Roman Rivera

rgrivera@berkeley.edu

UC Berkeley

PhD in Economics, Columbia University, April 2023

https://www.romangrivera.com/

My current research activity focuses on the criminal justice system in the United States. I study the police and crime in a broad set of environments, such as the influence of BLM protests on the financial aspects of the policing industry, the impact of race, gender, and peers on officer behavior, the importance of officer heterogeneity on officer activity, misconduct, and their ability to reduce crime, and how electronic monitoring influences defendant outcomes, among other questions. My work on policing and crime leverages tools and frameworks from labor economics, personnel economics, and applied econometrics to explore how law enforcement and related institutions work.

As a Hispanic American, I belong to an underrepresented group in economics, and I am passionate about ensuring that other underrepresented people feel that they belong in the economics community. I mentor underrepresented students and participate in diversity and inclusion initiatives. As a PhD student, I was the co-president of Columbia Economics's Diversity Initiative.

My research also contributes to the intellectual diversity of the economics profession. I study the criminal justice system and policing, which are not only rife with issues of racial disparities in themselves but also contribute to downstream disparities due to their influence on other socio-economic outcomes, and my research seeks to reduce such disparities by improving the quality of the criminal justice system and its agents. My methods are also distinct; my introduction to research was working with a journalism organization in the Southside of Chicago in 2015 to study and document police misconduct. The data for most of my work has come from Freedom of Information Act requests and collaboration with journalists and nonprofits, and much of my institutional knowledge and questions of interest stem from conversations and experiences with institutional outsiders. These experiences have allowed me to explore novel topics, such as whether there are racial disparities in how white supervisors treat Black officers. This novel research has given me a different perspective on the institutions and actors studied in the economics of crime.

The sessions in the final week of the NBER Summer Institute cover all of my interests, as my work intersects with four of the sessions: Crime, Labor, Law and Economics, and Personnel Economics.

Roman Gabriel Rivera

rgrivera@berkeley.edu

romangrivera.com | Google Scholar

Employment

2025- Assistant Professor, McCourt School of Public Policy, Georgetown University

2023-2025 Affiliate Researcher, Georgetown University

2024-2025 IRS Postdoctoral Research Associate, Princeton University

2023-2024 Postdoctoral Scholar, IRLE UC Berkeley

Education

2023 Ph.D. in Economics, Columbia University

2017 Master of Public Policy (Honors), The University of Chicago

2016 B.A. in Economics, The University of Chicago

Publications

"The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago", with Bocar Ba
[Accepted at *The Review of Economics and Statistics*]

"The Black-White Recognition Gap in Award Nominations", with Nayoung Rim, Andrea Kiss, and Bocar Ba
[Forthcoming at *Journal of Labor Economics*]

"Peer Effects in Police Use of Force", with Justin Holz and Bocar Ba
American Economic Journal: Economic Policy, May 2023

"The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago", with Bocar Ba, Dean Knox, and Jonathan Mummolo
Science, February 2021

"Disparities in Police Award Nominations: Evidence from Chicago", with Nayoung Rim and Bocar Ba
AEA Papers and Proceedings, May 2020

Working Papers

"Release, Detain, or Surveil? The Effect of Electronic Monitoring on Defendant Outcomes"

"The Effect of Minority Peers on Future Arrest Quantity and Quality"
[Revise and Resubmit at *American Economic Journal: Applied Economics*]

"Who are the Police? Descriptive Representation in the Coercive Arm of Government", with Bocar Ba, Jacob Kaplan, Dean Knox, Mayya Komisarchik, Rachel Mariman, Jonathan Mummolo, and Michele Torres
[Revise and Resubmit at *American Journal of Political Science*]

"Market Response to Racial Uprisings", with Bocar Ba and Alexander Whitefield
[NBER Working Paper No. 31606]

"Forecasting the Impact of Racial Uprisings, Market versus Stakeholders' Expectations", with Bocar Ba and Alexander Whitefield
[NBER Working Paper No. 31857]

"Police Officer Assignment and Neighborhood Crime", with Bocar Ba, Patrick Bayer, Nayoung Rim, and Modibo Sidibe
[NBER Working Paper No. 29243]

Honors, Fellowships, Awards, and Grants

2024	Efficacy and Effects of Pretrial Electronic Monitoring (The Laura and John Arnold Foundation)
2024	Officer Heterogeneity and the Incentive Problem of Policing (Institute for Humane Studies, IHS017836)
2023	Padma Desai Award for Best Dissertation (Columbia University, GSAS)
2022	Ford Foundation Dissertation Fellowship (The National Academies of Sciences, Engineering, and Medicine)
2022	Program for Economic Research Graduate Student Summer Fellowship (Columbia University)
2022	Graduate Student Workshop (WEAI Annual Conference)
2021	Lewis A. Sanders Endowed Fellowship in Economics (Columbia University, GSAS)
2021	Wueller 4th Year Pre-Dissertation Award, Winner (Columbia University, Economics Department)
2021	Best Student Paper (Urban Economics Association 2021)
2021	Program for Economic Research Graduate Student Summer Fellowship (Columbia University)
2020	Harriss 2nd Year Paper Award, Winner (Columbia University, Economics Department)
2020	Russell Sage Foundation Pipeline Diversity Grant (PIs: Bocar Ba and Nayoung Rim)
2018	Provost Diversity Fellowship (Columbia University, GSAS)

Conferences and Presentations

2024	ASSA, UC Berkeley, Goldman School of Public Policy, University of British Columbia, Texas Economics of Crime Workshop, NLS-E
2023	ASSA, Georgetown McCourt, NYU, UT Austin, UC Berkeley, LSE, U of I Geis, Notre Dame, Duke, UC San Diego, Wash U, UPenn Quattrone, SOLE, WEAI, NBER SI Crime, ViCE, APPAM
2022	Texas Economics of Crime Workshop, CELS, SOLE, WEAI, SDSU, APPAM, CEP/LSE Crime Seminar
2021	CIVICA, UEA (North America), HBS Rising Scholars, SEA
2020	APPAM, SEA

Other Employment Positions

2020-2021 W. Bentley MacLeod, Research Assistant, Columbia University

2020-2021 Diversity Initiative, Co-President, Columbia University

2016-2018 Robert J. LaLonde, Research Assistant, University of Chicago

2017-2018 Amy Claessens, Research Assistant, University of Chicago

2017-2018 Invisible Institute, Data Scientist

Referee Service

Quarterly Journal of Economics, Review of Economics and Statistics, American Economic Review, American Economic Journal: Applied Economics, American Economic Journal: Economic Policy, Journal of Labor Economics, Journal of Public Economics, Journal of Policy Analysis and Management, Journal of Economic Behavior and Organization, American Law and Economics Review, Industrial and Labor Relations Review, Criminology, Journal of Urban Affairs, Journal of Quantitative Criminology

Officer Heterogeneity

Roman Rivera*

UC Berkeley

March 8, 2024

[Preliminary. Do not cite or redistribute]

Abstract

To what extent do individual officers matter in preventing crime, making arrests, committing misconduct, and using force? And are there tradeoffs in the selection of officers due to correlations between costly actions, such as force and misconduct, and productive enforcement and crime prevention? This paper studies the degree of officer heterogeneity across enforcement, misconduct, and crime outcomes. I use extensive administrative data from the Chicago Police Department over a decade with granular information on officers' shifts, assignments, and activities in combination with granular crime reports to document a large number of facts about police officer heterogeneity, sorting between assignments, correlations in officer types across outcomes, and the relationship between demographics (e.g., race, gender, tenure, education, military experience) and officers' individual behaviors. Among other results, I find that officer heterogeneity is economically significant across most outcomes, officers are more critical than assignments/work environments in determining the use of force and misconduct, and officers tend to be heterogeneous in their behaviors (e.g., officers who make more serious arrests are not also high misconduct officers). Additionally, officer heterogeneity in influencing crime is comparable to the importance of officer quantity in prior work; however, officer heterogeneity matters more for property crime than for violent crime, with patterns most consistent with officers being differentially effective at deterring crime.

*For feedback and advice, I thank Bocar Ba, Patrick Kline, Raffaele Saggio, and the participants at UC Berkeley Labor Seminar and APPAM 2023. This work received support from the Institute for Humane Studies under grant number IHS017836.
Email: rgrivera@berkeley.edu

1 Introduction

Individual police officers can incur social costs through unnecessary contacts and misconduct or increase social benefits by reducing crime. Demands for better policing across the United States have led to research and policy focusing on influencing the quality of policing and minimizing of policing costs, such as police misconduct, use of force, and unnecessary stops or arrests, through recruitment, training, or oversight. Nevertheless, we know relatively little about the degrees to which individual officers vary, either focusing on group specific differences or on heterogeneity for a single enforcement outcome.[1] Furthermore, we know very little about how the importance of officer quality in relation to crime rates, even though the influence of the quantity of officers on crime is well established (Chalfin and McCrary (2017)).[2] This paper seeks to bridge these gaps by quantifying the influence of individual police officer on enforcement, force, misconduct, and crime reduction.

Quantifying the influence of an individual officer on enforcement, force, misconduct, and crime is empirically difficult as it requires extensive and granular data on police officers, their outcomes, and their assignments. I overcome this by leveraging 11 years of shift-level data on over 10,000 Chicago police officers, totaling over 14 million shifts. I exploit officers' operations calendar and fixed rotations, which makes the endogenous selection of when an officer works untenable and granular information on officers' assignments on each shift (e.g., location, role, time) to separately identify officer-specific factors and environmental/assignment-specific factors, which contribute to officers' enforcement activity.

I begin by focusing on officer-level enforcement outcomes — stops, nonindex arrests (more discretionary and lower-level), index arrests (less discretionary and more serious), uses of force, and complaints — and estimating the effect of individual officers and specific

[1] See Donohue and Levitt (2001), McCrary (2007), Miller and Segal (2018), Ba, Knox, et al. (2021), Goncalves and Mello (2021), Hoekstra and Sloan (2022) for demographic differences, and see Rozema and Schanzenbach (2019), Weisburst (2022), Adger, Ross, and Sloan (2022), and Gonçalves and Mello (2023) for heterogeneity.

[2] See McCrary (2002), Levitt (2002), Evans and Owens (2007), Weisburst (2019), Chalfin and McCrary (2018), Mello (2019), and Chalfin et al. (2022).

assignments on these outcomes. I adapt the Abowd, Kramarz, and Margolis (1999) (AKM)-style framework used widely to study worker and firm effects on economic outcomes (Card, Heining, and Kline (2013), Bonhomme et al. (2022), Lachowska et al. (2023b)), to modeling officer and assignment effects. These effects can be decomposed into the relative influences of observables, and their variances and covariances inform us on the magnitudes and relative importance of officer and assignment heterogeneity, the degree of sorting between officers into assignments, and how officers' behavioral effects covary across outcomes. I apply the Kline, Saggio, and Sølvsten (2020) (KSS) bias correction to each variance component and cross-outcome correlations following Lachowska et al. (2023b). With this setup, I answer a range of empirical questions related to police officer heterogeneity, with six main sets of results.

First, I find that differences across individual officers (officer heterogeneity) tend to explain more of the variation in outcomes than differences across assignments, with the exception of less discretionary arrests for index crimes. Importantly, although officers and assignments explain little of the overall variation in the use of force and complaints ($< 1\%$), officer effects explain over 50% more than assignment effects. This suggests that shifting officer behavior is highly important for rare and highly costly outcomes.

Second, across all officer-level outcomes, between-officer variation is economically significant. Compared with the overall means, a 1 standard deviation (SD) increase in officer effects each respective outcome increases stops by 41 (90% of the mean), 10 nonindex arrests (90%), 1.6 index arrests (95%), 0.45 uses of force (130%), and 0.29 complaints (120%) per 100 shifts. Compared with existing estimates, I find a much more significant influence of officer heterogeneity on arrests than prior work (Weisburst (2022)), but these estimates are similar to those on use of force in Adger, Ross, and Sloan (2022) and complaints in Rozema and Schanzenbach (2019).

Third, I find little evidence of officers sorting into assignments based on their effects on most outcomes, other than nonindex arrests. Effectively, officers who are receive more

complaints and use more force regardless of where they work tend to avoid sorting into assignments where such outcomes are more common. However, officers who make more nonindex arrests sort into assignments that allow for more nonindex arrests.

Fourth, officers' propensities for different outcomes are not strongly correlated. Two of the strongest correlations are between an officer's propensity to use force and their propensity to receive complaints, an officer's propensity to make stops, and their propensity to make nonindex arrests, which are both around 0.36. This suggests that rather than officers being highly active in all dimensions or lazy, they are relatively heterogeneous and multidimensional in their behavior.

Interestingly, officers' propensity to receive complaints is very weakly correlated to their propensity to make index arrests (around 0.05) and weakly negatively correlated with their propensity to make stops. This is evidence of police misconduct having little to do with effective or high-performance policing. This contradicts assumptions that officer activity in any sense necessitates misconduct, as is assumed by many models of police misconduct (Shi (2009), Premkumar (2022)), but is consistent with the evidence that oversight can be implemented without hampering arrests (Rivera and Ba (2023), Devi and Fryer (2020)). I also find that officers' propensity for 'serious' misconduct (e.g., physical abuse) is uncorrelated with their propensity for failing to provide services, consistent with work delineating between victim-complainants and suspect-complainants (Ba (2018), Prendergast (2021)).

Fifth, I find that the differences between groups (e.g., the average difference between Black and white officers' propensity to make nonindex arrests) are much smaller than the amount of officer heterogeneity within groups. Additionally, within-group heterogeneity differs across demographic groups. For example, Black officers, female officers, and more experienced officers make fewer nonindex arrests and use less force than their white, male, and less experienced peers, but they also display less heterogeneity (smaller demographic-specific variances).

Sixth, an officer's observable characteristics are strongly associated with differences in

policing behavior across different outcomes. I decompose officers' effects for stops, arrests, use of force, and complaints on the impact of race, gender, starting age, tenure, language ability (speaking Spanish), military experience, and educational attainment (lower than a BA or at least a BA). While prior literature has studied the relationship between subsets of these outcomes and subsets of these characteristics, I provide a more encompassing set of results. I confirm some previous results, such as minority, female, and more experienced officers making fewer stops, arrests, and using less force (Ba, Knox, et al. (2021), Goncalves and Mello (2021), Hoekstra and Sloan (2022), Ba, Bayer, et al. (2021)), but I also find demographic relationships with misconduct. Notably, the differences between higher education and lower education officers are economically small, particularly in comparison with other demographic factors, on almost all outcomes, suggesting that higher education alone may not yield better policing.

Interestingly, officers with military experience display a unique pattern in their officer effects. While they make significantly more arrests, stops, and uses of force, they do not receive more complaints for either serious allegations (e.g., excessive force) or failure to provide service/poor conduct. This is unusual as most demographics with strong positive or negative relationships with force and arrests also have a similar relationship with complaints. However, it suggests that military training or experience may be important for increasing officer activity but not misconduct.

In the second portion of the paper, I focus on how officers vary concerning their ability to prevent (or permit) crime rates in pursuit of understanding how officer 'quality' matters in crime reduction beyond the well-studied effects of officer quantity in economics (Chalfin and McCrary (2017)). Beyond traditional limitations for studying officer-level effects, recovering officer effects on crime is empirically challenging because a crime cannot be ascribed to a specific officer, only a specific geographic area, and so it can be the product of an entire team of officers' collective efforts (or lack thereof). To overcome this challenge, I use the shift-level data containing start and end times to construct the 'team' of officers working in a given

district at a given time, at the district-hour level. Then, I collapse outcomes, such as crime reports and arrests, to this level. I rely on the operations calendar to induce variation in the composition of officers working in a given district at a given hour, which is independent of shocks to other determinants of crime and conditional high-dimensional district-time fixed effects.

Given that teams (number of officers in a district-hour) of officers range in size from less than 20 to over 90, I borrow Bonhomme (2021)'s additive team production model (and associated pruning method) and model team effects as an average over the individual officer effects of the officers working in that team, resulting in thousands of parameters. I diverge from Bonhomme (2021) by employing an additive model of officer effects in which team effects are the mean of the participating officer effects without a team-size scaling factor.

Heterogeneity in team composition (i.e., which officers are working) has varying importance on crime rates depending on the crime type. Additionally, without bias correcting the variance (and covariances) of team effects using KSS, the importance of officer heterogeneity on crime rates would be significantly overstated. For violent crimes, heterogeneity across teams has a relatively small influence and tends to be primarily due to measurement error overstating the variance components. For property crime, however, variance in team effects explains around 13% of the variation in crime rates, primarily driven by their influence over larceny/theft, with a 1 SD increase in a team's effect on theft increasing theft by around 71% of the mean. Unsurprisingly, team heterogeneity is most impactful and explains the most variation on drug crimes (around 15%), which are almost entirely drug arrests, with a 1 SD increase in a team's effect increasing drug crimes reported by 140% of the mean. For other nonindex crimes, such as simple violent crime (e.g., domestic violence and weapon offenses), team effects explain a relatively small share of overall variation.

I probe the potential mechanisms explaining these effects, and attempt to differentiate heterogeneity in deterrence, incapacitation, and reporting/suppression of crimes. I find most evidence is inconsistent with incapacitation and with officers differentially reporting or

suppressing crime reports, with the exception of drug crimes (which are entirely arrests). For example, I find that variation in team effects on crime is largely determined by variation in uncleared crimes rather than arrests, consistent with low clearance rates for most crime types, and inconsistent with incapacitation. Additionally, team effects for arrests are generally positively correlated with uncleared crimes, which is not expected if higher incapacitation teams also reduced uncleared crimes. Finally, I find that team effects on crime reports are strongly correlated with related 911 call types, which is inconsistent with suppression and consistent with differential deterrence. Notably, team effects are (weakly) positively correlated across crime types, meaning teams which reduce one form of crime also reduce other forms, but not through higher arrests. Overall, this suggests that team effects influence non-drug crime rates through deterrence, consistent with most work on how the quantity of police influence crime (Chalfin and McCrary (2017), Chalfin et al. (2022)).

I compare these effects of officer 'quality' (team effects) on crime to existing estimates of the effect of officer quantity on crime, which have been widely studied at the department (McCrary (2002), Levitt (2002), Evans and Owens (2007), Weisburst (2019), Chalfin and McCrary (2018), Mello (2019), Chalfin et al. (2022)) or area level (Di Tella and Schargrodsky (2004), Klick and Tabarrok (2005), Draca, Machin, and Witt (2011), Donohue, Ho, and Leahy (2013)). Specifically, I compute the effect of a 1 SD decrease team effects (reducing crime) scaled by the average team size and mean of the crime rate, which approximates the effect of decreasing a single average officer's crime effect by 1 SD, and I compare this with the percent decrease in crime predicted by adding 1 additional average officer to a team (approximately a 1.7% increase) combined with elasticity estimates from Chalfin and McCrary (2018) and Mello (2019). I find that the quantity effects tend to be larger than quality effects for violent crimes, with the exception of aggravated assault where they are similar. For property crime, quality effects are relatively large or similar sized to quantity effects. This suggests that the quality of policing is similarly important to the quantity of policing, consistent with the importance of monitoring and policing strategies (Cheng and Long (2018)).

