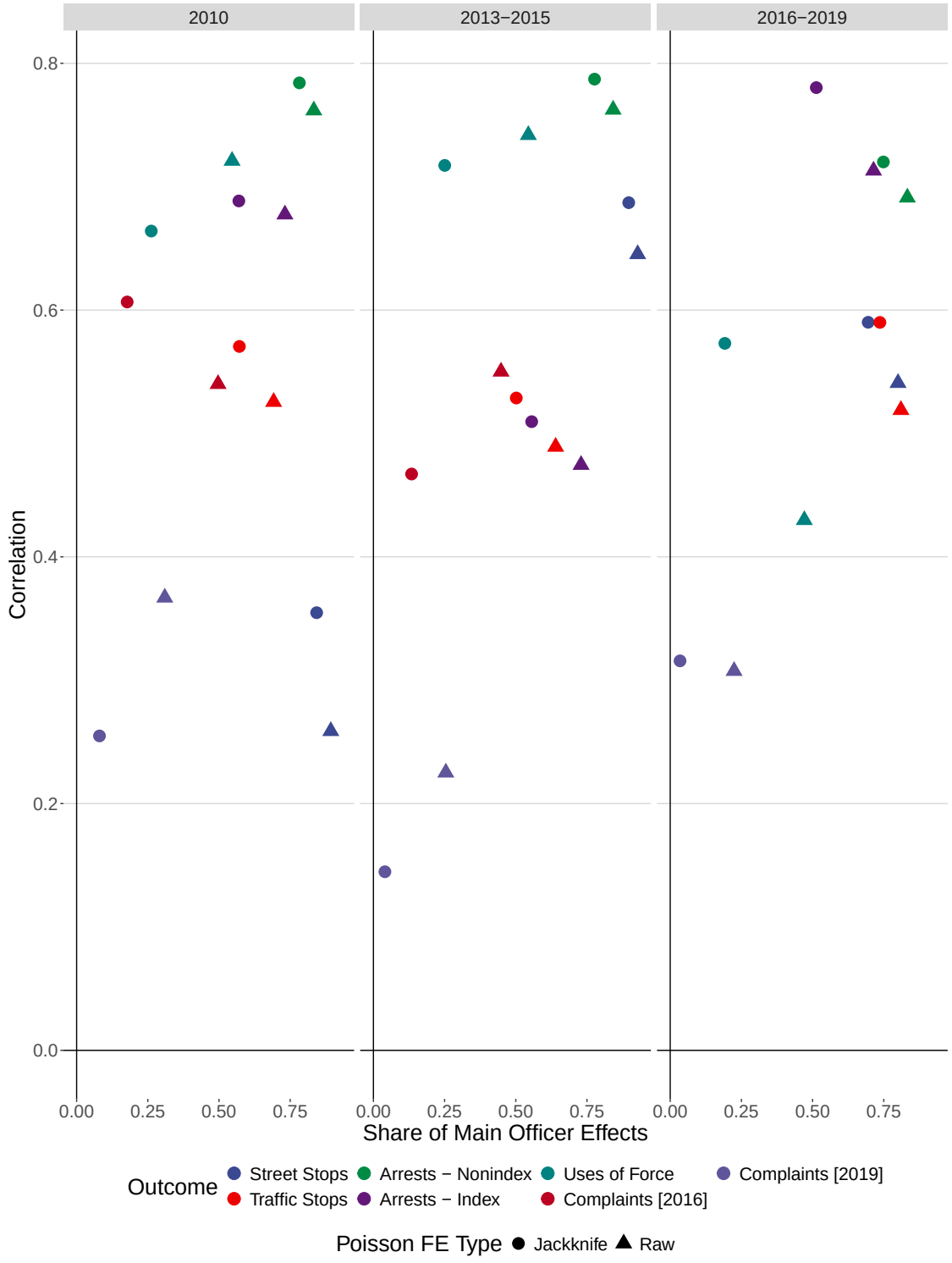
For the second I employ a version of Dhaene & Jochmans (2015) jackknife correction method for the multi-year samples by splitting each sample into two equally sized random samples and estimating fixed effects in each period (p), then bias-reduced 'jackknife' fixed effect for an officer: $\theta_i^{jk;p} = \theta_i^p - \frac{1}{2}(\theta_i^{p1} + \theta_i^{p2})$. However, because of the first issue (meaning

it is possible an effect can be recovered over the two-year period but not in one of the two when separated) $\theta_i^{jk;p1,p2}$ is more likely to be missing than θ_i^p.

Figure D.4 displays the correlations between officer effects from the main specification linear model and Poisson model officer effects (raw, not jackknifed), and Figure D.5 displays the correlation between main officer effects and Poisson model officer effects which were recovered for the long period samples (2009 - 2012, 2013 - 2015, 2016 - 2019) with both raw and jackknife fixed effects — the x-axis in both figures is the share of officer effects kept. For the 2-year main sample results, the correlations are weaker than those with other linear specifications, but still strong for high frequency outcomes, generally between 0.5 and 0.75 for stops and arrests with around 50% - 70% of the main officer effects being able to be recovered with the Poisson model. For use of force and complaints, the correlations are weaker, around 0.5 and the share kept is around 35%, though for the less frequent extended complaints the correlations are around 0.25 and only around 15% of effects could be recovered. For the longer sample results, the correlations with jackknife and raw Poisson fixed effects do not differ significantly, though there are a few cases of very low correlations generally due to the structural changes that occurred in the sample period (e.g., stops in 2009 are all zeros and traffic stops increased significantly between 2016 and 2019). Overall, the correlations tend to be between 0.4 and 0.8, and though the share of the sample that remains is lower with jackknife the correlations tend to be higher.

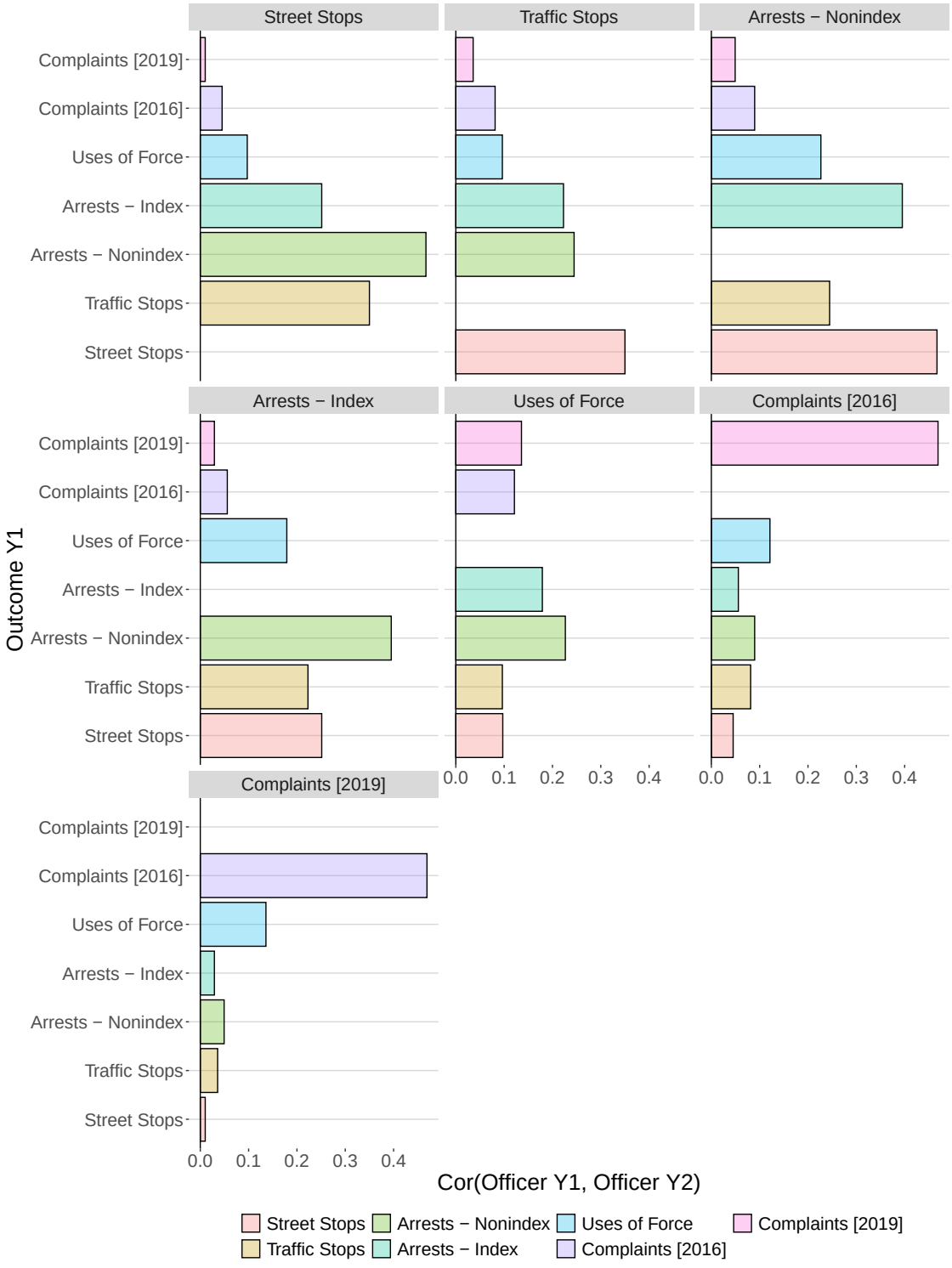
Figures D.6 and D.7 shows that the cross-outcome correlations and sorting patterns found using the Poisson model are similar to those of the main results, though biased downward as is expected due to not bias-correcting the variance and covariance components. In particular, officers only positively sort, consistently on nonindex arrests with negative correlations between officer and assignment effects from the Poisson model, and officer effects for complaints tend to be positive correlated with use of force and nonindex arrests, but only very weakly with stops and index arrests. Overall, the results are generally consistent using either a Poisson or linear model.