Finally, I estimate the relationship between crime and the demographic and behavioral (e.g., propensity to make arrests or stops) composition of officers on a team. I use multiple empirical strategies, including instrumenting for the composition of teams using the predicted composition of a team based on calendar rotations. Unlike with officer-level outcomes, I find relatively few robust relationships between officers' measurable characteristics and their influence on crime. For property crime, I find a suggestive evidence that more 'active' officers, who are less experienced and make more stops (holding other behaviors fixed), reduce property crime.

This paper proceeds as follows. Section 2 discusses the institutional background and data. Section 3 presents the empirical strategy and results for the officer-level outcomes, and Section 4 presents the empirical strategy and results for the crime outcomes. Section 5 concludes.

2 Background and Data

2.1 Background

The empirical analysis will focus 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. The study focuses on the period between 2009 and 2019.

The CPD's over 10,000 sworn officers are organized into units that perform different functions, ranging from general policing of a specific area to detective units, SWAT units, and training units. Officers rarely work in units other than their main assigned unit within a given year. The majority of officers work in patrol units, which correspond directly to large areas of the city known as "districts"—similar to precincts in other cities.

During this study, Chicago is divided into 22 (2013-2019) or 25 (prior 2009-2012) police districts. Districts are subdivided into 3 to 5 sectors which are subdivided into geographic

beats (usually 4, each around 1 square mile). The most common units are those that map to a specific police district, and officers are generally responsible for patrolling, answering 911 calls, and serving on special and tactical teams within the district. 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).

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. Changes in units are uncommon, 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). An officer's unit assignment is coupled with their regular watch (which generally determines their working hours) and their day off group (DOG). DOGs determine the cycling of present for duty officers working in a unit during a specific watch over the year, with many officers operating on a 4-2 rotation (4 days on duty followed by 2 days off). 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. For example, DOGs A, B, and C may be active today while tomorrow DOGs B, C, and D will be active, and DOGs C, D, and A will be active the following day. See A.1 for an example of the CPD's operations calendar.

In conjunction with an officer's stable unit assignment, they are also assigned to a specific watch within the unit. Watches 1-3 denote the time an officer works their regular 9-hour shift (e.g., night, morning, or day). Watch 4 denotes the officer is working as a tactical team officer in a specific district. I subdivide Watch 4 into two sub-watches by the time when the officer begins working (morning or night). All other irregular watches, which account for less than 10% of all observations, are grouped together.

Officers working on a given day work a specific assignment, also known as a 'beat' or a beat code. 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. For example, an officer may be a 'beat' officer, meaning they patrol a specific geographic beat; or they may be a rapid response officer, responding to high-priority 911 calls within a specific sector; or they may be a tactical team officer in a particular district, or they may be 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).

2.2 Data

2.2.1 Officer-Level 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 indicating 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, or they were in training), their start and end times, their title (e.g., police officer, sergeant, detective), and their assigned beat, watch, and unit. Initially, this 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). These filters result in approximately 14,400,000 observations, on 14,218 unique officers, in 44 units (25 of which are patrol units that correspond to districts).

Using the assignment data, I am able to measure when, where, and in what role an officer is working on a day-to-day basis over the course of multiple years for each officer, which I will use for the officer-level analysis. The data contains almost 20,000 unique beat

codes. I classify them into around 30 categories that are comparable across units — for example, a beat officer in sector 1 of district 11 has a different assigned beat as a beat officer in sector 1 of district 24, but they are both beat officers in sector 1 and thus have the same beat category. To construct assignments, I interact the unit, watch, and beat category which captures where, when, and in what role the officer is working, with around 2,700 unique assignments in the data.

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 1,504.65 shifts. Within a single year, the average officer works 182 shifts, given they work at least 1 day. Most assignments (80.2%) are in geographic (patrol) units, and they work in an average of 17.3 unique assignments, with 44.7% of an officer's shifts being in their most common assignment.

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, demographically, the sample of officers is plurality white (48.1%), with the remainder being Hispanic (25.1%), Black (23%), and Asian/Native American/other race (3.8%). As with many police departments, a majority of officers are male (80.6%). Officers tend to start in their late 20's around 28.9 years old, and the average officer in sample is around 41 years old and with 12.2 years of experience (tenure). Around 22.5% of officers were in the military, and 14.5% speak Spanish. For around 70% of the sample, I have educational history information, and 30.8% of officers have a BA or higher while 37.1% have less than a BA (e.g., associates degree, college credits).

For officer-level out, I link the daily assignment data to officers' arrests, stops, use of force, and misconduct records. This provides almost all possible officer-level activity dimensions considered in policy and academic research. In the primary analysis, I only consider an officer as credited with an arrest if they are a primary (1st or 2nd) arresting officer. I also split

arrests into two main categories: arrests for index (violent: murder, robbery, rape, aggravated assault; property: burglary, motor vehicle theft, larceny/theft, arson) and non-index crimes (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.

Panel C of Table 1 displays average outcomes per 100 shifts in the sample. Stops, which contain both street stops and traffic stops, are by far the most frequent at around 45 per 100 shifts. Next non-index arrests occur at around 11 per 100 shifts, followed by index arrests at 1.6 which are almost 10 times rarer. Use of force and complaints are both relatively infrequent, at 0.35 and 0.24 per 100 shifts. The main complaint data were produced in 2016, so I supplement this data with extended complaint ('Complaints (Ext.)') data obtained more recently which exists for the full sample period, however it is only a subset of the overall complaints included in the pre-2016 sample, at around 0.09 per 100 shifts.

Figure A.2 shows that nonindex arrests and index arrests have declined in both the full and geographic samples, and Figure A.3 displays the trends in arrests over time. Similar to the change in crime, 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 actually risen slightly.

Some outcomes not measured for the entire time period, namely stops (2010 - 2015) and complaints (2009 - 2015), as shown in Figure A.2. While nonindex arrests were declining, stops increased between 2010 and 2015 by around 40%, averaging at 38 per 100 shifts with 10 traffic and 28 street stops, but complaints declined from 0.3 per 100 shifts to 0.1 per 100 shifts between 2009 and 2015, averaging to around 0.16 per 100 shifts. Uses of force average

0.35 per 100 shifts and 0.07 subject injuries per 100 shifts — they also decline over the time period but only from around 0.4 to 0.3.

2.2.2 Crime

Table 2 displays the average crimes per district-hour broken into cleared and uncleared 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 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.

Reflecting broader trends in the U.S., the environment around crime and policing in Chicago changed significantly between 2010 and 2019. Figure A.4 displays the trends for reported crimes in Chicago over the time period broken into index and nonindex crimes which are further subdivided by whether they were cleared by an arrest. Most notably, the amount of crime declined over the time period by about 25%, but this was largely driven by moderate declines in property crime, less than 20% decline and largely driven by nonindex crime declining by around 30%, and violent crime remaining stable.

2.3 Internal Validity

The most direct threat to the identification of the parameters of interest in the following section 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. 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, Jordan, and Kim (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 B.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.

3 Effect of Officers on Enforcement, Force, and Misconduct

In this section, I explore the influence of officers and assignments on officer-level outcomes, the degree to which officers are correlated in their 'types' or 'propensities' across outcomes into assignments, and how strongly observable qualities of the officers are related to their underlying effects.

3.1 Empirical Strategy

I employ an AKM-style (Abowd, Kramarz, and Margolis (1999)) framework to disentangle the influence of officers, assignments, and their sorting patterns on officers' day-to-day outcomes. Given officer i working in assignment J on day t, the outcome $y_{i,t}$ (e.g., number of index arrests made by i) for officer i on date t is the sum of officer-specific effect θ_i, assignment-specific effect $\psi_{J(i,t)}$, time-varying observable characteristics $\beta x_{i,t}$ (e.g., date effects), and an error term $\epsilon_{i,t}$:

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

While this model is generally used to study the relationship between wages and workers and firms, we can easily adapt the model to policing. In this context, θ_i is all of the officer-specific factors that influence the relevant outcome or 'discretion', such as their personal biases and beliefs, friendliness, general aggression, how civilians see the officer, and their intrinsic motivation. Similarly, $\psi_{J(i,t)}$ is all assignment-specific factors which influence outcomes, such as the extrinsic incentive structure, 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. Finally, time varying factors, such as date effects or officer tenure may also influence outcomes due to trends (e.g., officers

may be less active as they get older and crime declined in aggregate over the sample period).

Under this model, identification of $\theta_{1,...,N}, \psi_{1,...,J}, \beta$ requires the following mean independence condition is satisfied:

$$E(\epsilon_{i,t}|x_{1,1}...x_{N,T}, J(1,1)...J(N,T), \theta_1...\theta_N, \psi_1...\psi_J) = 0 \tag{2}$$

that conditional on officers, assignments, and observables, the shocks to outcomes $\epsilon_{i,t}$, 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. Additionally, match-effects or drifting of effects are not allowed. I test for these issues in Appendix B.1 and the results suggest they are not a significant concern.

Prior to estimation, I restrict the sample to the largest connected component of officers and assignments, and further restrict it to the largest leave-one-out connected component (Kline, Saggio, and Sølvsten (2020)) which includes around 95% observations in a given sample, which partially reduces concerns over limited mobility bias.[3]

Given estimates of officer and assignment effects, I apply the classical variance decomposition, given equation (1) and letting $\overline{y}_{i,t} = y_{i,t} - \beta x_{i,t}$:

$$Var(\overline{y}_{i,t}) = Var(\theta_i) + Var(\psi_{J(i,t)}) + 2Cov(\theta_i, \psi_{J(i,t)}) + Var(\epsilon_i) \tag{3}$$

The relative magnitudes of $Var(\theta_i)$ and $Var(\psi_{J(i,t)})$ tell us how much of the variation in outcomes can be explained by the variation in officer types versus the influence of the assignments they work, while $2Cov(\theta_i, \psi_{J(i,t)})$ tells us the influence of sorting. For example, if we focus on complaints against officers as the outcome, then larger values of $Var(\theta_i)$ is

[3]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. Abowd, Creecy, and Kramarz (2002) shows that this condition is satisfied and effects are identified for workers and firms within a "connected set", which depends on the structure of the bipartite graph produced by the connections between firms and workers.

consistent with a 'bad apples' story: individual officer effects explain a larger amount of the variation in complaints, beyond their assignments. However, if $Var(\theta_i)$ is small and $Var(\psi_{J(i,t)})$ is large, then the receipt of complaints is more a function of the assignment type than the officer, and re-assigning low-complaint officers to replace high-complaint officers will have little effect. Importantly, if $Cov(\theta_i, \psi_{J(i,t)})$ is positive, then this indicates that high-misconduct officers tend to sort into assignments with high complaint (and potentially misconduct) opportunities. Overall, this allows us to gauge the degree of heterogeneity across officers and assignments.

Additionally, we are interested in how officer and assignment effects covary, for example $Cov(\theta_i^{IndexArrests}, \theta_i^{Complaints})$ tells us the degree to which officers who tend to receive many complaints due to their individual effect also tend to make more arrests for serious offenses. If this covariance is large and positive, it suggests a trade off in maintaining highly 'productive' officers, as they also produce more misconduct, consistent with models of policing which view misconduct as a byproduct of crime-fighting effort (Shi (2009)) rather than a counterproductive and potentially crime-increasing (Rivera and Ba (2023)).

All variance and covariance components are bias corrected using Kline, Saggio, and Sølvsten (2020)'s (KSS) heteroskedastic-consistent bias correction to all variance components, and I apply the Lachowska et al. (2023b) extension of KSS to construct bias-corrected estimates of cross-outcome correlations.[4]

The main specification for recovering officer and assignment effects is equation (1) with officer effects and GAB effects, with date-specific fixed effects in $x_{i,t}$. Because of the varying time periods and regimes which are captured in the data, we follow Card, Heining, and Kline (2013) and Lachowska et al. (2023a) by estimating equation (1) for overlapping time periods: 2009 - 2010, 2010 - 2011, ..., 2017 - 2018, 2018 - 2019. This allows us to ensure that officer effects are permitted to evolve over time as they gain tenure and assignment effects

[4]The components of the variance decomposition are biased even if point estimates are unbiased (Krueger and Summers (1988), Kline, Saggio, and Sølvsten (2020)) due to measurement error, and this bias is exasperated by weak connectivity or "limited mobility bias" (Andrews et al. (2008), Jochmans and Weidner (2019), Bonhomme et al. (2022)).

can change as the CPD's structure shifts over time. To explore within-group sorting and correlations, I re-estimate the main specification with 2 year periods separately for specific demographic groups: male, female, Black, white, Hispanic, and each 5 year tenure bin (<5, 5-10, 10-15, 15-20, >20).

To test the robustness of the main results to alternative specifications, I also estimate: (1) the main specification for single year periods only; (2) the main specification across three multi-year periods (2009-2012, 2013-2015, and 2016-2019); (3) aggregate data at the weekly level;[5] (4) match-effects model with officer-beat category interactions as θ_i and unit-watch interactions for J.

3.2 Magnitude of Officer Heterogeneity

Table 3 displays bias corrected standard deviations of officer and assignment effects (Columns (1) and (2)), and the total amount of variation explained by officer and assignment effects (Column (3)), and the share of each component of equation (3) explained by officer effects, assignment effects, and the sorting between the two (Columns (4)-(6)). 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.

Focusing on Column (1), differences across individual officers are economically significant. Over 100 shifts, 1 SD increase in officer effects for each respective outcome would increase stops by 41.2 (55 pre-2016 and 15.5 post-2016), nonindex arrests by 9.99, index arrests by 1.62, uses of force by 0.464, and accusations of misconduct by 0.29 and 0.167, for the regular and extended data respectively. For enforcement outcomes, these values are around 90-100% of the mean, while for force and misconduct they are generally over 120% of the mean. Figure 1 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.

[5]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.

Overall, these standard deviations indicate that officer heterogeneity is highly important in changing the policing environment for civilians in terms of enforcement, misconduct, and the use of force, with relatively everyday and discretionary activities such as stops and non-index arrests, relatively low-discretion outcomes such as index arrests, and rare outcomes such as misconduct and force.

We can compare some of these to prior work, as displayed in Figure 2. Adger, Ross, and Sloan (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 and Schanzenbach (2019) finds a 1 SD increase in officer effects for civilian allegations of misconduct is around 140% of the mean. 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 and pooling arrest types.

3.3 Comparison with Assignment Effects

Column (2) in Table 3 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 stops by 35.8 (44.8 pre-2016 and 15.6 post-2016), nonindex arrests by 8.52, index arrests by 1.73, uses of force by 0.287, and accusations of misconduct by 0.236 and 0.121.

For index arrests, assignment effects are generally more important with 1 SD officer effects being around 87% 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 20% for stops, 14% for nonindex arrests, and 3% for index arrests, which are relatively infrequent.

However, I find that officer heterogeneity is significantly larger than assignment heterogeneity for infrequent but highly impactful outcomes like use of force and complaints of misconduct: 1 SD of an officer effect is 162% and around 124% times 1 SD of assignment effects for use of force and complaints. Furthermore, officer and assignment effects only explain less than 1% of the variation in these outcomes. This suggests that highly impactful outcomes such as force and misconduct are driven primarily by random chance, though individual officer heterogeneity plays a much larger role than assignment heterogeneity.

3.4 Sorting

The final component of the variance decomposition in Columns (6) and (7) display the covariance between officer and assignment effects and their correlation. Positive correlations between θ_i^k and ψ_j^k are indicative of officers who causally increase outcome k regardless of where they work tending to work in assignments which also causally increase outcome k regardless of who works in them. For example, a positive correlation for complaints would suggest that officers prone to misconduct prefer working in assignments which permit more misconduct.

Interestingly, I do not find significant amounts of positive assortative matching between officers and assignments for most outcomes, and there is actually consistent evidence of negative assortative matching. The correlation between officer and assignment effects is only consistently positive for nonindex arrests with a correlation around 0.19, while is it near zero for stops, modestly negative for use of force and index arrests (around -0.1), and quite negative for complaints around -0.3. 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.[6]

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. For example, Black officers may use less force but also prefer working in districts with more Black civilians, which also happen to have higher crime rates and use of force due to socioeconomic segregation in Chicago. However, as shown in Figure A.6 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 A.5 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 A.7 displays correlations across outcome subtypes. While officers do no sort on stops overall, this is due to small negative sorting on traffic stops cancelling out small positive sorting on street stops. 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 (e.g., domestic violence 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.

[6]Notably, in comparison with Weisburst (2022)'s decomposition, a relatively small portion of variation in outcomes (arrests in a 911 call) is explained by officer effects (0.5%), but consistent with these findings on sorting for index arrests, officer and assignment effects are very weakly correlated, with a median correlation of -0.09.

3.5 Cross-Outcome Correlations

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, if effects are highly correlated, this suggests that officers are either highly active or lazy; however, if effects are uncorrelated this suggests that officers are specialists in one outcome and their behavior in one dimension is not informative of their behavior in another. Additionally, we are able to probe the common model of police misconduct of Shi (2009), in which 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 4 displays the cross-outcome correlations across officer effects. Interestingly, I do not find any very large (> 0.6) correlations between outcomes of unrelated types (other than complaints and the extended complaints). However, enforcement activities (stops and arrests) are modestly correlated. Officer effects for stops and non-index arrests have the strongest correlation at around 0.37, followed by arrests for index and nonindex offenses at around 0.34. Officer effects for index arrests and stops are modestly correlated at 0.2.

However, officer effects for enforcement activity are essentially uncorrelated with their effects on misconduct, with stops being generally negatively correlated with complaints, index arrests being essentially uncorrelated with complaints, and index arrests and stops are modestly correlated with use of force (0.21 and 0.11). Officers' propensities to make nonindex arrests are modestly positively correlated with complaints for misconduct at around 0.14 and 0.1 and more strongly correlated with use of force at around 0.31. Complaints are, however, most strongly correlated with officers' propensity to use force at around 0.33 and 0.37.[7]

Next, we can decompose the outcomes into subtypes to determine the degree to which

[7]Note that these cross-outcome correlations do not significantly differ across demographic groups, and most differences are not persistent across subsamples (as in Figure A.8).

heterogeneity is being masked by pooling outcomes. For example: index arrests are composed of both violent and property crime arrests; stops include both traffic and street stops; and complaints can be for 'serious' misconduct such as aggressive contact with civilians (excessive force, false arrest), or FPS/conduct misconduct, due to lack of service or poor conduct. Figures A.9-A.11 displays the correlations between these outcome subtypes and the primary outcomes.

Focusing on complaint subtypes, Figure A.11 shows that officer types are actually heterogeneous across forms of misconduct. For example complaints for serious offenses (verbal abuse, excessive force, false arrests), are strongly and positively correlated with the pooled 'complaint' metric, and serious complaint types are generally uncorrelated with service and conduct complaints. Interestingly, in the main results, officer effects for complaints and nonindex arrests are moderately positively correlated (around 0.15-0.2), and this correlation persists between nonindex arrests and serious misconduct, and the correlation between use of force and serious complaints is strong. However, nonindex arrests, index arrests, and use of force are negatively correlated with officer effects for service complaints. This suggests that officers are not only heterogeneous in their enforcement but even the type of misconduct they are accused of, and that officers who are relatively inactive (lower levels of enforcement) also are more likely to be accused of poor conduct and failing to provide service, while the opposite is true for more serious misconduct.

3.6 Influence of Demographics on Officer Heterogeneity

3.6.1 Decomposing Officer Effects

Significant work documents inter-group heterogeneity in police behavior, specifically stops, arrests, and use of force (Ba, Knox, et al. (2021), Goncalves and Mello (2021), Hoekstra and Sloan (2022)). As a result, policy proposals often emphasize the importance of hiring more officers of particular demographic groups in an effort to reduce such enforcement activities through personnel recruitment rather than training or instruction. However, if inter-group

heterogeneity is relatively small compared with intra-group heterogeneity, then drawing specific demographic groups may not actually result in relatively large shifts in enforcement outcomes.

First, we decompose officer effects into the influences of core demographic information (race, gender, start age), experiences (military, education, Spanish-speaking), and experience as a police officer (tenure) captured in Z_i by regressing:

$$\hat{\theta}_i = Z_i\alpha + \mu_i \tag{4}$$

I estimate α, and similar effects, in three ways. First, I estimate equation (4) for each period, then I perform a stacked regression with all periods and include period fixed effects which provides a single estimate of α across periods. For additional robustness, I add specifications with cohort fixed effects added to equation (4), as the month an officer started in the police force, to control for an police academy cohort effects (Rivera (2022)), and this controls for any changes in underlying composition of officers entering the police force (e.g., higher shares of military personnel in a specific year or lack of educational information for some cohorts). Finally, I compare these estimates with the results of using the assignment data to directly estimate the influence of officer demographics on outcomes by adapting the MDSB (month, day of week, shift, beat) estimation approach from Ba, Knox, et al. (2021) by interacting assignments ($J(i,t)$), which contains beat category, shift, and unit, with day of week (D) and year-month (M):

$$y_{i,t} = \alpha z_i + \Psi_{JDM(i,t)} + \epsilon_{i,t} \tag{5}$$

This exploits the variation in the pool of officers that are working on a given day of the week in a specific month, which allows for the identification of the causal effect of a specific demographic factor on outcomes. I estimate (5) for each period as well as on the complete sample.

3.6.2 Observables and Officer Effects

Table 5 displays the results of the main specification estimating equation (4), with the first row for each coefficient being the coefficient of the stacked regression and the second row containing the minimum, median, and maximum estimate over the ten period regressions in brackets — and all estimates are scaled by the appropriate mean of the dependent variable for easy comparison. Consistent with prior literature, Black officers make fewer stops, arrests, uses of force, but they also receive fewer complaints. By comparison, Hispanic officers are only slightly different from white officers making around 5% fewer stops and arrests, but the use force 14% less and receive complaints 12% less — though this shifts over time.

Female officers slightly fewer stops and index arrests (around -13% less) compared with male officers, but they make fewer nonindex arrests (-24%) and use force and receive complaints at a significantly lower rate (-36% - -47%). As a result, while they engage in more stops and index arrests compared with Black officers, they use less force, make fewer low-level arrests, and receive fewer 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 stops, between -2% and -4%, fewer nonindex arrests (-10% and -18%) and index arrests (-7% and -14%), but use substantially less force -24% and -41%, and receive much fewer complaints (-17% and -33%) — though the complaint effects also shift depending on the time frame.

For life experiences, military experience strongly influences officer effects. Officers who were in the military police more aggressively than non-military, which make more stops (+13%), arrests (+20%), and use more force (+12%), but unlike all other groups with more aggressive policing, they actually receive fewer complaints on average (-12% to -24%).

Spanish-speaking officers are negligibly different compared with non-Spanish speaking officers. Education proves to be a negligible factor as well, while officers with recorded educa-

tion information make more stops and arrests and use less force and receive fewer complaints, the difference between high education (BA and above) and low education (associates, credits) is relatively negligible in comparison with the effects on other characteristics.

Finally, officer tenure proves to be the largest factor in changing all outcomes. Officers with 5 - 10 years of experience make 7-10% fewer stops and arrests, but they use 27% less force and receive 14-18% fewer complaints than less experienced officers. However, these effects more than double for officers with 10-15 years of experience, who make 22-21% fewer stops and arrests, use 60% less force, and receive 43-48% fewer complaints. For officers with 15-20 and >20 years of experience, the effects increase in magnitude, though relatively moderately compared to the difference between 5-10 and 10-15, with around 36% and 50% reductions in stops and arrests and around 70% and 80% reductions in complaints and use of force for 15-20 and >20 officers respectively. So, 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.

Figure 3 compares these estimates across the additional specifications, including cohort fixed effects in equation (4), which makes tenure coefficients unable to be estimated for the period-specific samples, and employing the MDSB strategy from Ba, Knox, et al. (2021) as in equation (5). These additional results are highly consistent with the primary results with some minor differences in magnitudes across different outcomes and coefficients.

3.6.3 Influence on Subtypes

The large effects on outcomes may be driven by specific subtype components, which are useful in determining mechanisms. Figure A.12 displays the coefficients (scaled by their mean) for estimating equation (4) on outcome subtypes, such as arrests of various types (e.g., property, violent, drug), stop types (traffic, regular), force types (injured), and complaint types (serious, service). There are a few notable patterns of interest.

First, the differences between Black and white officers in terms of stops driven by fewer

street stops, and a smaller difference in traffic stops, while the differences in nonindex arrests are driven by drug, weapon, and traffic arrests, rather than simple violent crime arrests. Also Black officers make similar numbers of index property arrests, but around 25% fewer violent crime arrests, compared with white officers. For use of force, while Black officers use force at around -34% of the mean compared with white officers, they differ less with respect for force with injury rates at around -15%. For misconduct, Black officers receive significantly fewer complaints for serious misconduct (e.g., excessive force, abuse, false arrests) -25 to -50% of the mean, but they receive more complaints for service and conduct, around +10 - +25% relative to the mean. Collectively, Black officers engage in less discretionary enforcement and make fewer violent crime arrests, though they also use less force and receive fewer force related complaints; however, they appear to have worse conduct and service with (non-suspect) civilians, resulting in more FPS/conduct complaints.

Next, female officers display no differential effects across street or traffic stops, but their effect on nonindex arrests is driven primarily by reductions in drug and weapon arrests. Similar to Black officers, female officers make similar amounts of property arrests compared with male officers but make fewer violent arrests. Interestingly, female officers not only use less force overall, but injure suspects at a significantly lower rate (over -50% of the mean), and they receive significantly fewer complaints for serious and service/conduct complaints. Overall, this suggests that female officers, in comparison to male officers, only moderately reduce enforcement but significantly reduce their negative spillovers in force and misconduct while providing better service to civilians.

Military officers displayed a significantly different pattern from other demographics: while they engage in more enforcement and force activities, they nevertheless receive fewer complaints of misconduct. I find that military officers' higher rates of stops is driven by traffic stops, and they make significantly more nonindex arrests of almost all subtypes. They also make more violent, and to a lesser extent, more property arrests, and their force use is also more likely to result in injuries. Nevertheless, they receive fewer complaints for misconduct

due to near zero effects on serious complaints, but generally negative effects on conduct and service complaints. This suggests that officers with military experience are more aggressive on all accounts, however they maintain lower misconduct rates, which may be due to better training, more experience in high pressure situations, being more polite, or being better at following of procedures.

Finally, the lack of results for education prove to consistent across subtypes. Interestingly, tenure's effects on officer behavior is also generally symmetric across subtypes, with the exception of complaint subtypes. While complaints for serious misconduct decline as tenure increases, in line with an overall decline in officers' activity, service and conduct complaints are lowest for the 10-15 and 15-20 tenure bins, and the median effect of tenure actually increases (the negative effect becomes smaller) for the 20+ tenure bin. This is consistent with the oldest officers being increasingly inactive but also possibly less responsive to civilians compared with high tenure but slightly younger officers.

3.6.4 Intra-Group Heterogeneity

The estimates of α suggest that inter-group heterogeneity is significant in magnitude, however it is not clear if officers of groups with lower mean propensity are not also more varied in their behavior — for example, Black officers may have a lower average force rate but their distribution of force use may be wider.

Figure A.13 displays the standard deviations for officers of specific race, gender, and tenure groups divided by the standard deviation of the pooled sample. In general, white and male officer SDs are closest to the pooled SD, consistent with them being the plurality and majority of officers, and Hispanic officers do not tend to have significantly different SDs from white officers. Black officers have a tighter distribution, by comparison, with 1SD for Black officers being around 75% of an SD for the pooled sample for index arrests, nonindex arrests, complaints (though the extended complaints are around 85% of the pooled SD), stops, and force. Female officers by comparison have a similar SD to male officers for index

arrests, but much smaller SDs for other outcomes, particularly use of force, nonindex arrests, and complaints at around 60-70% of the pooled SD. These differences in SDs are consistent with the differences in averages as well, meaning that Black and female officers have lower propensities for enforcement and much lower propensities for force and misconduct on average and the distribution of officer effects within these groups also tends to be smaller.

For tenure, similar to the reduction in average officer effects as officers gain experience, we see a decline in variance among officers as tenure increases for most outcomes, however this decline becomes most striking after 10 years of experience, as variation is highest among officers with 5-10 years of experience (compared with <5) for stops, complaints (extended sample), and index arrests, possibly due to officers being constrained in their activity early on due to more senior partners or probationary periods. Interestingly, even as officers increase tenure, the variation in officer effects for index arrests remains substantial, being the lowest for officers with 10-15 and 15-20 years of experience, though both are close to the SD of the pooled sample.

4 Crime Effects and Team Production

Given the importance of officer heterogeneity on enforcement outcomes, we turn to the effect of officers on crime rates. The difficulty in quantifying an officer's effect on crime is largely a result of the difficulty in attributing a crime occurring or not occurring to an individual officers, rather than the work of a collective team of officers. Nevertheless, the effect of officers on crime is potentially more important than their influence on enforcement as crimes impose a high social cost, are much more frequent than enforcement, and are the primary focus of modern police forces. While prior work has found that the quantity of police influences crime (Evans and Owens (2007), Mello (2019), Klick and Tabarrok (2005), Di Tella and Schargrodsky (2004), Draca, Machin, and Witt (2011)), it is unclear if officers are heterogeneous in their crime-effects (e.g., does it matter who we hire to reduce crime?) and

how these crime-effects correlated with costly enforcement measures (e.g., do officers that reduce crime through deterrence further reduce costs by making fewer arrests?).

4.1 Empirical Strategy

We are interested in recovering the effect of specific officers on the number of crimes reported, cleared, and uncleared. Because officers are not generally assigned to specific locations and specific crimes cannot be directly tied to an individual officer, measuring the effect of a specific officer on the occurrence of crime is empirically difficult. To overcome these challenges, I model the number of crimes c of some type (e.g., uncleared robberies) which occur in district d in date-hour t as the additive effect of the team τ of officers working in district d at time t, $\Omega_{\tau(d,t)}$, the environment-specific effects, $\gamma_{w(d,t)}$, and additional features $X_{d,t}$, and exogenous shocks $\epsilon_{d,t}$

$$c_{d,t} = \Omega_{\tau(d,t)} + \beta X_{d,t} + \gamma_{w(d,t)} + \epsilon_{d,t} \tag{6}$$

The team effect, $\Omega_{\tau(d,t)}$, is a function of the officer-specific effects working in team $\tau(d,t)$. Specifically, I model this as an additive production function:

$$\Omega_{\tau(d,t)} = \frac{1}{N_{\tau(d,t)}} \sum_{i \in \tau(d,t)}^{N_{\tau(d,t)}} \omega_i \tag{7}$$

where $\gamma_{w(d,t)}$ captures fixed effects for the interaction between the year, quarter (or month), district, day of week, and shift (3 8-hour periods each day) captures the fixed calendar structure. $X_{d,t}$ include a second degree polynomial of the total number of officers on the team $N_{\tau(d,t)}$ and the number of officers working various functions in the district.

The identifying assumption is that $E[\epsilon|X, \Omega, \gamma] = 0$. 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, as discussed in Section 2.3, the calendar system significantly reduces the likelihood of such systematic selection, and γ controls for sorting across primary

assignments (unit and watch) as well as day-of-week interacted with quarter or month 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).

Under these assumptions, officer effects can be recovered by OLS with matrix form:

$$C = B\omega + G\gamma + X\beta + \epsilon \tag{8}$$

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 $1/N_{d,t}$, such that $B\omega = \Omega$. G is a matrix of indicators with one per row and γ contains the district-time fixed effects. Because we do not observe districts with no police, ω_i's can not be point-identified, rather they can only be identified relative to an excluded reference officer.

Similar to the connectivity issues in the AKM model, identification of ω requires that (at least) B is non-singular. 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 A.1).

I recover estimates of all parameters, ω, γ, β, then partial out low-dimensional controls: $\tilde{C} = C - X\beta$. Given estimates of $\hat{\omega}$, we can construct team effect $\hat{\Omega} = B\hat{\omega}$ and decompose the variance in (partialled) crime into:

$$Var(\tilde{C}) = Var(\Omega) + Var(\gamma) + 2Cov(\Omega, \gamma) + Var(\epsilon) \tag{9}$$

This allows us to determine the share of variation in crime that is attributable to officers: $Var(\Omega)/Var(\tilde{C})$. Higher values indicate that heterogeneity across officers explain more of the variation in crime — this does not tell us which officers increase or decrease crime in

absolute terms only relative to other officers. Additionally, we can compute correlations across outcomes, for example $Cov(\Omega^{property}, \Omega^{violent})$ tell us how the effect of teams on property crime covaries with their effect on violent crime — a positive value indicates that teams which reduce property crime more than average also reduce violent crime more than average.

As in the case of the officer-assignment model in Section 3.1, the estimated variance component or plugin estimates of $Var(\hat{\Omega})$ will be biased upwards while covariance components will be biased downwards. While Bonhomme (2021) proposes a homoskedasticity robust correction and utilizes team-specific scaling factors, in this environment the Kline, Saggio, and Sølvsten (2020) (KSS) assumption of the number of officers and number of teams tending to infinity is reasonable. So, I apply the KSS bias correction to $Var(\hat{\Omega})$ and the KSS extension to cross-outcomes correlations from Lachowska et al. (2023b).

4.2 The Effect of Teams on Crime

The central question in this analysis is if officers have any influence on overall crime and if so to what degree? Across specifications and periods of analysis, a few patterns are clear: the variation in the officers working in a given district at a given time (the team) explain a varying portions of the variation in crime and the bias correction has a significant impact on the interpretation. Table 6 summarizes the primary statistics of interest in this section.

For violent crime, the variation in teams explains little, on its own, of the variation across crime rates, with $Var(\hat{\Omega})/Var(\tilde{Y})$ being generally less than 1% across crime types, and along with unit-time fixed effects (year-unit-watch-day of week - month or quarter), team and unit-time effects only explain around 7% of the variation in violent crime overall, with the lowest amount being murder at around 0.3%. As shown in Column (4), if not for the KSS bias correction, the influence of teams would be significantly overstated as around 90% of the 'plug-in' $Var(\hat{\Omega})$ is due to bias, and the bias often exceeds 100%. In comparison with the mean, a 1 SD increase in a teams crime effect increases violent crime by around 27% of the mean, murder by around 90%, robbery by 38%, and aggravated assault by 54% of their

respective means.

For property crime, bias corrected variation in team effects explain significantly more of the variation than that of violent crime. Shares of bias in the plug-in estimates are relatively modest for pooled index property crime at around 35% for the median, ranging from 27% to 42%, but bias is much larger relatively for burglary (with a median of 61%, though never crossing 100%) and motor vehicle theft (median of 76% and maximum of 120%). The relative precision in property crime is driven by larceny/theft, which for which plug-in estimates are only around 40% bias. Reflecting this, the variation in teams alone explains around 13% of pooled property crime, 6% of burglary (with a minimum of 0.1%), 2.2% of motor vehicle theft (with a negative minimum), and 11% of theft relative to the variance of the outcomes — and the addition of unit-time effects does not generally add significant explanatory power. Relative to the mean, this relatively large contribution of teams to property crime means that a 1 SD increase in a teams' crime effects increases property crime by 62%, burglary by 85% (minimum of 13%), motor vehicle theft by 69%, and theft by 71% of their respective means. Overall, we can reject that which officers are working does not have an impact on reported crime rates.

Finally, for nonindex crimes, team effects prove to be slightly more influential than on other crime types. Unsurprisingly, drug crimes (which are synonymous with drug arrests as the clearance rate is nearly 100%), are strongly influenced by who is working, and the plug-in estimates are less biased than all other outcomes at around 27% bias. Team effects alone explain around 15% of the overall variation, and a 1 SD increase in team effects increases drug crimes by around 140% of the mean. Weapons crimes, which are also highly cleared, are similarly influenced by who is working, though team effects explain a much smaller share of the variation (less than 3%) and the share of bias in the plug-in estimates are much higher than for drug crime (around 76%).

Simple violent crime, such as domestic violence and misdemeanor battery, and other nonindex crimes are influenced by teams more similarly to violent crimes. Variation in team

effects explains around 5-6% of the variation in these crimes, and while the share of bias in plug-in estimates is around 50%, a 1 SD increase in crime effects increases these crimes by slightly less than 50% of their respective means.