Figure D.5: Correlation between Officer Effects and Raw and Jackknifed Poisson Effect using Multi-Year Samples



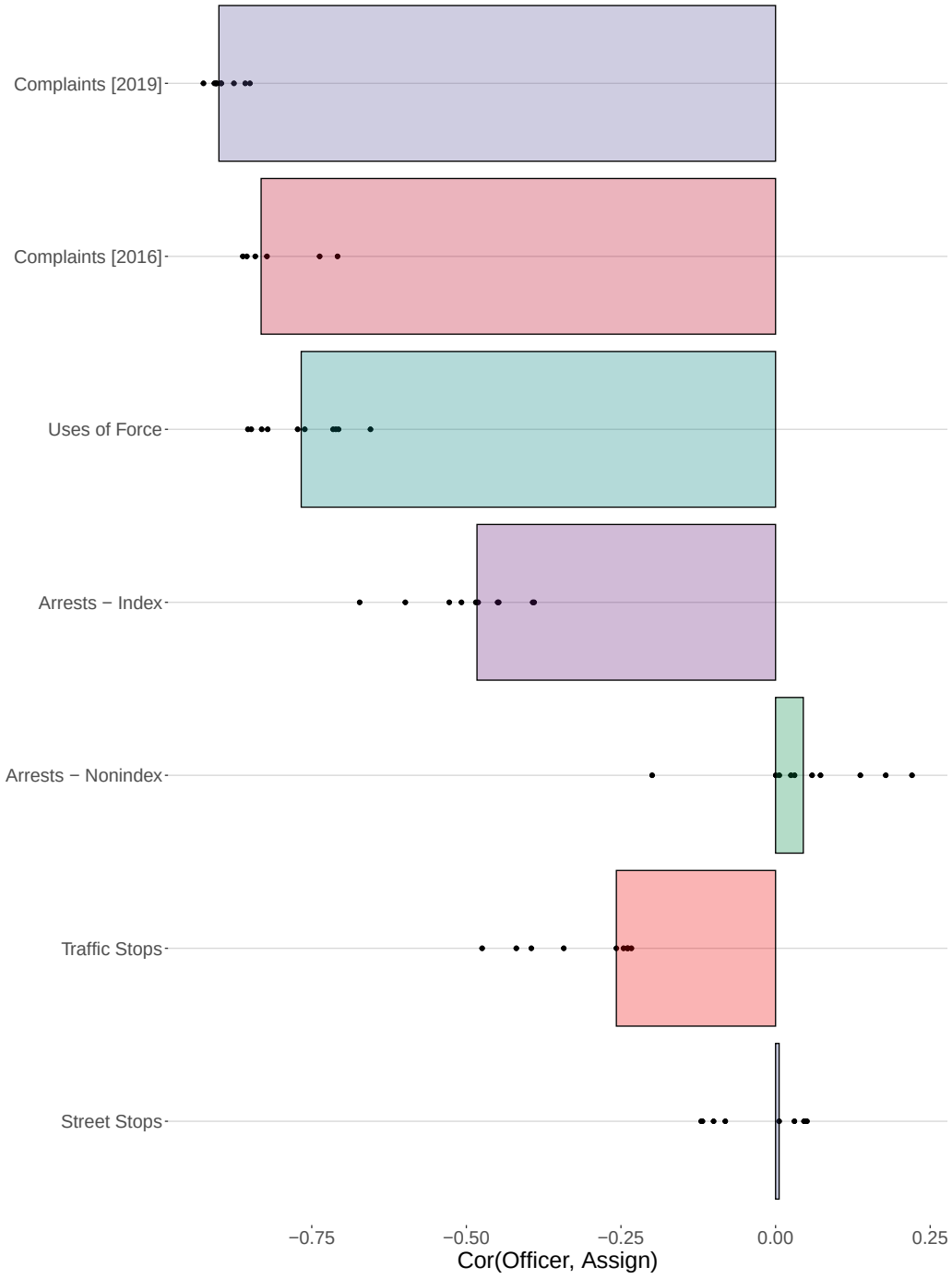
Note: Figure displays the (non-bias corrected) correlation (y-axis) between officer effects recovered from the main specification using multi-year samples (2009-2012, 2013-2015, 2016-2019) with officer effects recovered over the same sample period using a Poisson model with their raw fixed effects or jackknife fixed effects following Dhaene and Jochmans (2015). The x-axis contains the share of officer effects recovered in the main specification that were also recovered in the alternative specification.

Figure D.6: Cross-Outcome Correlations in Poisson Officer Effects



Note: Figure displays the median cross-outcome correlations among officer effects recovered using a Poisson regression across the main two-year samples.

Figure D.7: Correlation between Poisson Officer and Assignment Effects



Note: Figure displays the correlations between officer and assignment effects recovered using a Poisson regression across the main two-year samples, with bars denoting medians across samples (points).

D.4.3 Overtime

Officers earn overtime (generally 1.5 times their hourly rate) for any work after 40 hours a week, which can occur if the officer appears in court, works an extra event, or is completing an arrest which spills over beyond their regular shift. Naturally, this can create an incentive for officers to sort into assignments where they can make more overtime, and it is possible that overtime effectively extends the officers' shift which is not included in the primary analysis.

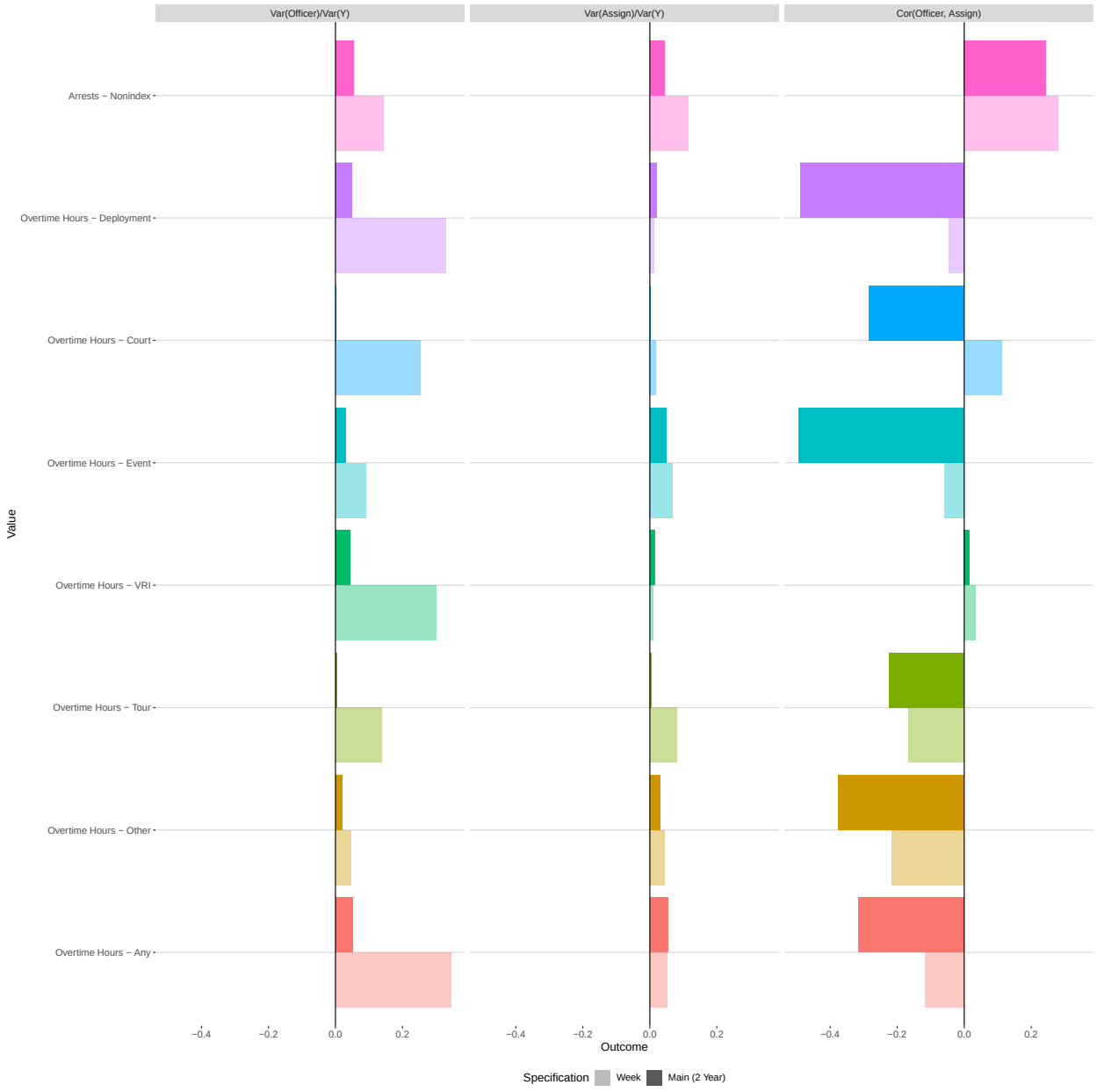
I employ additional data on overtime shifts from 2014 to 2019 to study the relationship between overtime and the prior results. I create a new outcome being the number of hours an officer works as overtime of various types. At the daily level, an officer's overtime hours are counted for their shift if they worked the overtime immediately following their shift (or overlapping with it). Because officers may also just work overtime on the next day, for things like court appearances following an arrest, I also aggregate the data at the weekly level and I count overtime as any overtime worked in that week, regardless of whether it was worked immediately after a regular shift.

Figure D.8 displays the share of variance in the number of overtime hours worked that is explained by officer and assignment effects and the correlation between officer and assignment effects for each overtime type and non-index arrests (for comparison). At the daily shift level, officer and assignment effects explain virtually none of the variation in overtime for court, but explain small amounts of variation for other types of overtime, and a small portions of overtime for deployments and VRI (violent reduction patrols). Additionally, at the shift-level, there are strong negative correlations between officer and assignment effects across overtime types. However, at the week-level, variation in officer effects explain sizable portions of the variation in overtime for deployments, court, events, VRI, and additional tours, while assignment effects explain relatively small portions. Additionally, while correlations between assignment and officer effects are generally negative (with the exception of court overtime which is small and positive), the negative correlations are much smaller than at the daily level.

These patterns indicate that officers do vary significantly in their desire to work overtime for additional policing and appearing in court, but this is generally not driven by the extension of regular shifts, but rather working additional overtime on different days or times disconnected from their regular shift. Additionally, officers are not clearly sorting into assignments which permit additional overtime, with the exception of overtime from court (at the weekly-level) consistent with officers sorting into assignments that allow for more arrests then those arrests leading to more court overtime.