We can probe the robustness of these results using the alternative specifications, by collapsing the data down to 8 hour time periods and 4 hour time periods. Both of these reduce the relative number of observations per parameter significantly. Figure 4 displays the standard deviations as a percent of the mean outcome for each specification. While increasing levels of aggregations tend to reduce the magnitudes of team variance and increase the likelihood of the plug-in estimates being entirely attributable to bias, these results do not significantly change the prior interpretation. Teams have relatively small or highly variable influences on violent crimes. Yet, team composition's influence on property crime remains strong, with 1 SD in team effects equating to around 40-70% of the mean for pooled index crimes, driven by relatively stable effects and similarly sized effects on theft. Team effects for burglary are relatively large compared with their mean, but occasionally have small effects, though almost all are between 50% and 120% of the mean. And, as aggregation increases, the estimated influence of teams on motor vehicle theft significantly declines. Finally, drug crimes remain strongly influenced by teams, though the relative magnitudes decline as aggregation increases (to around 70% of the mean per 1 SD for watch-level estimates), but effects on simple violent crime remain around 50% of the mean across specifications.

4.3 Quantifying Crime Effects

To quantify the magnitude of these effects, we can focus on how a 1 SD better officer changes crime outcomes. The prior results have focused on 1 SD in team effects, which is equivalent to increasing every officer's crime effect on a team by 1 SD by the additive model in equation (7), while the effect of a single 1 SD better officer is around 2% of a 1 SD better team. To make the results comparable to the prior section on officer-level outcomes we can compute the number of crimes 'prevented' by replacing a single average officer on a team with a 1

SD 'better' (lower) officer over the course of 100 shifts, given an average team size of 59 and average shift length of 9 hours.

Figure A.14 displays the number of crimes prevented by decreasing an average officer's crime effect by 1 SD over the course of 100 shifts across specifications. Over the course of 100 shifts, a 1 SD higher officer will increase reported drug crimes by around 2 crimes (driven entirely by drug arrests), which is around 80% of the mean number of drug arrests per 100 shifts at the officer level, and similar in proportion to 1 SD in officer effects on nonindex arrests in Section 3. Simple violent crime and other nonindex crimes increase by a similar amount, though weapons crimes only increase by around 0.25 crimes. For property crimes, 1 SD increase in officer effects increases property crime by around 4 per 100 shifts, composed of around 1 burglary, 0.5 motor vehicle thefts, and around 3 larceny/thefts. Finally, for violent crime 1 SD is equivalent to less than 0.5 violent crimes, primarily driven by aggravated assault.

Figure 5 also provides a comparison for the influence of the quantity and the 'quality' of officers. Much of the effect of police on crime literature studies the influence of the number of officers (quantity) on crime rates (Levitt (1997), McCrary (2002), Levitt (2002), Evans and Owens (2007), Weisburst (2019), Chalfin and McCrary (2018), Mello (2019), Chalfin et al. (2022)), with the results commonly reported in elasticities: the effect of a 1% increase in the number of officers causing some percent decrease in crime. This section focus on the 'quality' of officer in their individual crime prevention abilities. So, we can compare whether more crime would be prevented by hiring an additional officer (expanding the average team size by 1.7% from 59 to 60 officers) or by replacing a single officer on the team with a 1 SD 'better' officer. I use elasticity estimates from Mello (2019) and Chalfin and McCrary (2018) (CM) (which includes standard errors) to compute these quantity effects.

For violent crimes, both the Mello and CM quantity effects are moderately larger than the quality effects, and this is true across subtypes other than aggravated assaults, where the quality estimates are noticeably larger than the quantity estimates. For overall property

crime, the quality effects are between the CM estimates and Mello estimates. Across subtypes, the results suggest that while quantity of officers is more important than the quality for motor vehicle theft, that quality is more important for burglary. Overall, quantity of officers appears to be more influential for violent crime and motor vehicle theft, while quality is just as important for aggravated assaults, burglary, and theft.

4.4 Mechanisms: Deterrence, Incapacitation, and Reporting

While the prior results indicate that officers do influence the amount of crime that is reported during their shifts, the precise mode through which they influence it is unclear. First, officers may be reducing crime through contemporaneous deterrence: 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. In the literature focusing on how the number of officers influences crime rates, the most plausible channel is that the quantity of officers influences crime through deterrence rather than incapacitation (Mello (2019), Chalfin et al. (2022)), while Mastrobuoni (2019) finds that more officers reduces crime through arresting and incapacitating repeated offenders of robbery. Finally, it is possible that officers are influencing crime reports through increased reporting of crime or suppressing crime reports. In this section, I attempt to disentangle these potential mechanisms.

First, if officers are reducing crime through arrests, 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 = Arrests + Uncleared the variance of any crime type can be decomposed into:

$$Var(\Omega^{Crime}) = Var(\Omega^{Arrest}) + Var(\Omega^{Uncleared}) + 2 \times Cov(\Omega^{Arrest}, \Omega^{Uncleared}) \quad (10)$$

Figure A.16 displays the components of equation (10) scaled 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 would be consistent with incapacitation. However this covariance between arrests and uncleared crimes declines in magnitude as the levels of aggregation increases for theft and aggravated assault, suggesting incapacitation may play a secondary role for some outcomes.

Next, it is possible that individual officers report or suppress crime reports with different frequencies, and the effects on crime are not measuring the influence of officers on actual crimes occurring. To obtain a different proxy for crime occurrences, I use 911 call data as the outcome of interest. Figure A.17 displays SDs of team effects on 911 calls of different types and the share of variance that team effects explain. Consistent with officers actually influencing underlying crime rates, team effects have a non-negligible influence on 911 calls made relative to the means and overall variation.

Figure A.18 displays the correlations between team effects on 911 calls with their effects on crime reports — though these are not perfectly one-to-one as a disturbance call may be, for example, a domestic violence incident, but could also be vandalism or burglary. Consistent with officers who influence crime reports also influencing underlying crime rates, the correlations between 911 call and crime team effects are as we expect: team effects for 911 calls for assaults, battery, and domestic violence are highly correlated with crime reports of aggravated assault, simple violent crime, and index violent crime, while team effects calls for theft, burglary, and alarms are strongly correlated with crime reports of theft, burglary, index property crime. Overall, this is consistent with officers actually influencing crime rates,

as we would expect, if they were simply suppressing reports, that team effects on crime and team effects on 911 calls would be negatively correlated or uncorrelated.

Finally, crime effects appear to be positively correlated across outcomes, as shown in Figure A.15, meaning that officers who tend to reduce simple violent crime also reduce theft. Nevertheless, violent and property crime effects tend to be only weakly positively correlated, while violent crime effects for index (e.g., aggravated assault) and simple violent crime are positively correlated. This would suggest some amount of specialization in officer types rather than officers recoding higher forms of violent crime as lower types, which would constitute suppression. Additionally, drug crime effects are positively correlated with both violent and property crime effects, consistent with officers who make more arrests being less capable at projecting deterrence.

4.5 Demographics and Crime Effects

Given that the composition of teams of officers appears to have a robust influence on crime outcomes, it is important to determine if crime reduction effects can be predicted based on observable features of the officers and teams. To explore this, I employ multiple methods of estimate the effect of officer characteristics and officers' measured propensities for enforcement, force, and misconduct from Section 3.

First, I directly estimate the effect of a team's observable composition α, where z_i is a vector of officer i's observable (or measurable) characteristics, on crime by estimating

$$c_{d,t} = \frac{1}{N_{\tau(d,t)}} \sum_{i \in \tau(d,t)}^{N_{\tau(d,t)}} z_i \alpha + \beta X_{d,t} + \gamma_{w(d,t)} + \epsilon_{d,t} \tag{11}$$

where $X_{d,t}$ controls for the overall number of officers working in district d and day-hour t. This effectively replaces the team of Ω with average values of the team's observables from equation (6).

Next, I replace $c_{d,t}$ with estimated team effects $\hat{\Omega}_{\tau(d,t)}$ using the recovered estimates of

$\hat{\omega}_i$ in the above sections:

$$\hat{\Omega}_{\tau(d,t)} = \frac{1}{N_{\tau(d,t)}} \sum_{i \in \tau(d,t)}^{N_{\tau(d,t)}} z_i \alpha + \beta X_{d,t} + \gamma_{w(d,t)} + \epsilon_{d,t} \tag{12}$$

I include controls such as $\gamma_{w(d,t)}$ in this specification to account for the weighting of officers in teams which may be correlated with $\epsilon_{d,t}$ if not for the inclusion of $\gamma_{w(d,t)}$.

Next, I allow for team composition, $\frac{1}{N_{\tau(d,t)}} \sum_{i \in \tau(d,t)}^{N_{\tau(d,t)}} z_i$, to be endogenous, and I instrument for the composition by exploiting the calendar assignment. Specifically, I use the predicted rotation and working times based on the calendar and the method discussed in Section 2.3 to predict which officers should be working in a specific unit at a specific time. I construct $B^c Z = \frac{1}{N^c_{\tau^c(d,t)}} \sum_{i \in \tau^c(d,t)}^{N^c_{\tau^c(d,t)}} z^c_i$, where c denotes the team weights applied across officers as predicted by the calendar, and I instrument for the potentially endogenous team composition in equations (11) and (12) with the first stage:

$$\frac{1}{N_{\tau(d,t)}} \sum_{i \in \tau(d,t)}^{N_{\tau(d,t)}} z_i = \frac{1}{N^c_{\tau^c(d,t)}} \sum_{i \in \tau^c(d,t)}^{N^c_{\tau^c(d,t)}} z^c_i \rho + \beta X_{d,t} + \gamma_{w(d,t)} + \eta_{d,t} \tag{13}$$

To test the robustness of these results, I use both year-quarter-district-day of week-8 hour period fixed effects ('Shift') and year-month-district-day of week-hour ('Hour') fixed effects for $\gamma_{w(d,t)}$.

Finally, I focus the relationship between officer observables and their individual crime effects by estimating:

$$\hat{\omega}_i = z_i \alpha + \chi_i \tag{14}$$

which I estimate weighting by officer-shifts as in the officer-level analysis in Section 3.

For all above specifications, I estimate them separately for each 2.75 year sample period and in a stacked regression with the addition of sample period fixed effects when necessary. For the officer-level analysis, officer-level fixed effects used as regressors are shrunken using empirical Bayes following Weisburst (2022). For brevity and precision, I focus on three

primary outcomes: index violent crimes, index property crimes, and drug crimes.

The results for the stacked team-level regressions (OLS and IV) are displayed in Figure 6, which contains the coefficients on demographics, and Figure 7, which contains coefficients on officers' behavioral fixed effects. As shown in Figure 6, the addition of officer's behavioral fixed effects do not significantly change the coefficients on officer demographics, and often OLS and IV effects are highly similar in direction and magnitude. Furthermore, using team effects as the outcome does not significantly change the coefficients compared with using the actual crime rate, though the effects on team effects are artificially precise. However, the more stringent Hour fixed effects controls often significantly reduce the magnitudes of coefficients.

For violent crimes in Figure 6, no demographic has a significant and consistently precise influence. In Figure 7, we see that increasing a teams average officer-effects for stops, nonindex arrests, and use of force can reduce violent crimes by around 10% of the mean with stringent controls and using IV. This suggests that more 'active' officers can modestly reduce violent crime. However, the effects are only significant when using crime rates, not predicted team effects, as the outcome. Overall, this lack of robust effects is not entirely unexpected, given that the primary result from the prior subsections was that team effects explained little of the variation in violent crime.

For property crimes, where team effects exhibited a larger influence, IV estimates using Hour controls indicate that 1 unit increase in team composition of Black or Hispanic (instead of white) officers would decrease property crime by around 10% of the mean. The most consistent pattern is that higher tenure officers tend to increase property crimes, measured in both team effects and crime rates. A 1 unit increase in the share of 15-20 and 20+ tenure officers increases property crimes by around 10% and 20% of the mean— though the crime rates results are somewhat smaller than the team effects results. In Figure 7, again more 'active' officers appear to reduce property crimes when using crime rates as the outcome. In particular, increasing a teams average officer-level propensity for stops by 1 SD decreases

property crimes by around 20% (IV) and 5% (OLS) of the mean, when using the Hour fixed effects. On the other hand, officers with higher propensities for arrests do not display any large effects. Again, this is consistent with more 'active' officers reducing crime, though not through incapacitation and arrests, but through being visible and active.

For drug crimes (which are entirely due to arrests), using the Hour fixed effects and accounting for officer-level behavioral effects significantly reduces the magnitudes of any demographic effects in Figure 6. As shown in Figure 7, increasing the team's average behavioral effect for making drug arrests by 1 SD increases drug crimes by between 75% and 125% of the mean, depending on the specification and controls.

Finally, we turn to the officer-level results from estimating equation (14), as displayed in Figure 8 for officer demographics and Figure 9 for officer behavioral effects. The primary results are largely consistent with the prior analysis for demographics: most effects are not statistically or economically significant. However, the point estimates for Black officers on violent crime tend to be negative (around 10% of the mean for the pooled sample regression), and higher tenure officers tend to increase property crime (by 25% and 40% of the mean for 15-20 and 20+ years of tenure). For officers' behavioral effects, officers who make 1 SD more stops (all else equal) reduce property crime by around 5% of the mean, while officers who make 1 SD more drug arrests increase drug arrests by around 100% of the mean, which is expected.

Overall, I do not find much evidence for a significant relationship between officers' and teams' effectiveness at preventing crimes and many of their demographic characteristics or measurable propensities. The results I do find, however, are consistent with: 1. officer effects on violent crime being small or difficult to measure, 2. team-level measures of drug crimes being effectively proxied by officers' individual behavioral propensities for drug arrests, and 3. more 'active' officers who are younger and make more stops being better at reducing property crime.

5 Conclusion

This paper documents a number of facts about the degree of heterogeneity in police officers enforcement outcomes, their influence on crime, and their tendencies to sort into assignments. Overall, officer heterogeneity, beyond average differences between demographic groups, is substantial and an important consideration in both hiring and training policy. The results also indicate that officers vary in their ability to prevent crime beyond their propensities to make arrests.

6 References

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." *Longitudinal Employer-Household Dynamics Technical Papers*, March.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Adger, Chandon, Matthew Ross, and CarlyWill Sloan. 2022. "The Effect of Field Training Officers on Police Use of Force." Working Paper.

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 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): 673–97.

Ba, Bocar. 2018. "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago." Working Paper.

Ba, Bocar, Patrick Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé. 2021. "Police Officer Assignment and Neighborhood Crime." Working Paper 29243. National Bureau of Economic Research.

Ba, Bocar, Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. "The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science* 371 (6530): 696–702.

Bonhomme, Stephane. 2021. "Teams: Heterogeneity, Sorting, and Complementarity." arXiv. http://arxiv.org/abs/2102.01802.

Bonhomme, Stephane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2022. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, March.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics* 128 (3): 967–1015.

Chalfin, Aaron, Benjamin Hansen, Emily K. Weisburst, and Morgan C. Williams Jr. 2022.

"Police Force Size and Civilian Race." *American Economic Review: Insights* 4 (2): 139–58.

Chalfin, Aaron, and Justin McCrary. 2017. "Criminal Deterrence: A Review of the Literature." *Journal of Economic Literature* 55 (1): 5–48.

———. 2018. "Are u.s. Cities Underpoliced? Theory and Evidence." *The Review of Economics and Statistics* 100 (1): 167–86.

Cheng, Cheng, and Wei Long. 2018. "Improving Police Services: Evidence from the French Quarter Task Force." *Journal of Public Economics* 164 (August): 1–18.

Devi, Tanaya, and Jr Fryer Roland G. 2020. "Policing the Police: The Impact of 'Pattern-or-Practice' Investigations on Crime." Working Paper Series. National Bureau of Economic Research. https://www.nber.org/papers/w27324.

Di Tella, Rafael, and Ernesto Schargrodsky. 2004. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack." *American Economic Review* 94 (1): 115–33.

Donohue, John J., Daniel E. Ho, and Patrick Leahy. 2013. "Do Police Reduce Crime?: A Reexamination of a Natural Experiment." In *Empirical Legal Analysis*. Routledge.

Donohue, John J., and Steven D. Levitt. 2001. "The Impact of Legalized Abortion on Crime." *The Quarterly Journal of Economics* 116 (2): 379–420.

Draca, Mirko, Stephen Machin, and Robert Witt. 2011. "Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks." *American Economic Review* 101 (5): 2157–81.

Evans, William N., and Emily G. Owens. 2007. "COPS and Crime." *Journal of Public Economics* 91 (1): 181–201.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–41.

Gonçalves, Felipe M., and Steven Mello. 2023. "Police Discretion and Public Safety." Working Paper Series. National Bureau of Economic Research. https://www.nber.org/papers/w31678.

Gudgeon, Matthew, Andrew Jordan, and Taeho Kim. 2023. "Do Teams Perform Differently Under Black and Hispanic Leaders? Evidence from the Chicago Police Department." Working Paper. Working Paper.

Hoekstra, Mark, and CarlyWill Sloan. 2022. "Does Race Matter for Police Use of Force? Evidence from 911 Calls." *American Economic Review* 112 (3): 827–60.

Jochmans, Koen, and Martin Weidner. 2019. "Fixed-Effect Regressions on Network Data." *Econometrica* 87 (5): 1543–60.

Klick, Jonathan, and Alexander Tabarrok. 2005. "Using Terror Alert Levels to Estimate the Effect of Police on Crime." *The Journal of Law and Economics* 48 (1): 267–79.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out Estimation of Variance Components." *Econometrica* 88 (5): 1859–98.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (2): 259.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury. 2023a. "Do Firm Effects Drift? Evidence from Washington Administrative Data." *Journal of Econometrics* 233 (2): 375–95.

———. 2023b. "Work Hours Mismatch." Working Paper Series. National Bureau of Economic Research. https://www.nber.org/papers/w31205.

Levitt, Steven D. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *The American Economic Review* 87 (3): 270–90.

———. 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply." *American Economic Review* 92 (4): 1244–50.

Mastrobuoni, Giovanni. 2019. "Police Disruption and Performance: Evidence from Recurrent Redeployments Within a City." *Journal of Public Economics* 176 (August): 18–31.

McCrary, Justin. 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment." *American Economic Review* 92 (4): 1236–43.

———. 2007. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality

of Police." *The American Economic Review* 97 (1): 318–53.

Mello, Steven. 2019. "More COPS, Less Crime." *Journal of Public Economics* 172 (April): 174–200.

Miller, Amalia R, and Carmit Segal. 2018. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *The Review of Economic Studies* 86 (5): 2220–47.

Premkumar, Deepak. 2022. "Public Scrutiny, Police Behavior, and Crime Consequences: Evidence from High-Profile Police Killings." Working Paper.

Prendergast, Canice. 2021. "'Drive and Wave': The Response to LAPD Police Reforms After Rampart." Working Paper.

Rivera, Roman. 2022. "The Effect of Minority Peers on Future Arrest Quantity and Quality." SSRN Scholarly Paper 4067011. Rochester, NY: Social Science Research Network.

Rivera, Roman, and Bocar A. Ba. 2023. "The Effect of Police Oversight on Crime and Misconduct Allegations: Evidence from Chicago." *The Review of Economics and Statistics*, October, 1–45.

Rozema, Kyle, and Max Schanzenbach. 2019. "Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct." *American Economic Journal: Economic Policy* 11 (2): 225–68.

Shi, Lan. 2009. "The Limit of Oversight in Policing: Evidence from the 2001 Cincinnati Riot." *Journal of Public Economics* 93 (1): 99–113.

Weisburst, Emily K. 2019. "Safety in Police Numbers: Evidence of Police Effectiveness from Federal COPS Grant Applications." *American Law and Economics Review* 21 (1): 81–109.

———. 2022. "'Whose Help Is on the Way?': The Importance of Individual Police Officers in Law Enforcement Outcomes." *Journal of Human Resources*, March.

Table 1: Summary Statistics of Officer Data

	Pooled	By Year	By 2-Year and LCC
	(1)	(2)	(3)
A: Demographics and Experiences			
White	0.481	0.481	0.481
Black	0.23	0.23	0.232
Hispanic	0.251	0.251	0.249
Asian/Native American	0.0379	0.0379	0.0377
Male	0.806	0.806	0.808
Female	0.194	0.194	0.192
Age	41	41	41.2
Tenure	12.2	12.2	12.3
Military	0.225	0.225	0.223
Spanish Speaking	0.145	0.145	0.145
Edu >= BA	0.308	0.308	0.308
Edu < BA	0.371	0.371	0.372
Edu Missing	0.321	0.321	0.32
B: By Officer			
Avg. Total GABs	17.3	5.19	7.15
Avg. Total ABs	50	12	17.9
Avg. Total Units	2.66	1.26	1.46
Avg. Share Most Common GAB	0.447	0.695	0.636
Avg. Obs. (Shifts)	1,505	182	349
C: Pooled			
Geographic Unit	0.802	0.802	0.804
Total Units	44	39.7	39.9
Total Beats	19,288	5,962	7,577
Total Assignments (AB)	36,287	9,118	11,881
Total Assignments (GABs)	2,721	1,306	1,255
Unique POs	14,218	8,193	8,486
In Car	0.802	0.802	0.805
N Obs.	14,361,345	1,306,595	2,520,143
D: Outcomes (per 100 Shifts)			
Arrest Index	1.61	1.61	1.59
Arrest NonIndex	10.6	10.6	10.4
Stop	44.6	44.6	45.6
Force	0.347	0.347	0.347
Complaints	0.236	0.236	0.225
Complaints (Ext.)	0.0924	0.0924	0.0904

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 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	295,418.910	0.092	0.166	0.142	0.024	0.118	0.135	0.023	0.112
Murder	5,891.452	0.002	0.374	0.003	0.001	0.002	0.003	0.001	0.002
Robbery	127,653.805	0.040	0.091	0.061	0.006	0.056	0.059	0.006	0.053
Agg. Assault	144,906.357	0.045	0.229	0.069	0.016	0.054	0.066	0.015	0.050
Index Property									
Property	1,013,793.196	0.316	0.095	0.486	0.046	0.440	0.496	0.046	0.450
Burglary	175,588.964	0.055	0.054	0.084	0.005	0.080	0.089	0.005	0.085
Motor Vehicle Theft	129,950.538	0.041	0.070	0.062	0.004	0.058	0.065	0.004	0.061
Theft	703,707.512	0.221	0.110	0.337	0.037	0.300	0.340	0.037	0.303
Nonindex									
Nonindex	1,901,168.942	0.592	0.373	0.912	0.340	0.571	0.885	0.312	0.573
Drug	307,665.452	0.096	0.996	0.148	0.147	0.001	0.130	0.129	0.000
Simple Violent	659,176.777	0.207	0.217	0.316	0.069	0.248	0.306	0.067	0.239
Weapon	46,142.708	0.014	0.776	0.022	0.017	0.005	0.020	0.016	0.005
All Other	888,184.005	0.279	0.252	0.426	0.107	0.318	0.428	0.100	0.329

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 3: Decomposition of Officer-Level Results

	Per 100 Shifts		Share of Var(Y)				
	SD(PO)	SD(Assign)	Total Exp.	Var(PO)	Var(Assign)	2 x Cov(PO, Assign)	Cor(PO, Assign)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stops	41.2	35.8	0.204	0.111	0.0816	0.0007	0.00314
	[14.2, 60.8]	[15.2, 54.5]	[0.154, 0.22]	[0.0677, 0.134]	[0.0686, 0.0991]	[-0.0158, 0.0188]	[-0.0719, 0.135]
Arrests - Nonindex	9.99	8.52	0.144	0.0658	0.0527	0.0225	0.186
	[6.79, 11.8]	[6.25, 11.1]	[0.116, 0.154]	[0.0517, 0.0782]	[0.0375, 0.0678]	[0.0148, 0.0311]	[0.111, 0.253]
Arrests - Index	1.62	1.73	0.0258	0.0131	0.0159	-0.00261	-0.1
	[1.12, 2.03]	[1.53, 2.1]	[0.0223, 0.0313]	[0.00756, 0.0188]	[0.0142, 0.0206]	[-0.00905, 0.0005]	[-0.261, 0.0227]
Uses of Force	0.464	0.287	0.00718	0.0059	0.00247	-0.0009	-0.124
	[0.351, 0.544]	[0.248, 0.336]	[0.00606, 0.00821]	[0.0045, 0.00712]	[0.00156, 0.00318]	[-0.00201, -0.0003]	[-0.224, -0.0493]
Complaints	0.29	0.236	0.00427	0.00376	0.00244	-0.00177	-0.299
	[0.241, 0.389]	[0.198, 0.295]	[0.0037, 0.00607]	[0.00305, 0.00493]	[0.00205, 0.00285]	[-0.00218, -0.00128]	[-0.339, -0.203]
Complaints (Ext)	0.167	0.121	0.0029	0.00314	0.00178	-0.00143	-0.31
	[0.112, 0.256]	[0.0845, 0.231]	[0.00246, 0.00435]	[0.00179, 0.00687]	[0.00109, 0.00618]	[-0.0102, -0.0005]	[-0.786, -0.165]

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 and assignment effects recovered from estimating equation (1), multiplied by 100 to refer to the effect of a 1 SD increase in officer/assignment effects per 100 shifts. Columns (3)-(6) display the components of equation (2) scaled by 1 over 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 4: Cross-Outcome Correlation in Officer Effects

	Arrests - Nonindex	Arrests - Index	Use of Force	Complaints	Complaints (Ext)
	(1)	(2)	(3)	(4)	(5)
Stops	0.372	0.214	0.129	-0.0988	-0.0172
	[0.32, 0.49]	[0.14, 0.31]	[-0.057, 0.22]	[-0.21, 0.059]	[-0.29, 0.2]
Arrests - Nonindex		0.354	0.318	0.192	0.146
		[0.22, 0.54]	[0.25, 0.44]	[-0.15, 0.34]	[-0.04, 0.27]
Arrests - Index			0.219	0.0458	0.0374
			[0.13, 0.3]	[-0.11, 0.13]	[-0.097, 0.15]
Uses of Force				0.357	0.372
				[0.16, 0.42]	[0.11, 0.52]
Complaints					0.722
					[0.17, 0.78]

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 5: Effect of Observables on Officer Effects

	Stops	Arrests - Nonindex	Arrests - Index	Uses of Force	Complaints	Complaints (Ext)
	(1)	(2)	(3)	(4)	(5)	(6)
Demographics						
Black	-0.307	-0.182	-0.13	-0.342	-0.271	-0.176
	[-0.32, -0.23, -0.12]	[-0.27, -0.18, -0.11]	[-0.21, -0.14, -0.06]	[-0.63, -0.29, -0.18]	[-0.26, -0.19, 0.008]	[-0.45, -0.12, 0.11]
Hispanic	-0.052	-0.05	-0.04	-0.143	-0.115	0.011
	[-0.08, -0.04, -0.01]	[-0.08, -0.04, -0.02]	[-0.12, -0.04, 0.06]	[-0.28, -0.12, -0.08]	[-0.14, -0.11, 0.07]	[-0.17, 0.05, 0.33]
Other Race	-0.041	-0.086	-0.007	-0.172	-0.077	0.018
	[-0.08, -0.02, 0.0002]	[-0.17, -0.1, -0.03]	[-0.2, -0.03, 0.2]	[-0.23, -0.17, -0.08]	[-0.22, -0.07, 0.16]	[-0.25, 0.06, 0.44]
Female	-0.131	-0.242	-0.14	-0.467	-0.45	-0.356
	[-0.19, -0.11, -0.08]	[-0.37, -0.23, -0.13]	[-0.27, -0.14, -0.03]	[-0.56, -0.45, -0.38]	[-0.37, -0.32, -0.24]	[-0.42, -0.38, -0.19]
S. Age = [25,28]	0.009	-0.058	-0.04	-0.117	-0.011	0.007
	[-0.05, 0.004, 0.05]	[-0.07, -0.03, -0.004]	[-0.1, -0.03, 0.02]	[-0.17, -0.11, -0.02]	[-0.03, 0.004, 0.04]	[-0.08, 0.02, 0.18]
S. Age = [29,32]	-0.02	-0.101	-0.067	-0.243	-0.168	-0.071
	[-0.12, -0.02, 0.03]	[-0.15, -0.06, -0.01]	[-0.13, -0.06, 0.09]	[-0.37, -0.24, -0.08]	[-0.14, -0.12, -0.03]	[-0.18, -0.04, 0.26]
S. Age = >33	-0.045	-0.182	-0.136	-0.413	-0.331	-0.237
	[-0.12, -0.04, 0.03]	[-0.22, -0.16, -0.06]	[-0.2, -0.13, -0.008]	[-0.5, -0.37, -0.27]	[-0.26, -0.23, -0.17]	[-0.4, -0.26, 0.15]
Experiences						
Spanish Speaking	-0.002	-0.018	0.022	0.017	0.121	0.042
	[-0.04, -0.003, 0.04]	[-0.06, -0.03, 0.004]	[-0.03, 0.009, 0.11]	[-0.02, 0.01, 0.11]	[-0.06, 0.06, 0.26]	[-0.24, -0.02, 0.31]
Military	0.134	0.228	0.199	0.117	-0.119	-0.242
	[0.1, 0.2, 0.4]	[0.14, 0.37, 0.48]	[0.1, 0.24, 0.39]	[0.08, 0.24, 0.39]	[-0.15, -0.03, 0.04]	[-0.33, -0.09, 0.05]
Edu >= BA	0.154	0.024	0.073	-0.007	-0.092	-0.185
	[0.04, 0.09, 0.16]	[-0.05, -0.02, 0.04]	[-0.003, 0.08, 0.13]	[-0.17, -0.06, 0.00003]	[-0.21, -0.04, -0.007]	[-0.37, -0.2, -0.07]
Edu < BA	0.092	0.017	0.062	-0.019	-0.069	-0.127
	[-0.02, 0.03, 0.13]	[-0.06, -0.01, 0.03]	[0.03, 0.06, 0.11]	[-0.13, -0.09, -0.02]	[-0.18, -0.06, 0.06]	[-0.38, -0.16, 0.03]
Tenure						
Tenure 5-10	-0.097	-0.071	-0.089	-0.274	-0.183	-0.135
	[-0.38, -0.21, 0.06]	[-0.25, -0.06, 0.22]	[-0.31, -0.09, 0.36]	[-0.51, -0.32, 0.08]	[-0.2, -0.11, -0.02]	[-0.58, -0.07, 0.02]
Tenure 10-15	-0.309	-0.256	-0.223	-0.604	-0.479	-0.43
	[-0.54, -0.34, -0.04]	[-0.53, -0.22, 0.09]	[-0.46, -0.22, 0.25]	[-0.89, -0.53, -0.21]	[-0.39, -0.33, -0.01]	[-0.55, -0.35, -0.05]
Tenure 15-20	-0.37	-0.355	-0.358	-0.78	-0.66	-0.608
	[-0.6, -0.36, -0.11]	[-0.53, -0.32, -0.07]	[-0.61, -0.31, 0.07]	[-1.1, -0.7, -0.41]	[-0.55, -0.42, -0.23]	[-0.8, -0.6, -0.08]
Tenure >20	-0.484	-0.451	-0.541	-0.906	-0.829	-0.591
	[-0.76, -0.46, -0.26]	[-0.64, -0.46, -0.18]	[-0.75, -0.53, -0.1]	[-1.3, -0.86, -0.53]	[-0.65, -0.62, -0.26]	[-0.91, -0.53, 0.01]

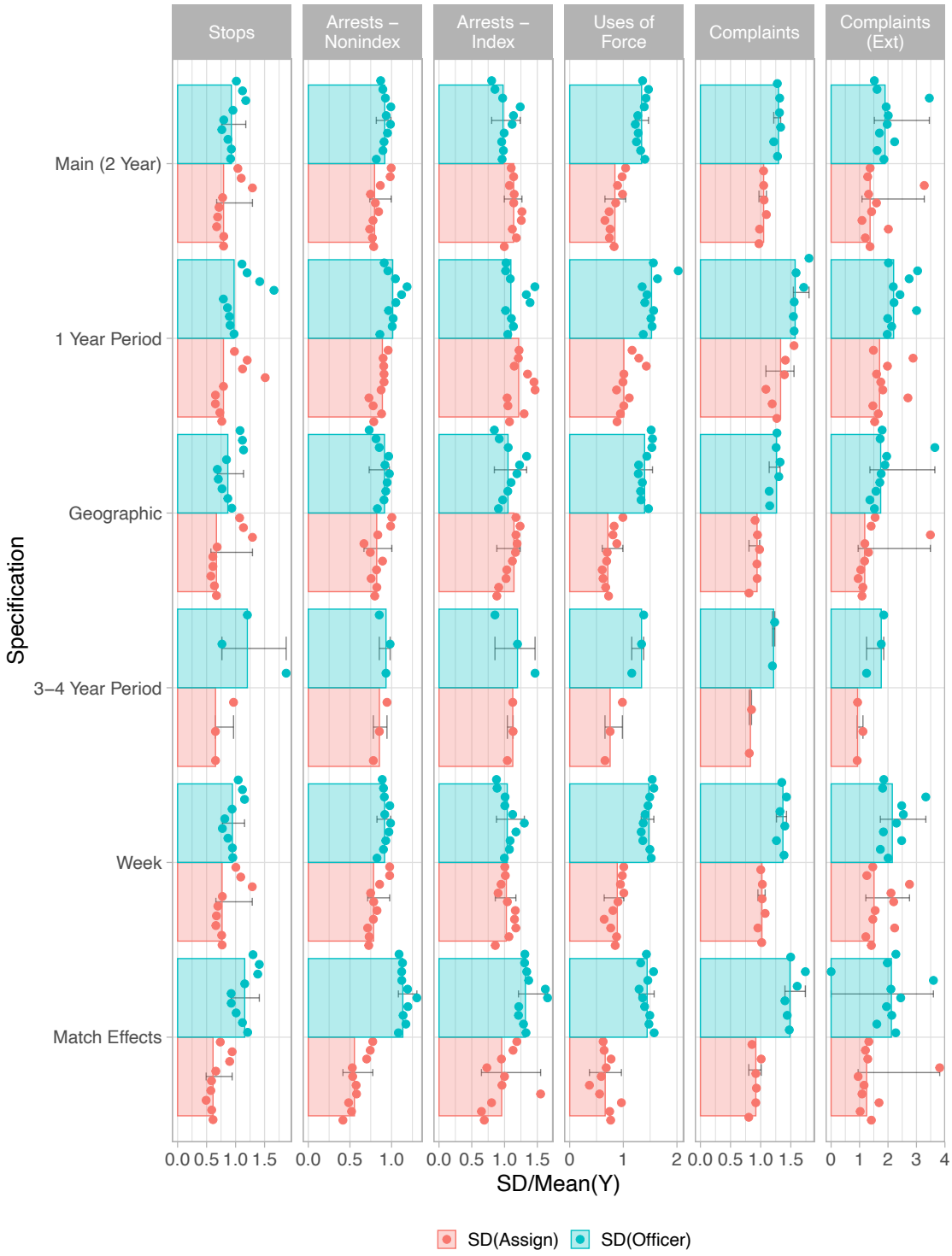
Note: Table presents results from estimating equation (4) on officer effects scaled by the mean of the relevant outcome, with the regression weighted by the officer's observations. Rows reflect the coefficient on each officer-specific variable, with the first row reflecting the stacked regression coefficient and the second row containing the minimum, median, 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.