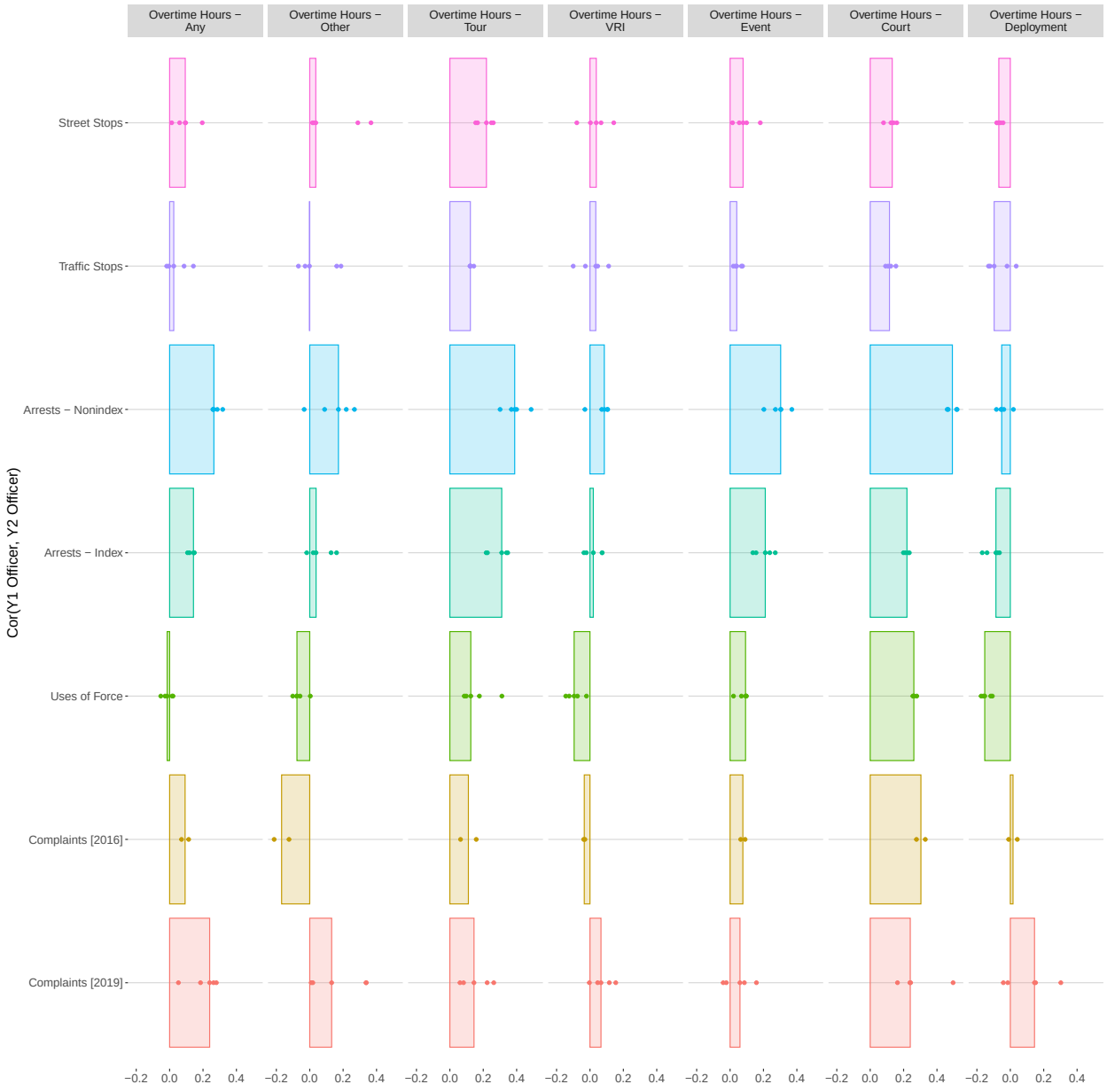
Next, given that overtime is a product of officers working additional hours, we can determine if the officers who work more overtime are productive during their regular shifts. Figure D.9 displays the cross-outcome correlations in officer effects with the other main outcomes using the weekly-level data. Overtime for activities such as deployments and VRI have only weak relationships with the primary activities. However, overtime for court, events, and tours are correlated with all primary activities. Court and tour overtime is most strongly correlated with other activities, around 0.15 for traffic stops, stops, complaints and index arrests, and around 0.2 for use of force and over 0.4 for nonindex arrests, while event overtime is less strongly correlated with such metrics though index and nonindex arrests are the largest (0.15 and 0.2, respectively). Overall, officers who are more active do work more overtime, and some of this is driven by arrests leading to court overtime, however variation in officers' arrest propensity by no means explains all of the variation in their overtime propensities.

Figure D.8: Decomposition of Variation in Officer Overtime



Note: Figure displays the median KSS bias corrected variance and correlation components across 2 year sample periods from estimating equation (1) on officer overtime, with the same statistics for nonindex arrests displayed for comparison. Shading denotes the specification, either the main specification using shift-level data or data aggregated to the week level, though both use 2-year sample periods.

Figure D.9: Correlation between Officer Effects on Main Outcomes and Overtime Hours



Note: Figure displays the median KSS bias corrected correlation between officer effects on officer-level outcomes (rows) and overtime hours worked by the officer (columns) across 2 year sample periods from estimating equation (1), with each point representing the estimate from a sample period.

D.5 Cross-Crime Correlations

We can recover the correlations between team effects on different crime types, as shown in Table D.2. For this, I use the first difference results (pooling actual and predicted compositions), as the bias correction tends to produce more reasonable correlations, though the main specification correlations are displayed in Figure D.10. Violent crime effects are correlated more strongly with effects on simple violent crime (0.45), and calls relating to domestic disturbances, assaults, and robberies (0.51, 0.4, 0.43) — similarly team effects on robberies and assaults are most strongly correlated with effects on 911 calls about robberies and assaults. Additionally, violent crime effects are positively but not very strongly correlated with property crime effects (0.26). Property crime effects are most strongly correlated with other nonindex offenses, which contains vandalism and nonindex property crimes. Notably, team effects on theft and MV theft are positively, but not strongly, correlated with corresponding calls for 911 calls (between 0.15 and 0.27). In general, team effects tend to be positively correlated across crime types, suggesting teams that are better at reducing one type of crime are not, instead, increasing another type of crime.

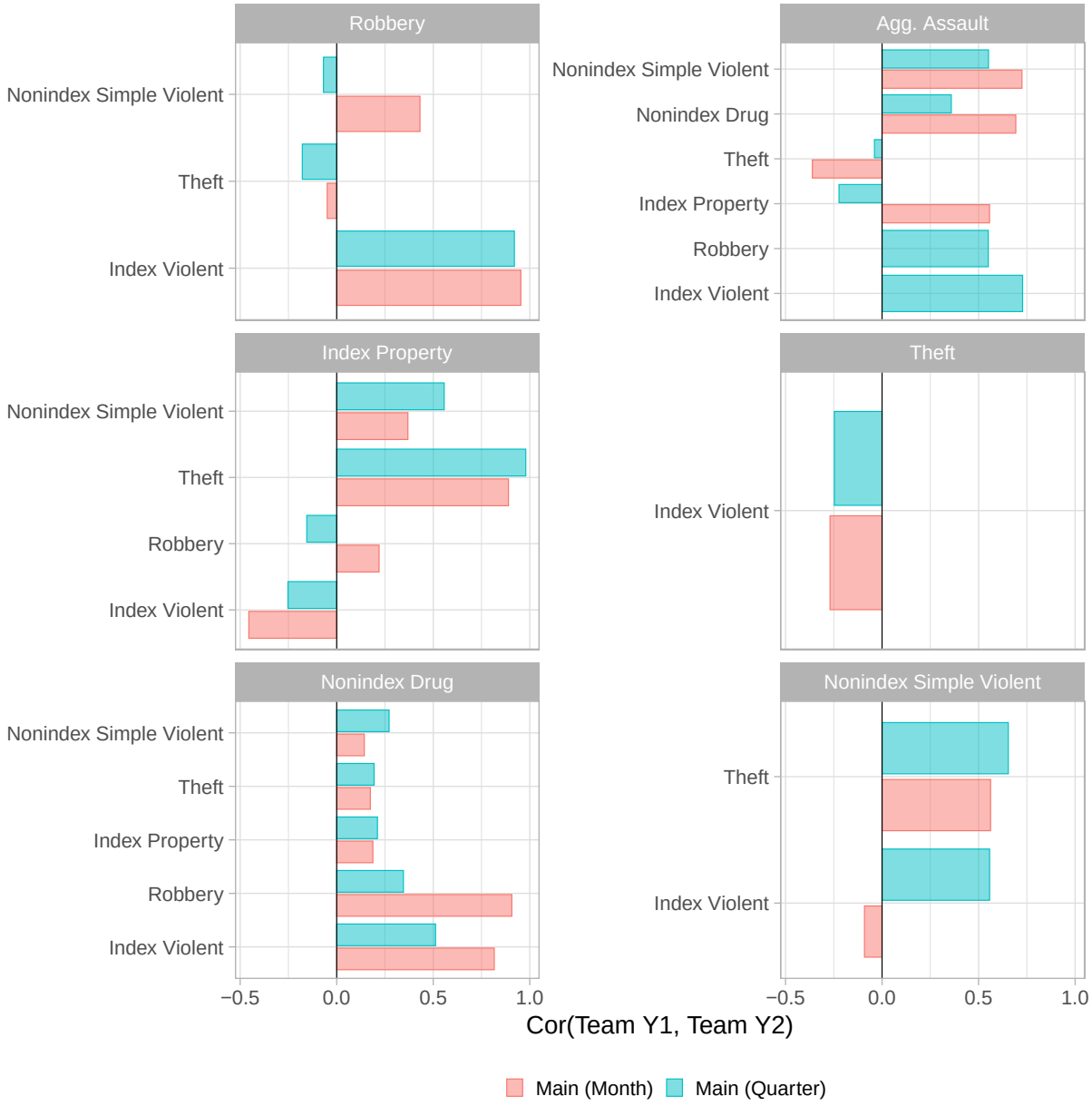
Interestingly, team effects on drug crime, which are driven entirely by discretionary arrests, are relatively weakly but positively correlated with other crime types, suggesting teams which are 'good' at making more drug arrests are not also reducing other types of crime through contemporaneous deterrence or incapacitation. Rather, drug crime effects are most strongly correlated with other 'discretionary' crime, in particular, weapons crimes with a median correlation of 0.39.