Table 6: Decomposition of Team Effects on Crime

	SD(Team)/Mean(Y)	Var(Team)/Var(Y)	Share Var(Y) Explained	Share Bias
	(1)	(2)	(3)	(4)
Violent				
Index Violent	0.271	0.00762	0.0718	0.912
	[0, 0.71]	[-0.0061, 0.053]	[0.069, 0.08]	[0.64, 1.1]
Murder	0.935	0.0024	0.00329	0.967
	[0, 2.9]	[-0.052, 0.013]	[0.0014, 0.0046]	[0.89, 1.8]
Robbery	0.377	0.00674	0.0323	0.926
	[0, 0.99]	[-0.014, 0.045]	[0.026, 0.037]	[0.66, 1.2]
Agg. Assault	0.542	0.0163	0.047	0.844
	[0, 0.97]	[-0.018, 0.044]	[0.041, 0.051]	[0.66, 1.3]
Property				
Index Property	0.623	0.128	0.16	0.353
	[0.52, 0.67]	[0.082, 0.16]	[0.14, 0.17]	[0.27, 0.42]
Burglary	0.85	0.0579	0.041	0.613
	[0.13, 1.1]	[0.001, 0.077]	[0.019, 0.062]	[0.49, 0.99]
Motor Vehicle Theft	0.685	0.022	0.0224	0.755
	[0, 0.93]	[-0.022, 0.047]	[0.017, 0.024]	[0.68, 1.2]
Theft	0.713	0.114	0.174	0.399
	[0.53, 0.75]	[0.06, 0.13]	[0.16, 0.2]	[0.28, 0.48]
Nonindex				
Drug	1.4	0.152	0.246	0.273
	[1.3, 1.9]	[0.094, 0.24]	[0.21, 0.27]	[0.23, 0.35]
Simple Violent	0.486	0.0517	0.12	0.553
	[0.41, 0.63]	[0.037, 0.089]	[0.11, 0.13]	[0.52, 0.67]
Weapon	1.33	0.0262	0.023	0.76
	[0, 2]	[-0.0025, 0.095]	[0.019, 0.048]	[0.41, 1]
All Other	0.469	0.0624	0.116	0.534
	[0.39, 0.51]	[0.047, 0.079]	[0.11, 0.12]	[0.48, 0.64]

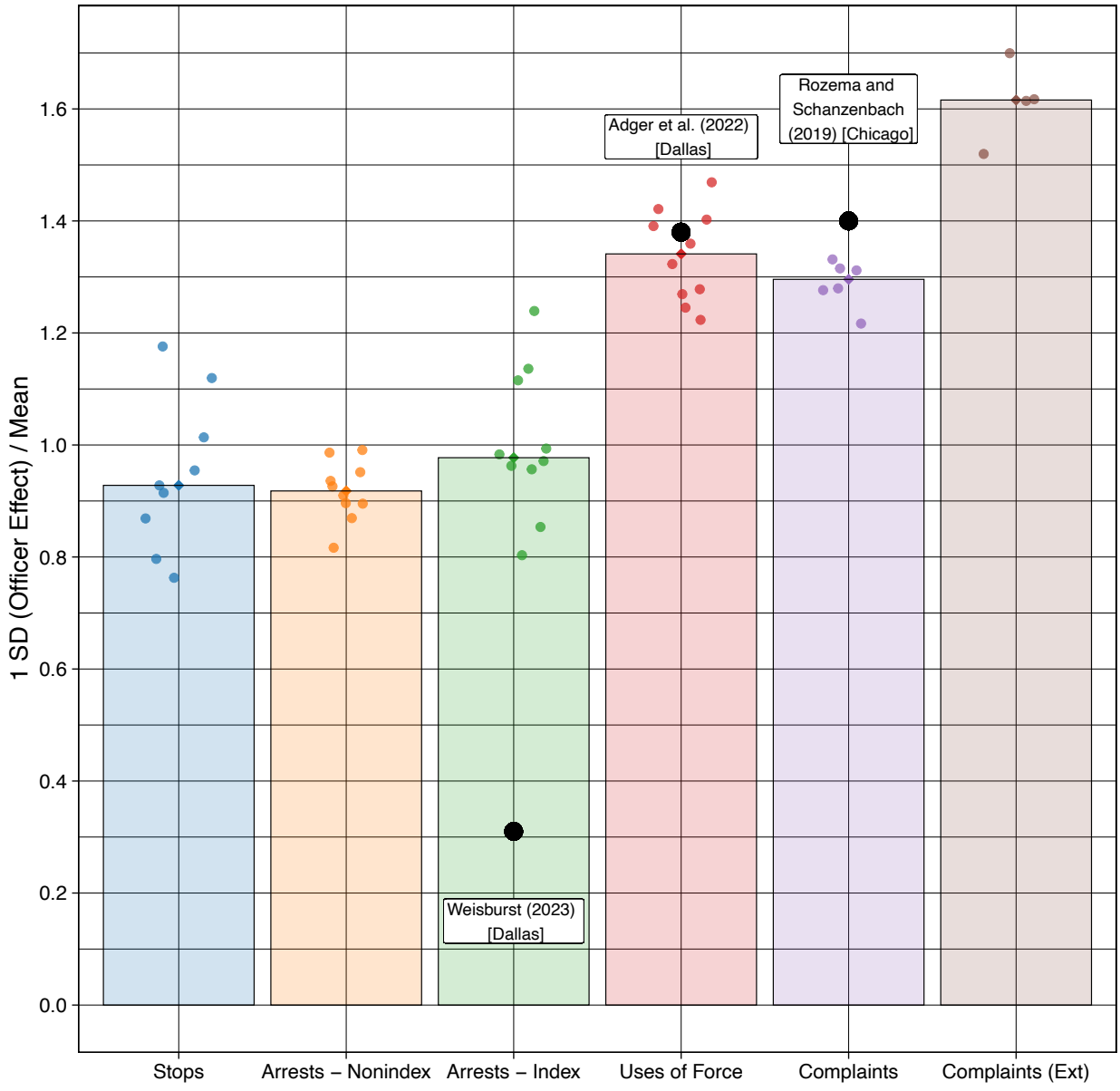
Note: Table displays summary of various statistics (columns) by outcome variable (rows) which are crimes per district-hour per 100,000 population. The first row for each outcome is the median estimate across four 2.75 year sample groups while the second row contains the minimum and maximum across sample groups. Estimates are bias corrected using KSS other than Column (4). Team effects 'Team' are recovered from estimating equation (6) with team effects being the average of officer effects (equation (7)). SD(Team) and Var(Team) refer to the standard deviation and variance of team effects across observations within each sample group. Column (3) contains the overall share of variance explained by team effects and district-time fixed effects. Column (4) contains the share of the uncorrected ('plug-in') variance of team effects which is attributable to bias. Zero or negative values in Columns (1)-(3) and values of 1 or more in Column (4) reflects plug-in estimates being entirely attributable to bias.

Figure 1: Magnitude of Officer and Assignment Effect Heterogeneity by Specification



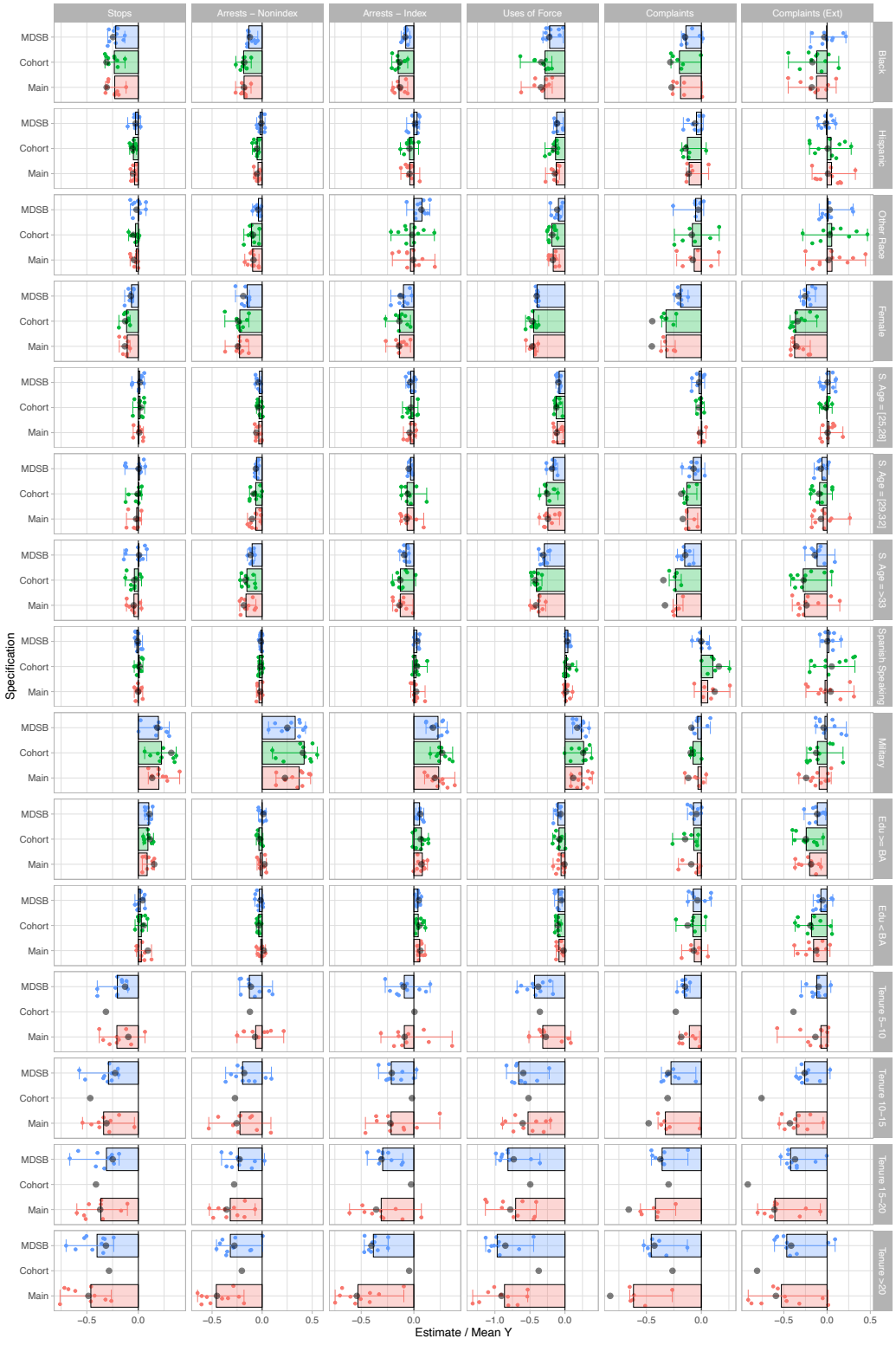
Note: Figure displays the KSS bias-corrected standard deviations of officer and assignment effects (recovered from equation (1)) across different specifications scaled by the mean of the outcome, with each point denoting an estimate for a specific time period, the bar denoting the median across these 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; 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.

Figure 2: Comparison of Officer Heterogeneity with Prior Estimates



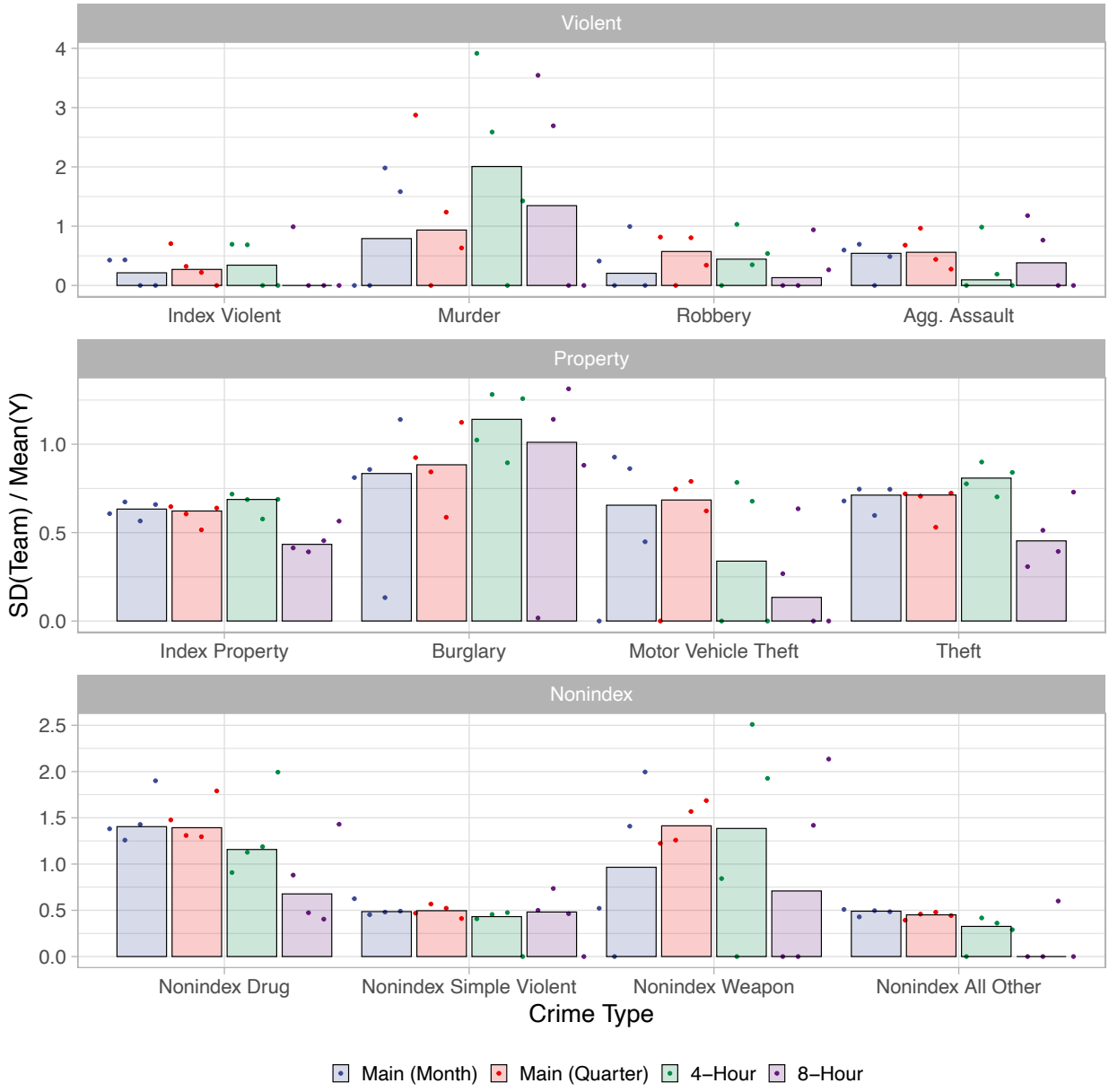
Note: Figure displays one KSS bias corrected standard deviation in officer effects scaled by the mean of the outcome, with officer effects recovered from the main specification in equation (1), and each colored point denotes an estimate for a 2 year period and bars denote the median of these estimates. Black dotes and associated labels denote estimates of standard deviations of officer effects divided by their respective means for associated outcomes in prior literature along with the city of the study.

Figure 3: Effect of Officer Characteristics on Officer Effects and Outcomes by Specification



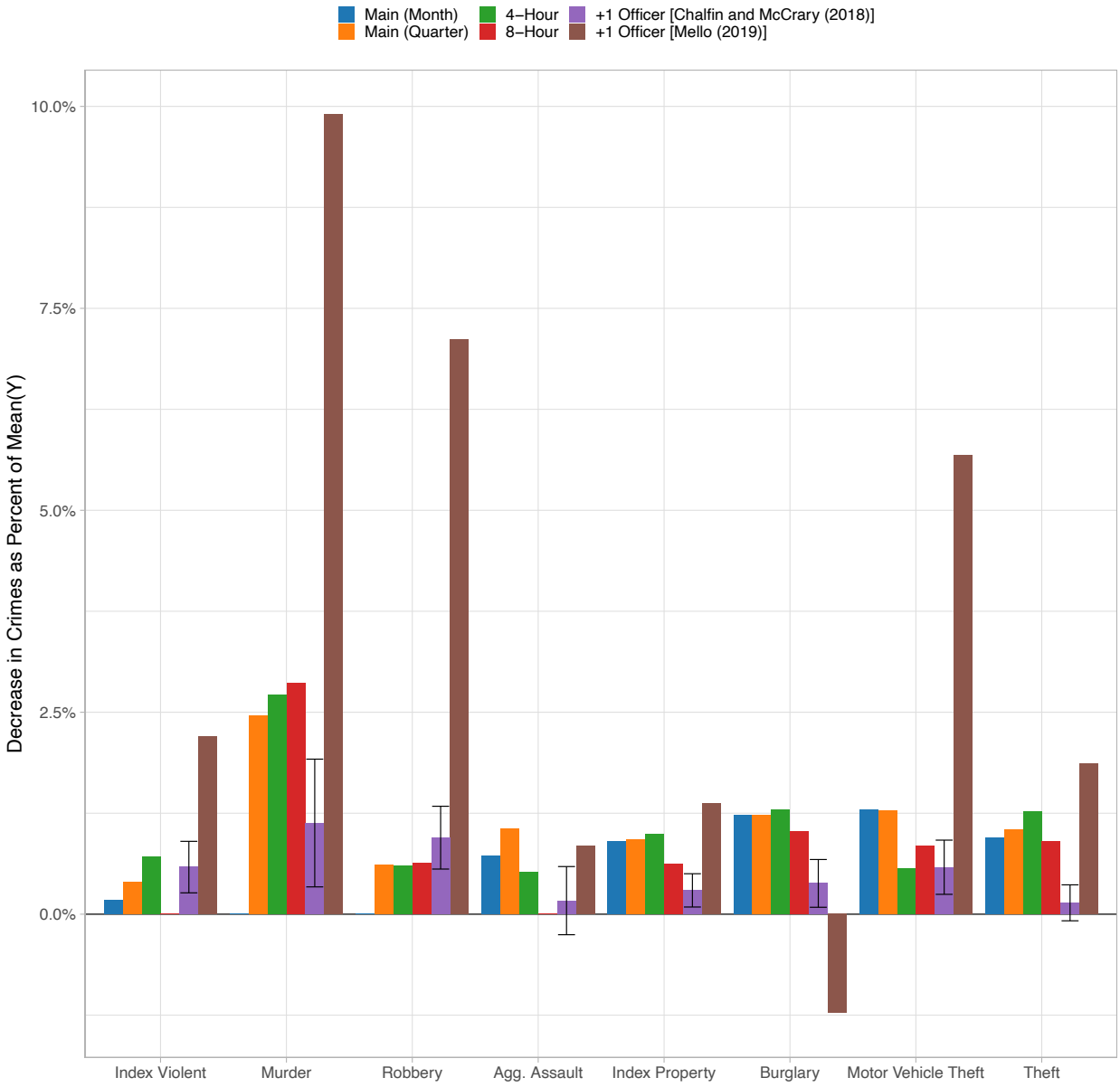
Note: Figure displays the coefficients of the effect of officer characteristics on officer effects (recovered by estimating equation (1)) and outcomes, which are the results of estimating equation (4) for 'Main' and 'Cohort' specifications and equation (5) for 'MDSB' specification. All estimates are scaled by the mean of the outcome. Each smaller point is an estimate from the 10 2-year period samples, the bars are the median of these estimates, and the lines span the minimum and maximum of the estimates, while the larger points denote the estimate of the stacked regressions using all years. Cohort specifications add cohort fixed effects to the main specification, while MDSB regresses officer characteristics on outcomes rather than officer fixed effects and uses GAB - year - month fixed effects as controls.