Table D.2: Cross-Crime Correlations in Officer Effects

	Reported Crimes											911 Calls				
	Murder	Robbery	Agg. Assault	Property	Burglary	MV Theft	Theft	Drug	S. Violent	Weapon	All Other	Domestic	Assault	Robbery	MV Theft	Theft
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Reported Crimes																
Violent	0.208	0.755	0.831	0.26	0.125	0.251	0.266	0.16	0.454	0.274	0.319	0.51	0.399	0.429	0.137	0.244
	[0.1, 0.51]	[0.47, 0.92]	[0.73, 1]	[-0.15, 0.45]	[-0.12, 0.18]	[-0.029, 0.58]	[-0.24, 0.41]	[0.02, 0.44]	[0.32, 0.76]	[0.074, 0.65]	[-0.044, 0.44]	[0.3, 0.67]	[0.25, 0.6]	[0.3, 0.8]	[-0.17, 0.41]	[0.013, 0.53]
Murder		-0.0176	0.0894	0.0111	0.0882	-0.0834	0.0334	0.0202	0.0332	0.0903	0.0414	0.0989	0.0526	0.0702	-0.0572	0.125
		[-1.6, 1.7]	[-0.14, 0.64]	[-0.21, 0.15]	[-0.14, 0.44]	[-1.2, 0.13]	[-0.27, 0.14]	[-0.036, 0.11]	[-0.036, 0.6]	[-0.09, 0.23]	[-0.042, 0.22]	[-0.25, 0.64]	[-0.34, 0.37]	[-0.33, 0.97]	[-0.2, 0.5]	[-0.45, 0.37]
Robbery			0.449	0.0454	-0.036	0.0817	0.0598	0.143	0.434	0.377	0.0687	0.369	0.279	0.598	0.387	0.257
			[0.22, 0.84]	[-0.52, 0.34]	[-0.16, 0.5]	[-0.13, 0.57]	[-0.85, 0.33]	[0.034, 0.38]	[0.21, 0.9]	[-0.13, 1]	[-0.35, 0.36]	[-0.003, 0.43]	[0.025, 0.59]	[0.43, 0.97]	[0.003, 1.2]	[0.014, 1.4]
Agg. Assault				0.145	0.0749	0.18	0.128	0.117	0.367	0.209	0.138	0.485	0.413	0.18	-0.00313	0.147
				[-0.14, 0.33]	[-0.58, 0.13]	[-0.15, 0.41]	[-0.083, 0.32]	[-0.012, 0.49]	[0.21, 0.71]	[-0.045, 0.33]	[-0.1, 0.31]	[0.23, 0.59]	[0.3, 0.63]	[0.079, 1]	[-0.63, 0.28]	[-0.15, 0.41]
Property					0.658	0.621	0.961	0.0518	0.289	0.102	0.759	0.192	0.225	0.0623	0.0405	0.199
					[0.56, 0.68]	[0.35, 0.84]	[0.92, 0.99]	[-0.11, 0.14]	[-0.055, 0.4]	[-0.055, 0.34]	[0.6, 0.86]	[0.033, 0.35]	[0.13, 0.4]	[-0.13, 0.36]	[-0.024, 0.26]	[0.09, 0.45]
Burglary						0.281	0.478	-0.00745	0.172	-0.0473	0.461	0.0858	0.0976	-0.00778	0.0181	-0.0485
						[-0.027, 0.38]	[0.27, 0.62]	[-0.092, 0.078]	[-0.17, 0.37]	[-0.18, 0.24]	[0.34, 0.59]	[-0.24, 0.33]	[-0.002, 0.18]	[-0.15, 0.27]	[-0.18, 0.39]	[-0.22, 0.12]
MV Theft							0.491	0.057	0.149	0.139	0.537	0.288	0.0789	0.0177	0.153	-0.0218
							[0.22, 0.92]	[-0.14, 0.28]	[-0.14, 0.49]	[-0.12, 0.9]	[0.28, 0.97]	[0.12, 0.34]	[-0.018, 0.46]	[-0.12, 0.52]	[-0.17, 0.3]	[-0.2, 0.035]
Theft								0.0542	0.266	0.128	0.699	0.18	0.244	0.0814	0.0166	0.272
								[-0.097, 0.12]	[0.005, 0.38]	[-0.014, 0.22]	[0.55, 0.87]	[0.063, 0.29]	[0.14, 0.43]	[-0.25, 0.33]	[-0.13, 0.38]	[0.16, 0.57]
Drug									0.147	0.387	0.164	0.145	0.137	0.191	0.0176	0.0639
									[0.07, 0.39]	[0.17, 0.74]	[-0.033, 0.24]	[0.1, 0.34]	[0.067, 0.42]	[-0.052, 0.46]	[-0.086, 0.47]	[-0.029, 0.43]
S. Violent										0.337	0.318	0.604	0.669	0.246	0.244	0.361
										[0.045, 0.99]	[0.23, 0.39]	[0.49, 0.69]	[0.56, 0.9]	[0.006, 1]	[-0.023, 0.69]	[-0.013, 1.4]
Weapon											0.165	0.288	0.238	0.325	-0.0967	-0.187
											[-0.006, 0.34]	[0.27, 0.86]	[0.099, 0.48]	[0.041, 0.66]	[-1.3, 0.33]	[-0.37, 0.41]
All Other												0.284	0.244	0.0461	0.0505	0.0647
												[0.21, 0.44]	[0.06, 0.5]	[-0.008, 0.22]	[-0.054, 0.11]	[-0.016, 0.24]
911 Calls																
Domestic													0.581	0.171	0.101	0.198
													[0.45, 0.71]	[0.067, 0.46]	[-0.11, 0.2]	[0.16, 0.42]
Assault														0.318	0.151	0.375
														[0.073, 0.56]	[-0.009, 0.27]	[0.13, 0.97]
Robbery															0.104	-0.000479
															[-0.07, 0.94]	[-0.28, 0.52]
MV Theft																0.14
																[-0.98, 0.33]

Note: Table presents cross-outcome correlations for team membership effects on crime and 911 calls, using the first difference specification in equation (6) with actual and predicted team membership pooled. The first row for each statistic refers to the median estimate across three sample periods and pooled across specifications, and the second row displays the minimum and maximum values.

Figure D.10: Cross-Crime Correlation in Team Effects



Note: Figure displays the KSS bias corrected correlations between team effects (Ω) on crime for a subset of crime types. Estimates of team effects recovered from estimating equation (6) on three multi-year samples, and the bars denoting the median of these estimates. Main specifications are at the district-hour level with controls for year-district-day of week-8 hour period interacted with either month or quarter.

D.6 Alternative Team Specifications

We can probe the robustness of these results and quality of the additive specification by employing alternative empirical strategies. First, there is the concern that using a single effect for each officer (ω_i) causes significant misspecification because the officers may be working in different roles. Similarly, it is possible that officers may be correlated with their supervisors, i.e., the sergeants on duty, which are actually driving the effects.

To account for these potential issues, I extend the model to allow for two teams — e.g., one for regular officers and one for tactical team officers, one for officers and one for sergeants — such that:

$$\Omega^1_{\tau(d,h)} + \Omega^2_{\tau(d,h)} = \frac{1}{N^1_{\tau(d,h)}} \sum_{i \in \tau^1(d,h)}^{N^1_{\tau^1(d,h)}} \omega^k_{i,1} + \frac{1}{N^2_{\tau(d,h)}} \sum_{i \in \tau^2(d,h)}^{N^2_{\tau^2(d,h)}} \omega^k_{i,2} \tag{19}$$

So, in the regular vs. tactical team case, each officer in the original specification can have two ω_i's, one for when they work as a beat officer and one for when they work as a 'special' (non-beat) officer. In the case of adding sergeant controls, this simply involves constructing a separate 'team' of the sergeants on duty.

If the initial model is misspecified, then we expect that after bias correction, the total amount of variation explained in the model and the variance attributable to the officer components (after bias correcting all variance and covariance components) should increase. For example, we expect that given

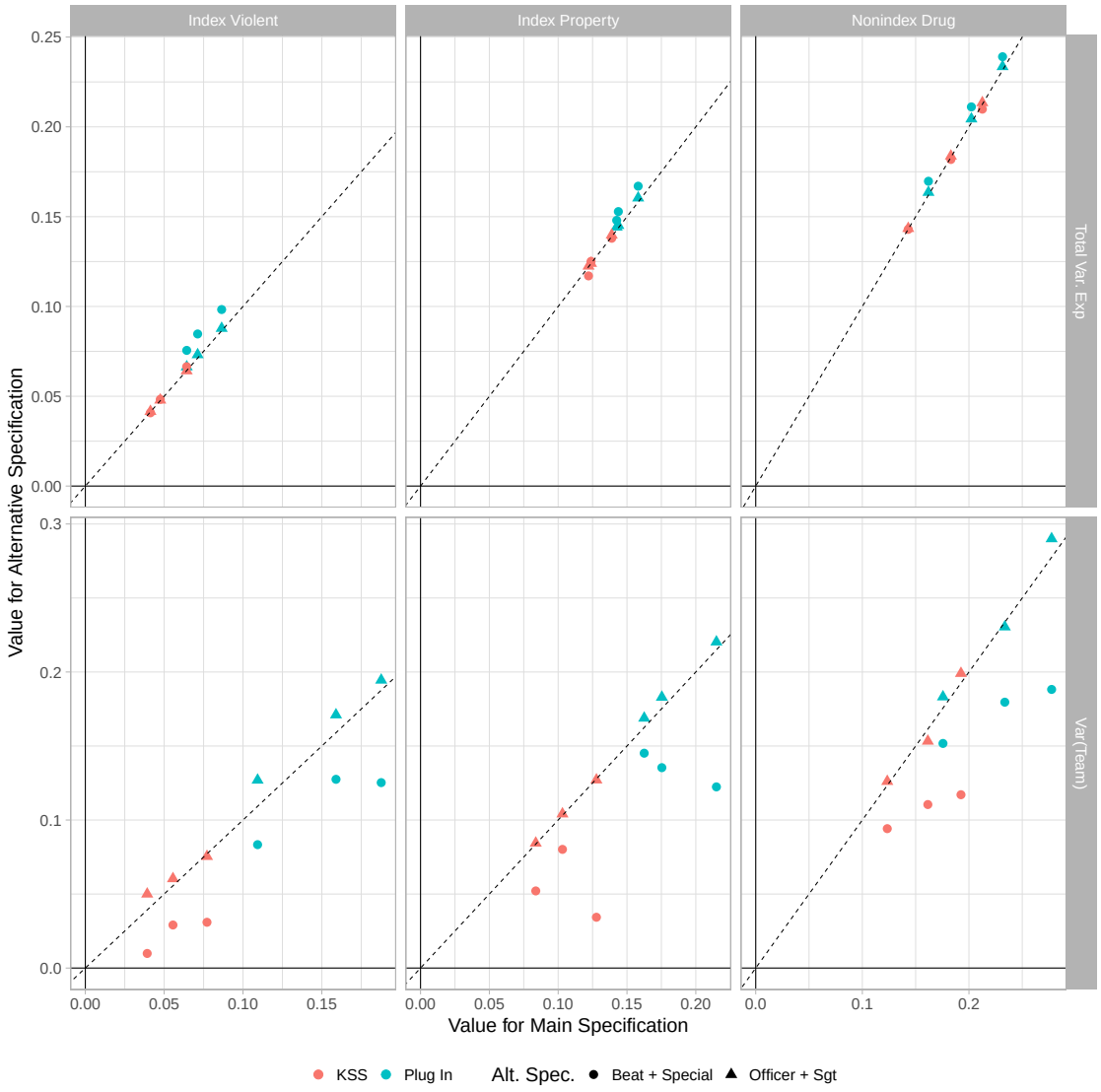
$$Var(Y) = \underbrace{Var(\Omega^1) + Var(\Omega^2) + Var(\gamma) + 2Cov(\Omega^1, \Omega^2)}_{Var(\Omega)} + 2[Cov(\Omega^1, \gamma) + Cov(\Omega^2, \gamma)] + Var(\epsilon)$$

then $Var(\Omega)$ should be larger than $Var(\Omega)$ recovered from the model in which there is no additional officer type (as in equation (7)) and total variation explained by $(Var(Y) - Var(\epsilon))$ should also be larger with more officer types.

Figure D.11 displays both total variation explained $(Var(Y) - Var(\epsilon))$ and the total

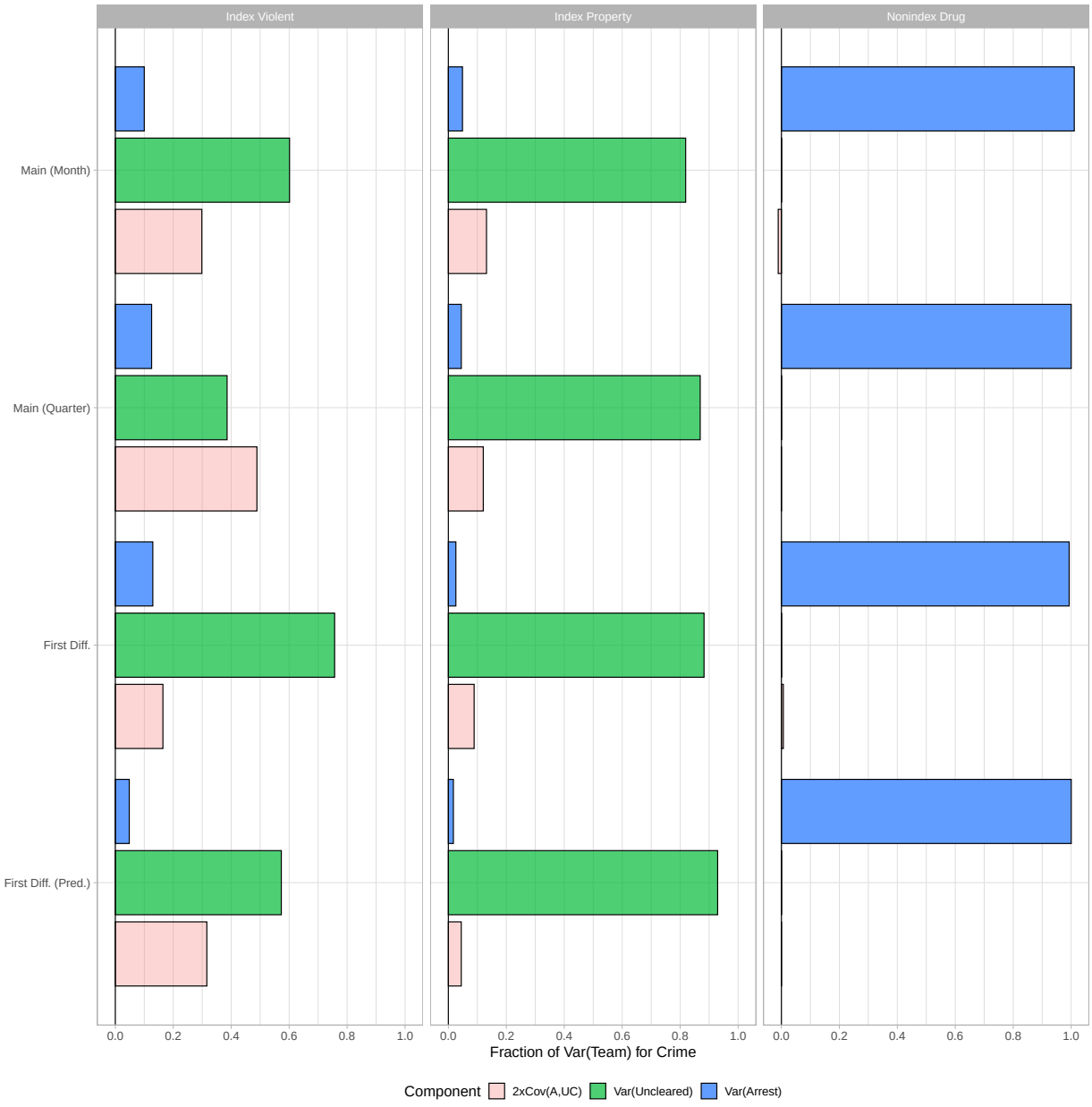
variation of officer components ($Var(\Omega)$) compared with these values in the main specification with a single officer type. As expected given the larger number of parameters, the multi-officer model has higher variation explained before bias correction. After bias correction, there is little evidence that the multi-officer models explain more variation than the single officer model. Additionally, the total variation in the officer components is not consistently and significantly larger after bias correction, suggesting that while multi-officer models can increase variation explained it the single officer model is more efficient given finite data and not significantly misspecified.

Figure D.11: Comparison between Main and Alternative Team Constructions



Note: Figure displays comparison of total variation explained and the variance of the team composition ($Var(\Omega)$) when using the main specification (teams composed of officers, equation (5)) on the x-axis with the same estimates when using alternative team constructions. Colors denote whether the estimate is KSS bias corrected or not ('Plug In'). Shapes denote the alternative team constructions. Each point corresponds to a specification with or without KSS bias correction in a specific sample (multi-year periods). All specifications use DYQDW fixed effects.

Figure D.12: Decomposing Team Effects on Crime into Arrests and Uncleared Crimes



Note: Figure displays the KSS bias corrected variance and covariances for team effects (Ω) for the components of total crime (arrests and uncleared crimes) scaled by the overall variance in team effects on crime. Estimates of team effects recovered from estimating equation (6) on 3 multi-year samples, and the bars denoting the median of these estimates. Main specifications are at the district-hour level with controls for year-district-day of week-8 hour period interacted with either month or quarter. Variance components can be negative due to bias correction.

E Appendix E - Data

E.1 Assignment and Attendance Data

The primary data for this study is daily assignment and attendance records for all units in the CPD between 2009 and 2019. This data contains officer-by-shift information, where a shift is a working spell by the officer generally for around 9 hours. This data indicates whether the officer was or was not present for duty, the reason for their absence (e.g., it was their regular day off, a vacation day, they were in training), their start and end times, their title (e.g., police officer, sergeant, detective), and their assigned beat, watch, and unit, among other information.

Initially, this data contains over 35 million entries including days off. I restrict my sample to days where the officer is present for duty with a known assignments and to the sample to officers who are the lowest and most common rank of "police officer" (around 80% of the CPD). They must also be on regular assignments (e.g., not on desk duty, training, SWAT teams, or serving on lockup). I remove all observations from units in which are uncommon and where officers do not do much measurable police work (e.g., no arrests occur in training units), and I remove all officers who could not be matched to other data sets. Finally, I remove probationary officers or those in the academy (less than 1.5 years of tenure) and a small number of observations with officers with reported tenure of 30 years or more. These filters result in 13,928,210 officer-shift observations, with 13,400 unique officers, in 54 units (25 of which are patrol units that correspond to districts).

Table 1 displays summary statistics for this sample at the officer-shift level, with Column (1) summarizing the full analysis sample and Column (2) averaging over each year summarized individually. The average officer is observed for 1,547.092 shifts. Within a single year, the average officer works 182 in-sample shifts, given they work at least 1 day. Most assignments (78.5%) are in geographic (patrol) units, and they work in an average of 15.4 unique assignments (GABs), with 45.7% of an officer's shifts being in their most common

assignment.

E.2 Officer Demographics

I link this assignment data to officer personnel records, containing demographic information (race, gender, age, tenure), as well as supplementary information on officers' experiences and skills (whether they have military experience, their education level, and whether they speak Spanish). As shown in Table 1, the sample of officers is plurality white (48.4%), with the remainder being Hispanic (24.6%), Black (23.3%), and Asian/Native American/other race (3.7%). As with many police departments, a majority of officers are male (80.8%). Officers tend to start in their late 20's around 28.9 years old, and the average officer in sample is around 41.5 years old and with 12.6 years of experience (tenure). Around 20.1% of officers are classified as having military experience and 1.2% served in the military from 2014 to the present, and 14.8% speak Spanish.[26] For around 70% of the sample, I have educational history information, and 30.7% of officers have a BA or higher while 37.6% have less than a BA (e.g., associates degree, college credits).

E.3 Officer Behavior and Performance

To measure the many dimensions of officer-level behavior and performance, I collected data on nearly all possible officer-level activity dimensions considered in policy and academic research: arrests, stops, use of force, and misconduct. I link the daily assignment records to these officer activities. Panel D of Table 1 summarizes the means of the primary outcomes per 100 shifts, while Figure E.1 displays the average outcomes per 100 shifts by year. Table E.1 displays means and standard deviations for the primary and supplementary outcomes

[26]I use two incomplete measures for military experience. First, labeled 'Military' denotes whether an officer was included in a roster provided by the CPD in response to a request for all officers who were hired with military experience — while the CPD did officially deny the request, they did provide a roster and the roster only includes a subset of officers who were actually hired within the time period. Additionally, this military roster overlaps with the set of officers who had military absences during the same period. Second, 'Military (Active)' denotes that the officer was listed in a response by CPD on which officers served in the military (presumably actively) between mid-2014 to May of 2024.

per 100 shifts.

E.3.1 Street and Traffic Stops

First, the most common outcomes are 'street stops' and 'traffic stops'. Both types of stops can be considered a metric for contact with suspects and non-suspects alike, and they are often used in "broken windows" and proactive policing to measure officer performance, as was the case in Chicago under Superintendent McCarthy (2011 - 2015).[27] The data for both of these outcomes starts in 2010 in the records obtained for this project.