Figure 4: Magnitude of Team Heterogeneity for Crime Outcomes



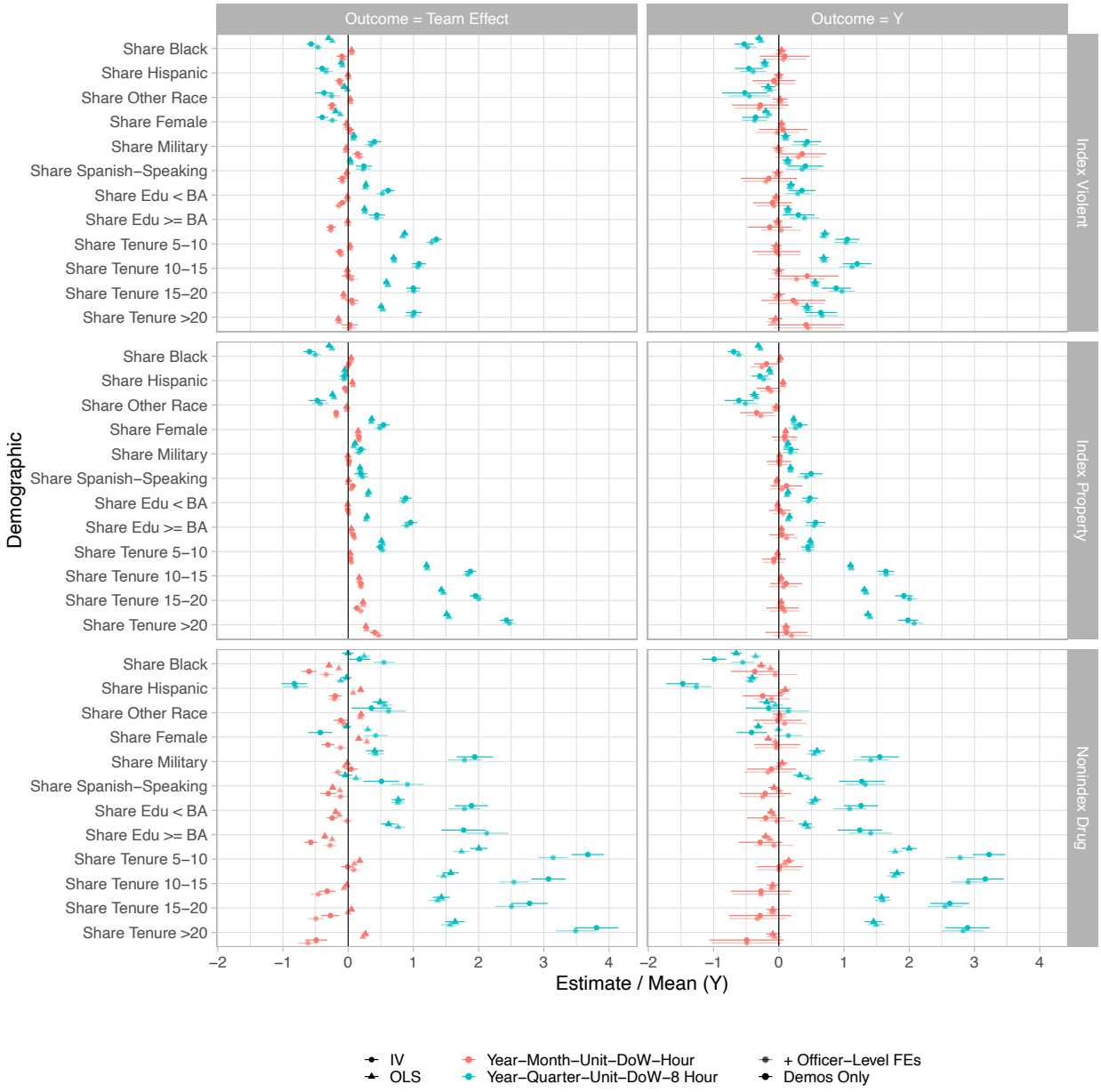
Note: Figure displays the KSS bias corrected standard deviations of team effects (Ω), recovered by estimating equation (6) scaled by the mean of the outcome variable for each crime type and across different specifications, across the 3 main crime categories. All outcomes used were scaled at crimes per 100,000 population. 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, while 4-Hour and 8-Hour aggregate the data to 4 and 8-hour periods using the same level of controls interacted with quarter. Each point represents an estimate recovered from one of four sample periods each 2.75 years long, and bars denote the median of these estimates. Values of zero indicate that the bias correction produced a zero or negative variance for team effects.

Figure 5: Effect of Officer Quantity versus Quality



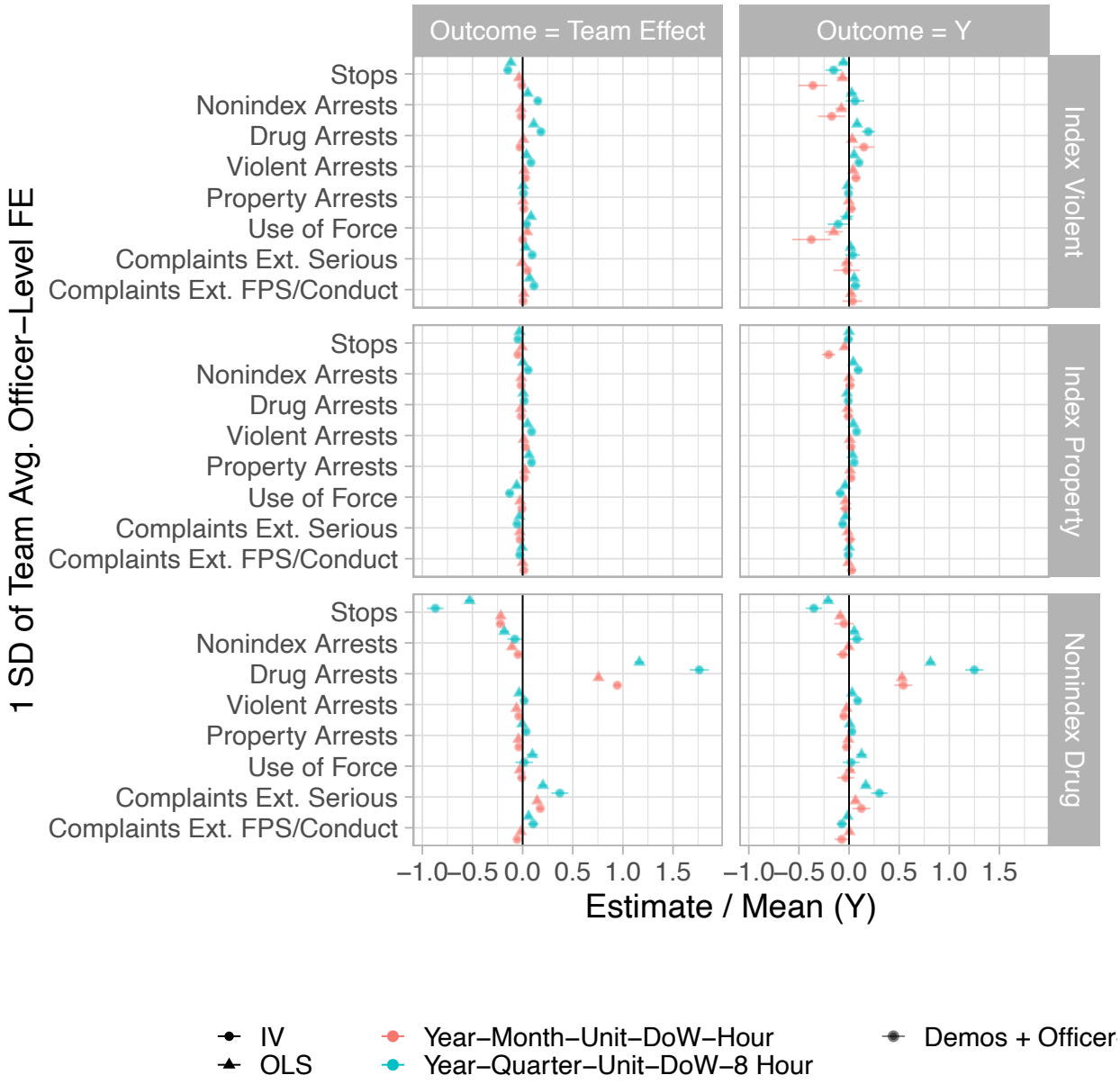
Note: Figure displays comparison between the effect of 1 additional officer added to a team using elasticities of crime with respect to number of officers from estimates from Mello (2019) and Chalfin and McCrary (2018) across index crime types ('quantity' effect) compared with the effect of reducing 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 1/59 (the percent increase in number of officers by adding 1 officer to the average team of 59 officers). Quality estimates are computed from the KSS bias corrected standard deviations of team effects (Ω), recovered by estimating equation (6), with each estimate being divided by 59 (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, while 4-Hour and 8-Hour aggregate the data to 4 and 8-hour periods using the same level of controls interacted with quarter. Each bar for the non-quantity estimates represents the median of estimates recovered from each of the four 2.75 year long 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.

Figure 6: Effect of Team Demographics on Team Effects and Crime Rates (1)



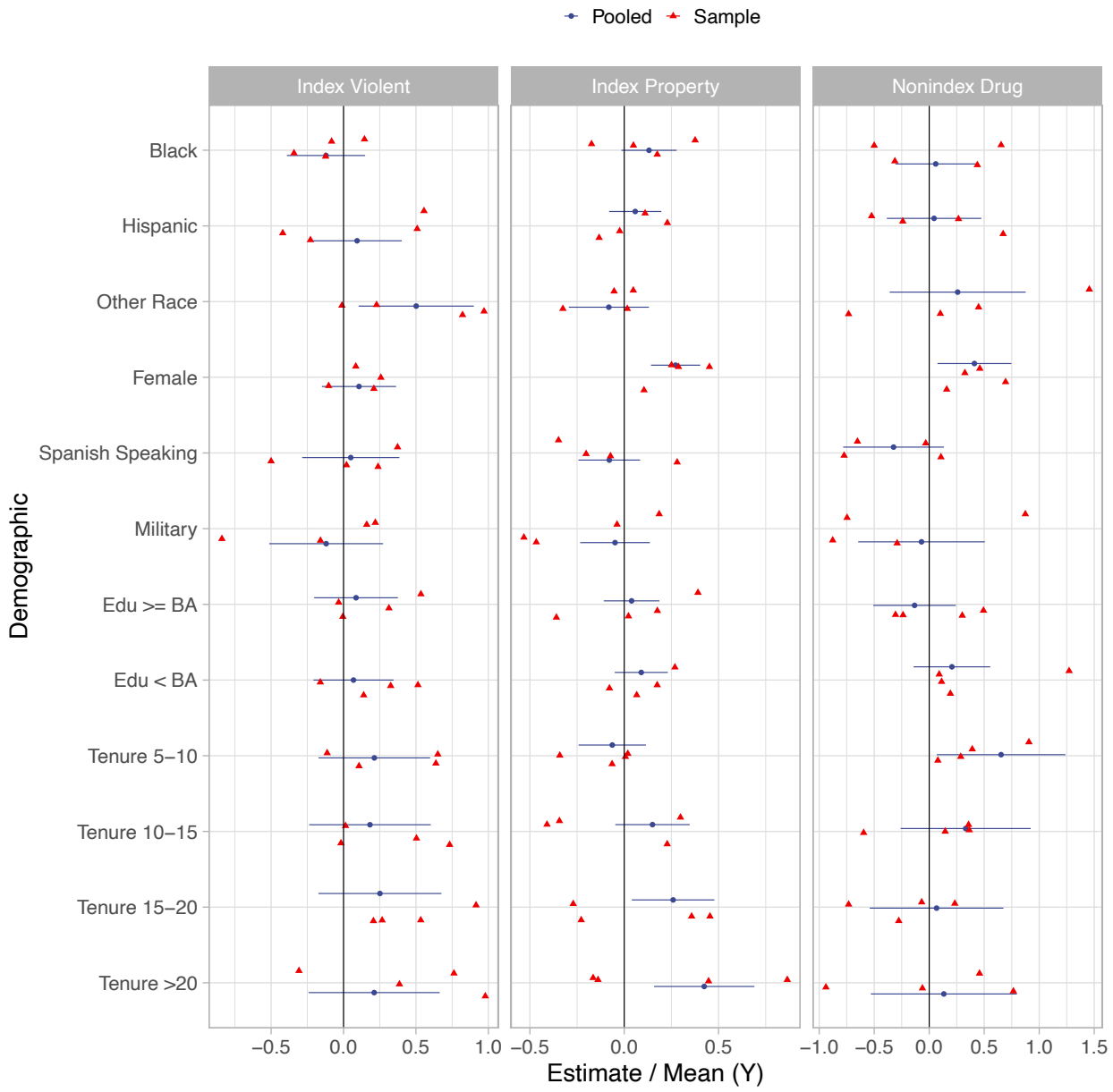
Note: Figure displays the coefficients from regressing predicted team effects (recovered from estimating equation (6)) (left panel) or crime rates (right panel), at the district-hour level, on the demographic composition of the team of officers, scaled by the mean crime rate, for three crime types. Multiple specifications are used: OLS refers to estimating equations (11) and (12) directly, IV instruments for the composition of officers using the predicted set of officers working using the rotation calendar (equation (13)); time fixed effects are the interaction of year-quarter-unit-day of week and 8-hour time periods or year-month-unit-day of week and 1-hour time periods; the specifications either include only officer demographics or also include officers' individual behavior fixed effects recovered in Section 3 and displayed in the following figure. Regressors are the effect of a 1 unit increase in the share of officers of a given type on a team. Confidence intervals are 95% confidence intervals computed from standard errors clustered at the Hour or Shift fixed effect level – confidence intervals for team effect estimates are not bias corrected and artificially precise.

Figure 7: Effect of Team Average Officer-Effects on Team Effects and Crime Rates (2)



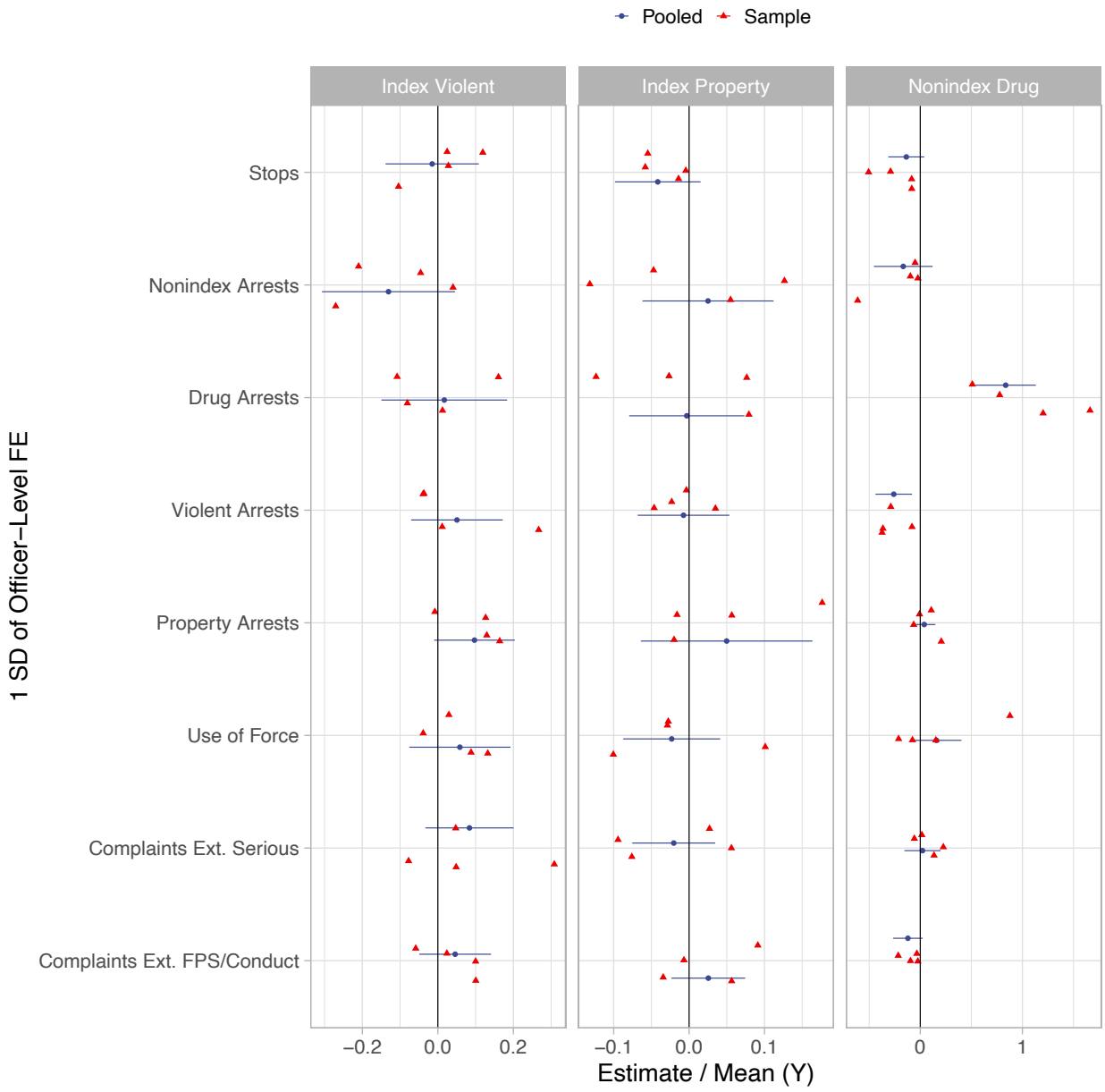
Note: Figure displays the coefficients from regressing predicted team effects (recovered from estimating equation (6)) (left panel) or crime rates (right panel), at the district-hour level, on the demographic composition and average behavior fixed effects (recovered in Section 3) of the team of officers, scaled by the mean crime rate, for three crime types. Multiple specifications are used: OLS refers to estimating equations (11) and (12) directly, IV instruments for the composition of officers using the predicted set of officers working using the rotation calendar (equation (13)); time fixed effects are the interaction of year-quarter-unit-day of week and 8-hour time periods or year-month-unit-day of week and 1-hour time periods; the specifications also include officers' demographics with those coefficients displayed in the previous figure. Effects correspond to the effect of increasing a teams' average officer-level fixed effects for a given outcome by 1 SD. Confidence intervals are 95% confidence intervals computed from standard errors clustered at the Hour or Shift fixed effect level – confidence intervals for team effect estimates are not bias corrected and artificially precise.

Figure 8: Effect of Officer Demographics on Crime Effects (1)



Note: Figure displays the coefficients from regression predicted officer-specific effects on crime (ω) (recovered from estimating equation (6)) scaled by the mean crime rate, by estimating equation (14) at the officer - sample group level (each sample group contains 2.75 years and sample group fixed effects are included). Regressors include officer-level demographics and standardized measures of officer-level behavioral fixed effects (recovered in Section 3) and shrunken using empirical Bayes – the fixed effect coefficients are displayed in the next figure. Each red point denotes an estimate from estimating equation (14) on a single sample group, while the pooled point and 95% confidence interval (from heteroskedastic robust standard errors) were produced from a stacked regression including sample group fixed effects.