'Street stops' refers to contact cards (CC), which were used from 2010-2015, and investigator stop reports (ISR), which were used from 2016-2019 excluding those classified as 'traffic'. In the sample of shifts, there are 35 street stops per 100 shifts. In late 2015, following a report on CPD stop and frisk (ACLU (2015)) by the ACLU, the CPD agreed to collect additional information on their stops replacing CC with ISR, which took effect on January 1st, 2016. As in Figure E.1, street stops rose significantly from 2010 to 2015 from just around 30 to over 60 per 100 shifts. However, in 2016, stops declined massively to less than 15 per 100 shifts and increased only modestly to just less than 20 per 100 shifts in 2019.

'Traffic stops' are the second most common outcome, with an overall mean of 12.5 per 100 shifts. This data was obtained separately FOIA-ing for traffic stops rather than contact cards, and the CPD often categorizes them separately. Traffic stops became significantly more common in 2016, rising from less than 10 per 100 shifts between 2010 and 2015, to 20 in 2017 and around 35 by 2019. As discussed in Hausman & Kronick (2023), this shift occurred because the ISR reform removed 'pedestrian' stops (CC then ISR) as a feasible activity (ISR cards were significantly more involved) to be used as a performance metric or target suspects. (Note that while a 'stop' can be attributed to up to two officers, the traffic stop data only

[27] In 2011, Garry McCarthy started as the Superintendent of Police in Chicago. This is notable because McCarthy brought a different perspective on policing from his experience under Bill Bratton in the New York Police Department, which involved implementing CompStat and following "broken windows" policing (Moser (2012)). This resulted in an emphasis on proactive policing and measuring performance, and the CPD began relying heavily on stop-and-frisk (Skogan (2022)). In 2012, under McCarthy, there was a districts reorganization with three smaller districts being dissolved and merging with larger districts.

includes on officer per event.)

E.3.2 Arrests

Next, I link data on officers' arrests from 2009 to 2019. Arrests are the primary metric for enforcement are and often used to measure both departmental and officer productivity. Arrest data include both arrestee information, crime types, and arresting officer type (primary, secondary, and auxiliary/supporting). In the main results, I credit an officer for an arrest if they are listed as a primary or secondary arresting officer, and I divide arrests into two general categories: arrests for index crimes, 'index arrests', (violent: murder, robbery, rape, aggravated assault; property: burglary, motor vehicle theft, larceny/theft, arson) and arrests for non-index crimes, 'nonindex arrests', (e.g., drugs, prostitution, simple battery, vandalism, municipal code violations, traffic offenses). Nonindex arrests are considered more discretionary and often do not have a victim, where as all index crimes have a victim and are considered less discretionary.

Index arrests are less common on average than nonindex arrests, at around 1.6 per 100 shifts compared with 11 per 100 shifts, and follow different patterns over time. Index arrests decline from around 2 to 1.5 per 100 shifts between 2009 and 2011, then are relatively stable until 2017 after which they decline slightly. In contrast, nonindex arrests steadily decline from 15 per 100 shifts in 2009 to just under 10 per 100 shifts in 2015; in 2016 they significantly drop to around 7 per 100 shifts and trend upward from there until 2019.[28]

[28]Figure E.2 displays the trends in index and nonindex arrests and their subtypes over time. Total arrests have fallen over the decade in the sample. Index arrests have fallen by around 25%. However, this is due to high numbers during 2009, but from 2011 to 2019 they have not changed significantly, declining by around 15% for property arrests and no real change for violent arrests. Nonindex arrests have fallen consistently over the time period, by around 30%. While arrests for warrant and simple violent crimes have stayed relatively constant, the overall decline has been driven by declines in drug, municipal, and other nonindex offenses, while weapon and traffic arrests have risen slightly.

E.3.3 Use of Force and Misconduct

Beyond contacts and enforcement, police use of force and misconduct are notable metrics for negative outcomes which policy attempt to mitigate and avoid whenever possible. Uses of force by officers are collected using Tactical Response Reports (TRR) data from 2009 to 2019 which contain all instances when officers report using force on a subject, ranging from relatively minor force (e.g., pressure points, defensive techniques) to gunshots and tasers (which are relatively rare). Uses of force occur at around 0.36 per 100 shifts, but were close to 0.4 between 2009 and 2012 the declined until 2018 to around 0.25 with the sharpest drops being after 2012, 2015, and 2017. See Holz et al. (2023) for more details on use of force in the CPD.

Officer misconduct is measured using complaint data which contains complaints filed against police officers by either civilians or fellow officers — though civilian complaints comprise the vast majority of complaints ($> 85\%$). The data on complaints is sourced from two different oversight regimes. The complaint data which covers 2009 to 2015 (in full) was obtained from Independent Police Review Authority (IPRA) in November of 2016 by the Invisible Institute, which are referred to as 'Complaints [2016]', and occur at a mean of 0.24 per 100 shifts between 2009 and 2015. In 2017 IPRA was largely replaced by Civilian Office of Police Accountability (COPA) following the post-Laquan McDonald reforms,[29] and I obtained an extended data set on complaints ('Complaints [2019]') from the CPD which contains the full breadth of the data. The 2019 data include fewer complaints overall than the 2016 data, with a mean of 0.096 per 100 shifts. The dynamics differ across the data sets, but both decline from 2009 to 2016, and the 2019 complaints increase slightly around 0.06 to 0.09 per 100 shifts after.

[29] In late 2015, the Laquan McDonald scandal began following the release of the video of his murder at the hands of CPD officer Jason Van Dyke. This resulted in significant political changes in Chicago, a DOJ investigation and report, and the major reform of the oversight agency in Chicago among other changes. The change in oversight agencies resulted in a shift in the complaints data as the different regimes produced different data — see Rivera & Ba (2023) for more information.

E.4 Police Teams

To connect officers working at a given time in a given place with the outcomes that are not directly observed at the officer-level (e.g., crimes), I construct teams of police officers who are working together within the same location (district) at the same time (hour). Police teams are constructed using the raw assignment data with similar filters but with important differences. The unit of observation is the district-hour (e.g., district 12 on June 15th, 2013, at 2pm), and I allocate officers to 'teams' (the set of officers working in a specific district at a specific hour) based on the start and end times of their shifts when they are present for duty, which is recorded in the assignment data.

I restrict the analysis to only the geographic units which map directly to police districts (1-25), and count officers as active if they are present for duty at a specific time based on their start and end times (rounded to the nearest hour).[30] For team composition, I only count officers working as police officers (the focus of the prior sample), though I do record the number of other ranks active by unit-hour to use as controls (e.g., number of detectives).

After preliminary cleaning, the team data contains 2,032,449 teams (district-hour observations). There are 11,686 unique officers, with most officers working for 9 consecutive hours, and the average officer being observed on 8,485.699 teams. The average team contains 48.8 officers, however the size of teams can vary significantly within the same district over the course of the day as shown in Figure B.3.

Average composition of teams along with standard deviations are displayed in Table 2. The average composition of teams reflects the composition of officers, similar to Table 1. Most officers are regular officers rather than tactical team officers (40 vs. around 10). 4 officers are assigned to desks on average, and while others are working overtime of some sort.

[30] I remove officers who do not have reported start or end times, and include observations officers working overtime (only available for 2014-2019) but within the unit. I remove officers who are reported to be in court, meetings, lockup, absent, desk officers, or working as an field training officer, but I record total manpower to use as a control (e.g., number of desk officers). To ensure sufficient observations, I remove a small number of officers within a year that have fewer than 40 observations (hours), and I remove a small number of unit-months which do not have sufficient observations, days which have too few observations, and all observations in the months where units were collapsed.

On average there are very few detectives working in these patrol districts, though there are on average 6 sergeants and 3-4 other non-police officers working in a given hour, and 4 officers excluded from the team due to tenure (e.g., probationary officers).

E.5 Crime

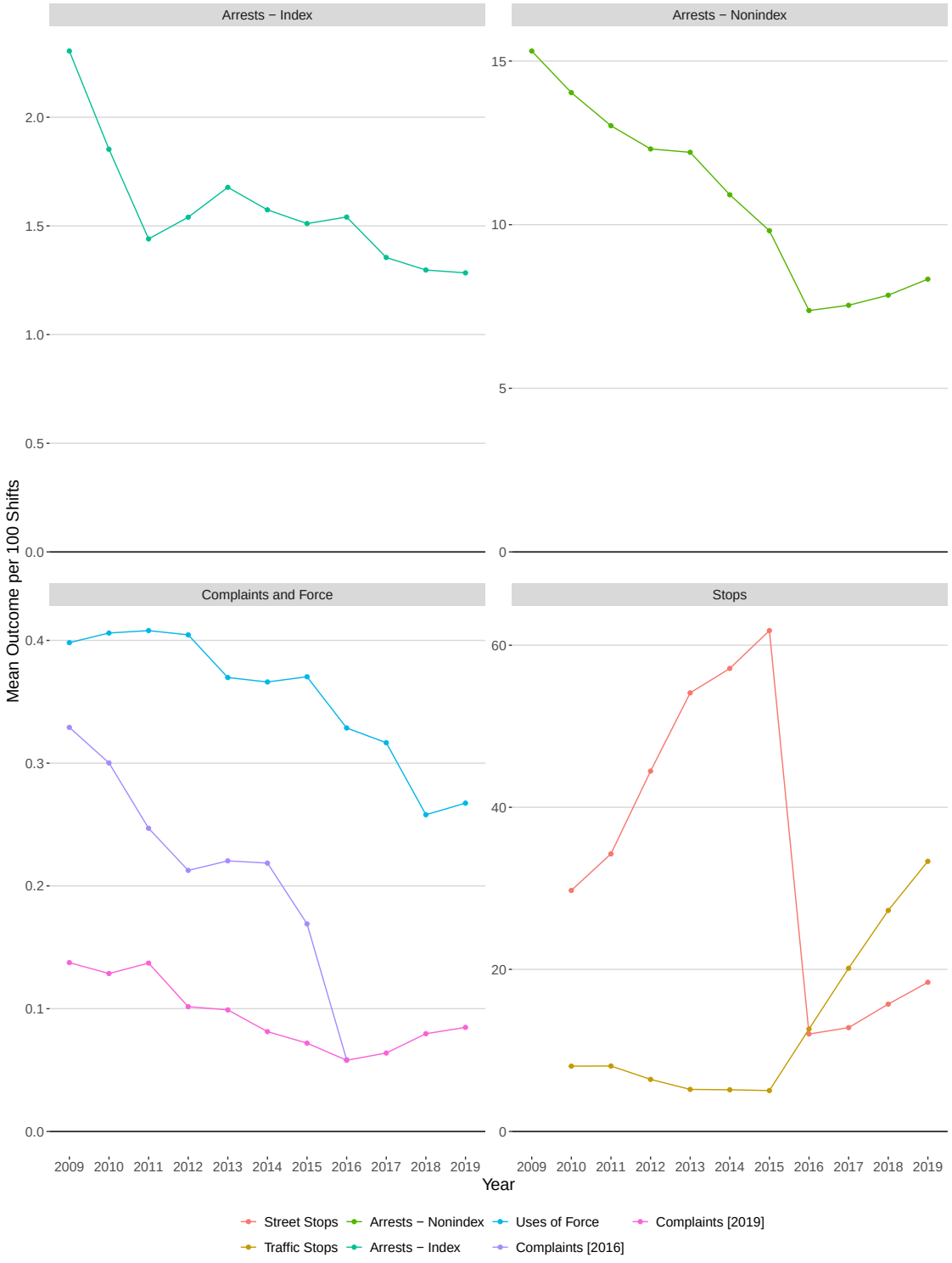
Reported crime is primary outcome of interest which cannot be directly assigned to an individual officer but can be linked to a specific place and time. Geo-located and time-stamped crime data were obtained from Chicago's public data portal. This data contains information on reported crimes in Chicago including the district, coordinates, date and time, whether an arrest was made, crime type, and other information.[31]

Table 3 displays the average crimes per district-hour broken into cleared (an arrest was made) and uncleared (no arrest was made) crimes with corresponding clearance rates. Drug crimes, with around 0.15 occurring every hour in the average district, are almost all cleared, consistent with them generally being 'victimless' crimes that civilians have little incentive to report, so reports of drug crime come almost all from arrests made at the discretion of officers. Weapon offenses, which occur at a rate of 0.02 per district-hour, are also generally cleared by arrests at 77%, for similar reasons as with drug crime though they are more likely to be reported by civilians than drug crimes. Clearance rates are significantly lower for other crime types. For example, simple violent offenses (nonindex), which occur at around 0.31 per district-hour and consist of simple assault and battery, often due to domestic violence, are only cleared 22% of the time, with a similar rate of other nonindex crimes (occurring at 0.43 per district-hour). Index crimes are less likely to be cleared with arrests. Only 17% of violent index crimes (murder, robbery, rape, and aggravated assault) are cleared with arrests and

[31]Reflecting broader trends in the U.S., the environment around crime and policing in Chicago changed significantly between 2009 and 2019. Figure E.3 displays the trends for reported crimes in Chicago over the time period broken into index and nonindex crimes and corresponding clearance rates. Crime across categories declined significantly between 2009 and 2012, while clearance rates remained relatively stable. From 2013 to 2015, crimes also declined but more modestly, while clearance rates increased by a small amount for index crimes. Following 2015, nonindex crimes continued to decline and plateau and index property crimes saw a modest increase but index violent crimes increased significantly into 2017 and declined after; clearance rates during this crime increase declined across crime types and only increased slightly following 2017.

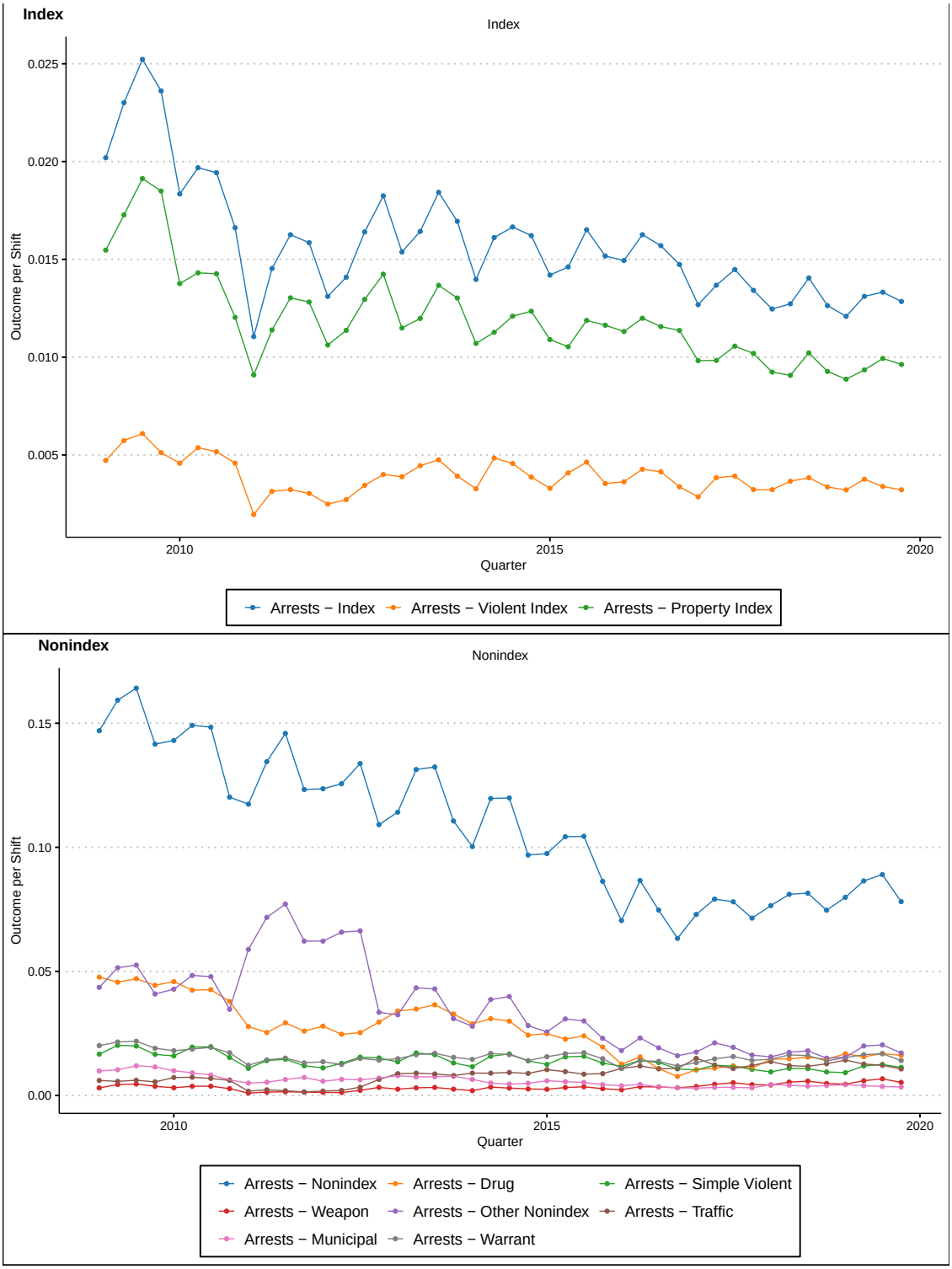
they occur at a rate of 0.14 per district-hour. Index property crimes (motor vehicle theft, theft, burglary, arson) are rarely cleared at only 9% but are fairly common at almost 0.5 per district-hour.

Figure E.1: Average Officer-Level Outcomes per 100 Shifts over Time



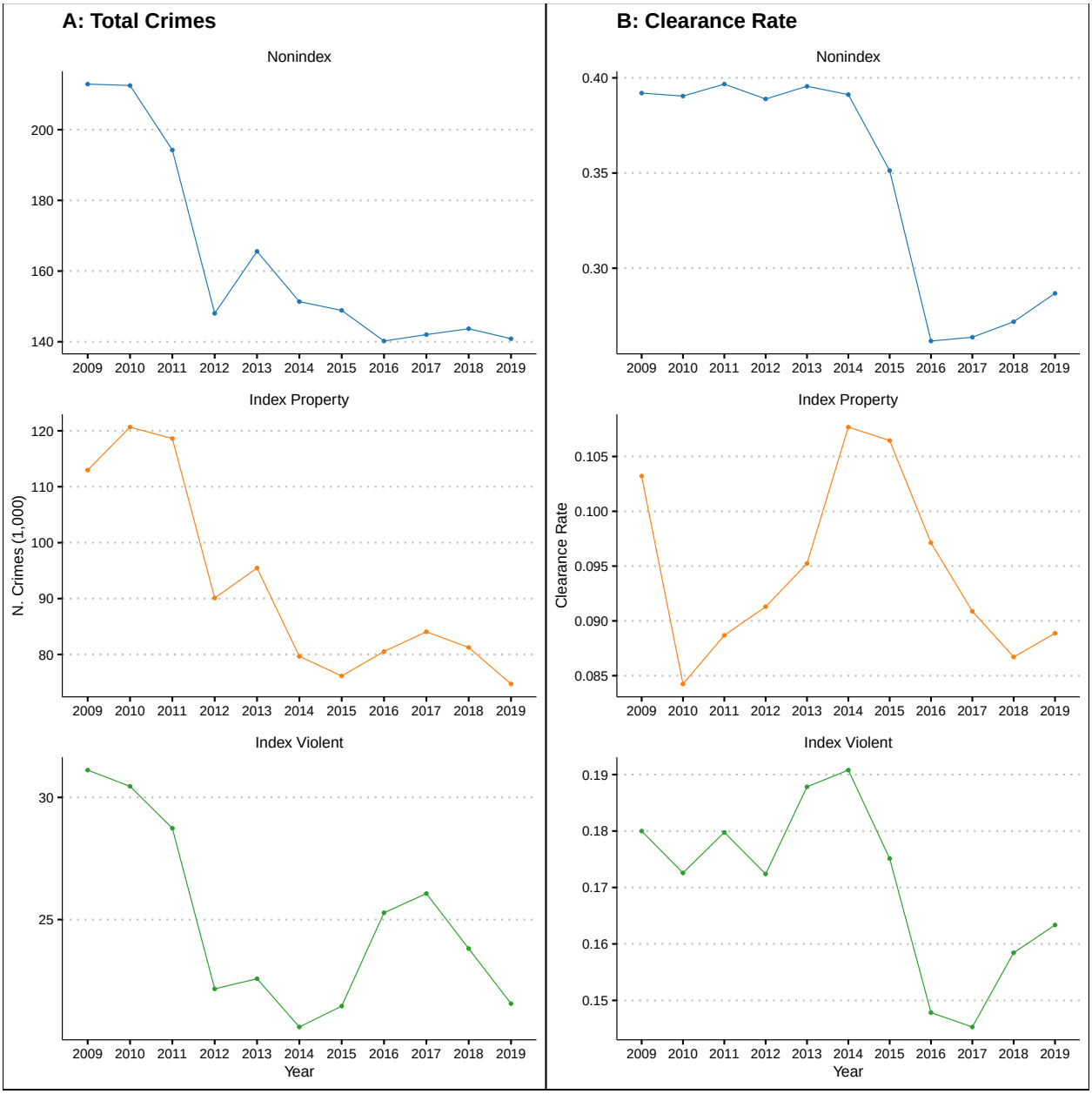
Note: Figure displays the average officer-level outcomes per 100 shifts by year. Street refers to non-traffic contact card (2010 - 2015) and ISR (2016-2019), and complaints [2016] end in 2016 while complaints [2019] end after 2019.

Figure E.2: Timeseries of Arrests by Subcategory



Note: Figure displays the timeseries of mean arrests per 100 shifts of the officer-day data with panels for index arrests and nonindex arrests broken into subtypes at the year-quarter level.

Figure E.3: Timeseries of Crime and Clearances



Note: Figure displays the total number of crimes in 1,000 reported in Chicago between 2009 and 2019 by year broken into subtypes denoted by color (index violent, index property, and nonindex) in Panel A and the clearance rate (share of crimes was reported to have been cleared by an arrest) in Panel B.

Table E.1: Means and Standard Deviations of All Outcome per 100 Shifts by Sample Period

	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Stops	-	46.31 (127.14)	55.19 (143.12)	68.98 (165.1)	79.68 (178.85)	77.87 (176.74)	43.88 (132.45)	12.36 (55.49)	14.22 (59.25)	17.11 (66.03)
Street Stops	-	31.91 (102.48)	39.21 (117)	48.91 (134.11)	55.57 (144.73)	59.19 (151.28)	37.66 (120.94)	12.36 (55.49)	14.22 (59.25)	17.11 (66.03)
Stops - Traffic	-	14.4 (65.59)	15.98 (71.03)	20.08 (83.03)	24.11 (91.81)	18.67 (81.43)	6.22 (46.72)	0 (0)	0 (0)	0 (0)
Traffic Stops	-	8.04 (43.5)	7.14 (41.51)	5.84 (38.76)	5.18 (37.17)	5.1 (37.31)	8.8 (45.4)	16.27 (60.33)	23.58 (75.33)	30.65 (88.4)
Traffic Stops - Search	-	0.53 (7.89)	0.42 (6.88)	0.29 (5.7)	0.24 (5.06)	0.27 (5.4)	0.26 (5.29)	0.23 (4.98)	0.22 (5.04)	0.23 (5.07)
Traffic Stops - Ticket	-	4.65 (33)	4.27 (32.65)	3.95 (32.89)	3.94 (33.09)	3.65 (32.58)	3.38 (29.25)	3.44 (26.89)	3.22 (25.87)	3.11 (25.74)
Stop - Traffic (Primary)	-	11.52 (56.08)	11.95 (58.11)	13.71 (65.38)	15.59 (70.31)	12.45 (62.92)	4.51 (37.32)	0 (0)	0 (0)	0 (0)
Stop - Street (Primary)	-	17.4 (70.02)	21.14 (80.02)	26.21 (92.17)	29.73 (99.4)	31.69 (103.95)	20.33 (82.63)	6.94 (38.76)	7.81 (40.72)	9.29 (44.52)
Arrests - Nonindex	14.6 (46.67)	13.46 (44.31)	12.62 (42.38)	12.2 (41.44)	11.42 (39.63)	10.21 (36.66)	8.6 (32.82)	7.42 (29.63)	7.69 (29.82)	8.05 (30.57)
Arrests - Drug	4.38 (23.7)	3.43 (20.53)	2.68 (17.97)	3.02 (19.34)	3.07 (19.56)	2.51 (17.59)	1.71 (14.49)	1.11 (11.61)	1.29 (12.41)	1.54 (13.66)
Arrests - Simple Violent	1.8 (14.26)	1.53 (13.15)	1.33 (12.19)	1.44 (12.67)	1.49 (12.89)	1.45 (12.65)	1.35 (12.16)	1.2 (11.35)	1.08 (10.7)	1.09 (10.74)
Arrests - Weapon	0.36 (6.31)	0.23 (4.99)	0.16 (4.14)	0.24 (5.08)	0.27 (5.49)	0.28 (5.54)	0.3 (5.82)	0.37 (6.41)	0.47 (7.15)	0.53 (7.53)
Arrests - Other Nonindex	4.52 (29.44)	5.52 (30.43)	6.21 (30.92)	4.73 (27.8)	3.52 (24.32)	2.99 (21.66)	2.33 (18.39)	1.88 (15.83)	1.76 (15.01)	1.74 (14.78)
Arrests - Traffic	0.63 (8.93)	0.43 (8.04)	0.26 (6.16)	0.59 (8.33)	0.88 (9.92)	0.92 (10.02)	1.03 (10.48)	1.18 (11.14)	1.26 (11.35)	1.23 (11.24)
Arrests - Municipal	0.96 (12.04)	0.71 (10.34)	0.62 (9.59)	0.7 (9.9)	0.64 (9.45)	0.51 (7.97)	0.44 (7.1)	0.34 (6.16)	0.35 (6.37)	0.39 (6.63)
Arrests - Warrant	1.94 (14.34)	1.6 (13.06)	1.37 (12.02)	1.48 (12.47)	1.55 (12.74)	1.55 (12.7)	1.43 (12.2)	1.34 (11.78)	1.47 (12.38)	1.54 (12.64)
Arrests - Index	2.09 (16.17)	1.65 (14.19)	1.5 (13.58)	1.61 (14.09)	1.63 (14.26)	1.54 (13.91)	1.53 (13.85)	1.46 (13.43)	1.33 (12.71)	1.3 (12.47)
Arrests - Violent Index	0.52 (7.94)	0.39 (6.84)	0.3 (5.99)	0.37 (6.78)	0.42 (7.22)	0.4 (6.92)	0.39 (6.78)	0.37 (6.5)	0.35 (6.26)	0.35 (6.22)
Arrests - Property Index	1.57 (14.09)	1.26 (12.44)	1.2 (12.18)	1.24 (12.35)	1.21 (12.29)	1.14 (12.05)	1.14 (12.07)	1.09 (11.75)	0.98 (11.07)	0.95 (10.81)
Uses of Force	0.4 (6.55)	0.4 (6.56)	0.4 (6.53)	0.39 (6.37)	0.37 (6.19)	0.37 (6.21)	0.35 (6.04)	0.32 (5.78)	0.29 (5.47)	0.26 (5.27)
Uses of Force - Injuries	0.1 (3.13)	0.1 (3.21)	0.1 (3.2)	0.1 (3.11)	0.09 (3.04)	0.09 (3.05)	0.08 (2.8)	0.05 (2.2)	0.03 (1.77)	0.03 (1.86)
Complaints [2016]	0.31 (5.63)	0.27 (5.25)	0.23 (4.8)	0.22 (4.68)	0.22 (4.7)	0.19 (4.42)	-	-	-	-
Complaints [2016] - Serious	0.21 (4.62)	0.18 (4.3)	0.15 (3.9)	0.14 (3.82)	0.14 (3.8)	0.11 (3.29)	-	-	-	-
Complaints [2016] - Failure to Provide Service	0.06 (2.39)	0.05 (2.28)	0.05 (2.16)	0.04 (2.1)	0.04 (2.03)	0.04 (2.08)	-	-	-	-
Complaints [2016] - OPV/Other	0.05 (2.13)	0.04 (1.95)	0.03 (1.77)	0.03 (1.69)	0.03 (1.84)	0.04 (2.06)	-	-	-	-
Complaints [2016] - Sustained	0.01 (0.99)	0.01 (0.9)	0.01 (0.89)	0.01 (0.93)	0.01 (0.94)	0.01 (0.76)	-	-	-	-
Complaints [2019]	0.13 (3.64)	0.13 (3.64)	0.12 (3.45)	0.1 (3.18)	0.09 (3.01)	0.08 (2.8)	0.07 (2.6)	0.06 (2.49)	0.07 (2.71)	0.08 (2.89)
Complaints [2019] - Serious	0.1 (3.19)	0.1 (3.15)	0.09 (2.95)	0.07 (2.65)	0.06 (2.42)	0.05 (2.16)	0.03 (1.86)	0.03 (1.77)	0.04 (1.9)	0.04 (2.02)
Complaints [2019] - FPS/Conduct	0.01 (1)	0.01 (0.96)	0.01 (0.89)	0.01 (0.95)	0.01 (1.04)	0.01 (0.98)	0.01 (1.02)	0.01 (1.13)	0.02 (1.33)	0.02 (1.47)
Complaints [2019] - OPV	0.01 (0.81)	0.01 (0.79)	0.01 (0.79)	0.01 (0.74)	0.01 (0.73)	0.01 (0.74)	0.01 (0.79)	0.01 (0.85)	0.01 (0.91)	0.01 (0.94)
Complaints [2019] - Other	0.02 (1.42)	0.02 (1.55)	0.02 (1.56)	0.02 (1.46)	0.02 (1.44)	0.02 (1.48)	0.02 (1.47)	0.02 (1.29)	0.02 (1.36)	0.02 (1.42)
Complaints [2019] - Sustained	0.01 (0.89)	0.01 (0.86)	0.01 (0.94)	0.01 (0.97)	0.01 (1)	0.01 (1.01)	0.01 (1.04)	0.01 (1.1)	0.02 (1.32)	0.02 (1.45)
Arrests - Index (Primary)	1.08 (11.55)	0.85 (10.11)	0.76 (9.64)	0.82 (9.98)	0.83 (10.05)	0.78 (9.78)	0.78 (9.73)	0.74 (9.46)	0.69 (9)	0.68 (8.83)
Arrests - Nonindex (Primary)	7.44 (32.96)	6.82 (31.16)	6.39 (29.84)	6.18 (29.18)	5.78 (27.86)	5.17 (25.75)	4.38 (23.09)	3.78 (20.94)	3.93 (21.06)	4.14 (21.52)
Arrests - Index (Supp.)	1.61 (15.67)	1.22 (13.44)	1.13 (12.97)	1.43 (14.62)	1.66 (15.83)	1.77 (16.34)	1.97 (17.27)	2 (17.24)	2 (17.03)	2.05 (17.13)
Arrests - Nonindex (Supp.)	9.81 (46.21)	9.45 (45.37)	9.23 (43.64)	9.61 (43.73)	9.97 (43.49)	9.78 (41.72)	9.01 (39.47)	8.54 (36.87)	9.47 (37.47)	10.41 (39.01)

Note: Table presents the means and standard deviations of outcomes across each rolling 2-year sample period with each sample filtered to be the analysis sample, meaning the observations belong to the leave-one-out largest connected component. Primary refers to the officer being the first officer, while Supp. refers to them being supplementary or auxilary. By default, the outcomes only count for primary and secondary officers (non-supplementary) when relevant.