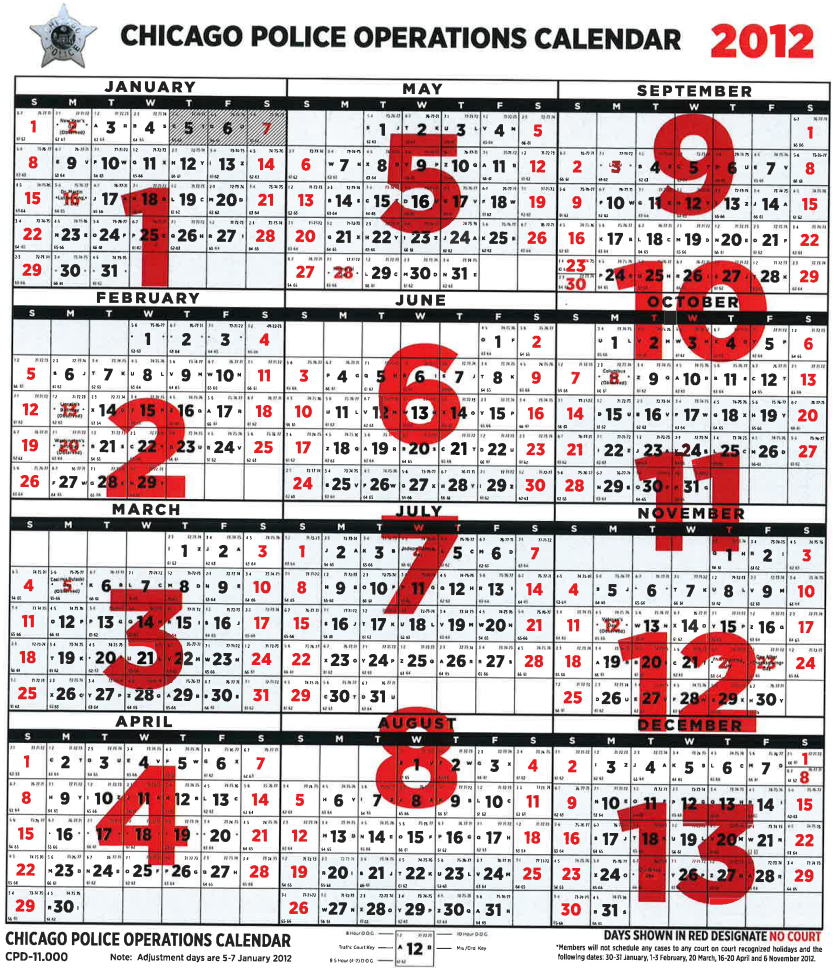
Figure 9: Effect of Officer-Effects on Crime Effects (2)



Note: Figure displays the coefficients from regression predicted officer-specific effects on crime (ω) (recovered from estimating equation (6)) scaled by the mean crime rate, by estimating equation (14) at the officer - sample group level (each sample group contains 2.75 years and sample group fixed effects are included). Regressors include officer-level demographics and standardized measures of officer-level behavioral fixed effects (recovered in Section 3) and shrunken using empirical Bayes – the demographic coefficients are displayed in the prior figure. Each red point denotes an estimate from estimating equation (14) on a single sample group, while the pooled point and 95% confidence interval (from heteroskedastic robust standard errors) were produced from a stacked regression including sample group fixed effects.

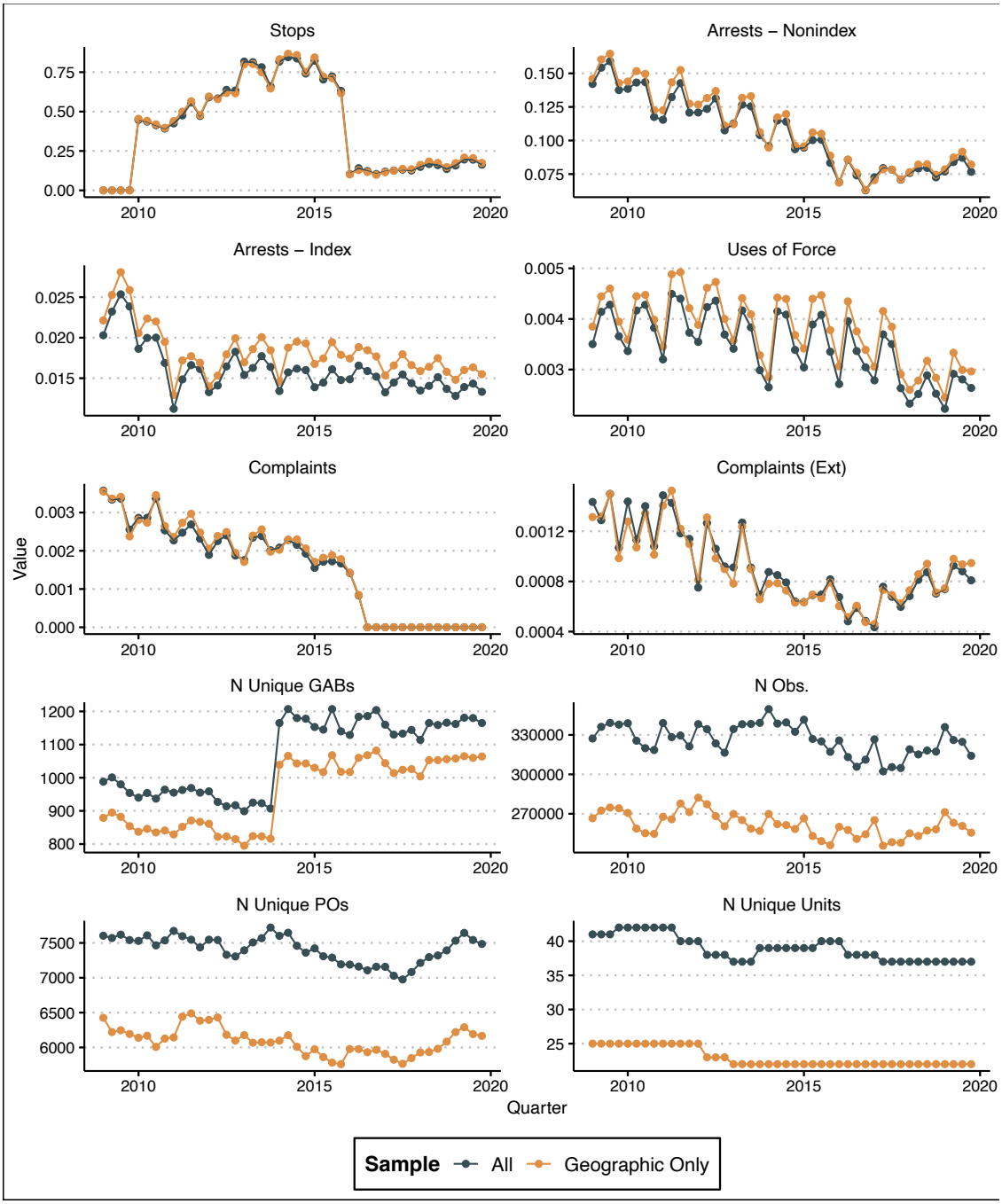
A Appendix A: Additional Figures and Tables

Figure A.1: CPD Operations Calendar (2012)



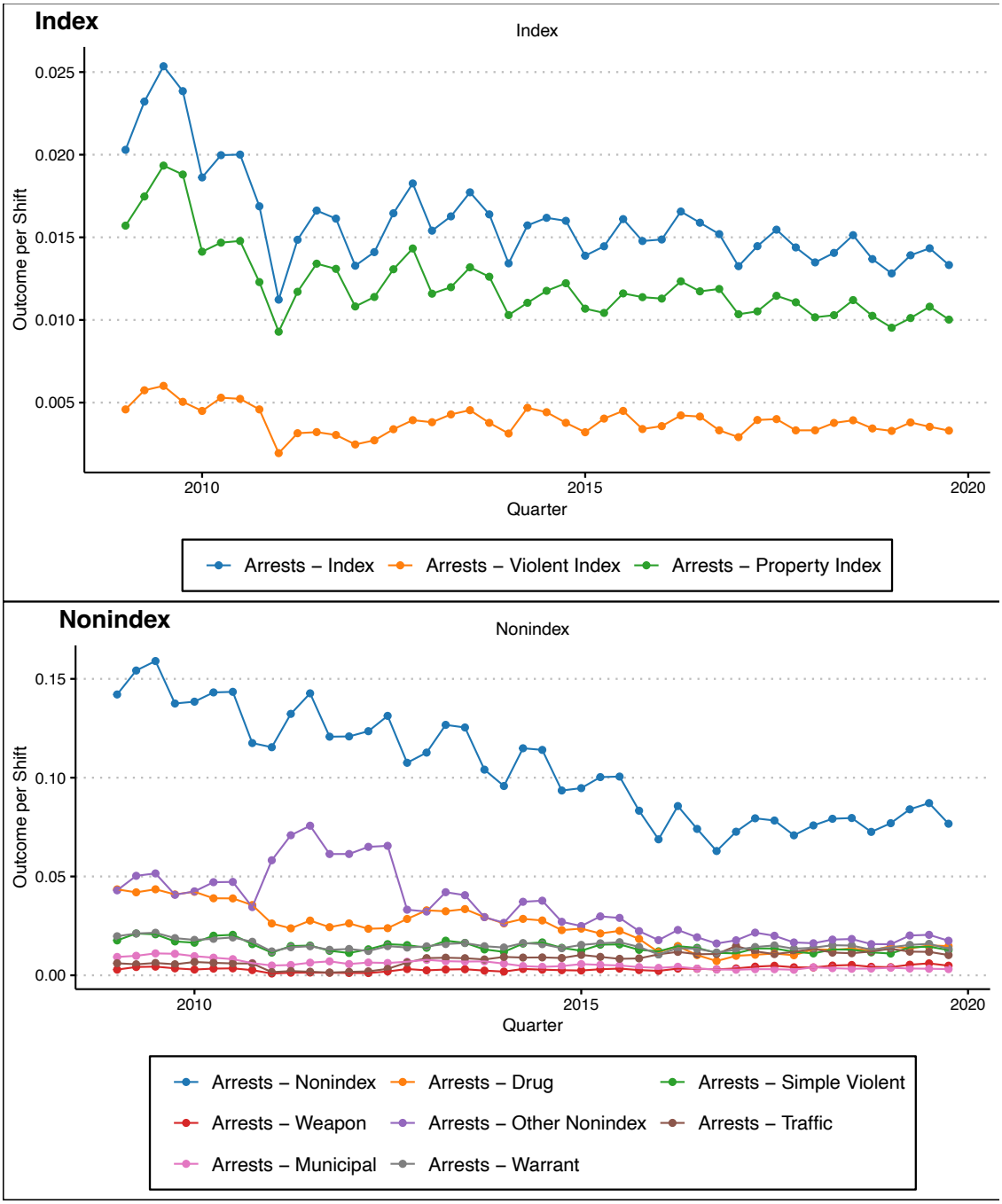
Note: Figure displays an example of the CPD operations calendar for the year 2012.

Figure A.2: Timeseries of Officer-Level Outcomes and Assignments



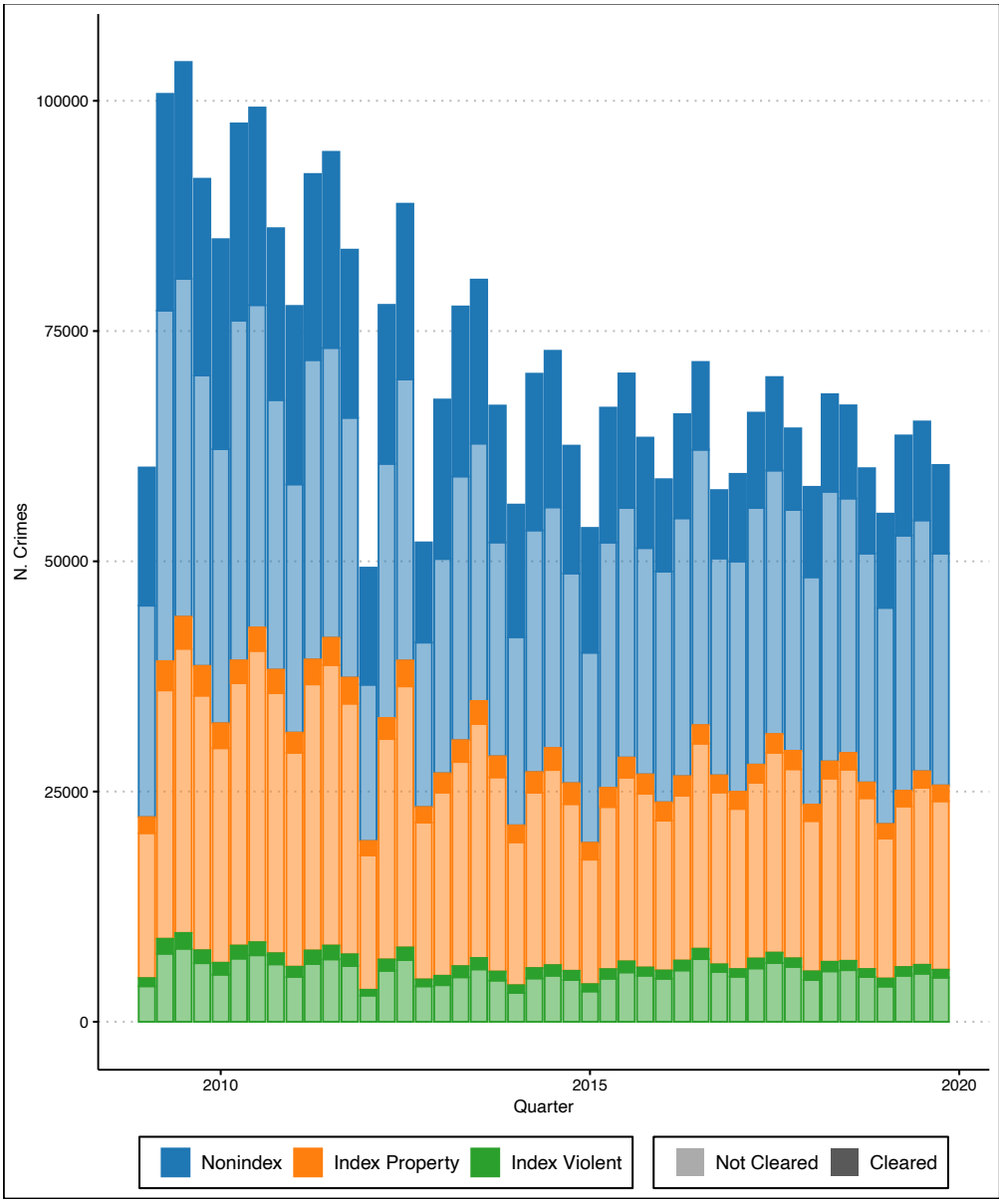
Note: Figure displays the timeseries of outcomes and other statistics at the year-quarter level for officer-day observations for all observations and observations only in geographic units. For outcome variables, the y-axis refers to the mean outcome per 100 shifts. For the count variables (beginning with 'N' referring to the number or count), the y-axis refers to the total count.

Figure A.3: 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 A.4: Timeseries of Crime and Clearances



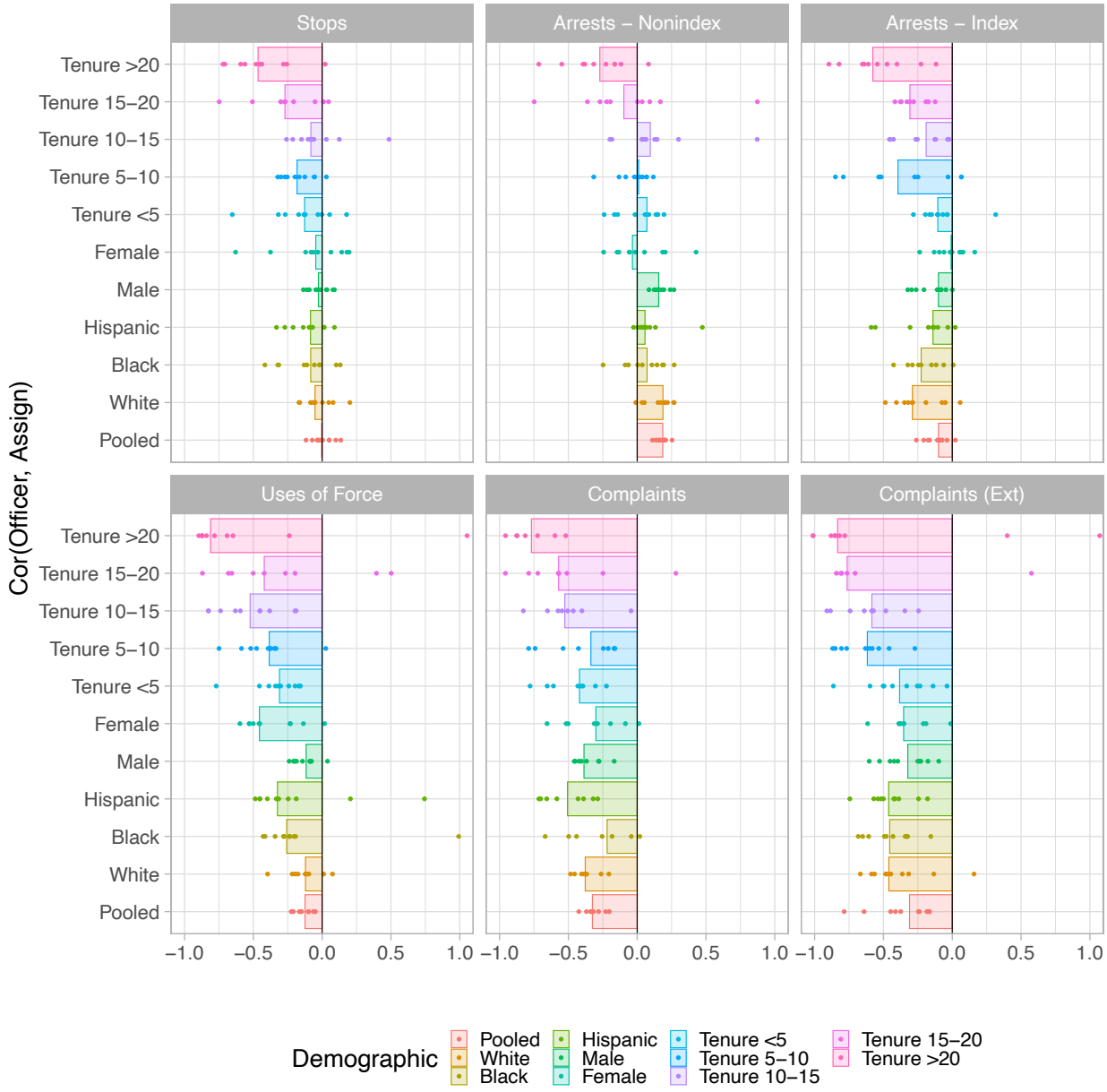
Note: Figure displays the total number of crimes reported in Chicago between 2009 and 2019 by quarter broken into subtypes denoted by color (index violent, index property, and nonindex) and whether or not the crime was reported to have been cleared by an arrest (shaded for cleared, unshaded for not cleared).

Figure A.5: Degree of Officer-Assignment Sorting over time by Specification



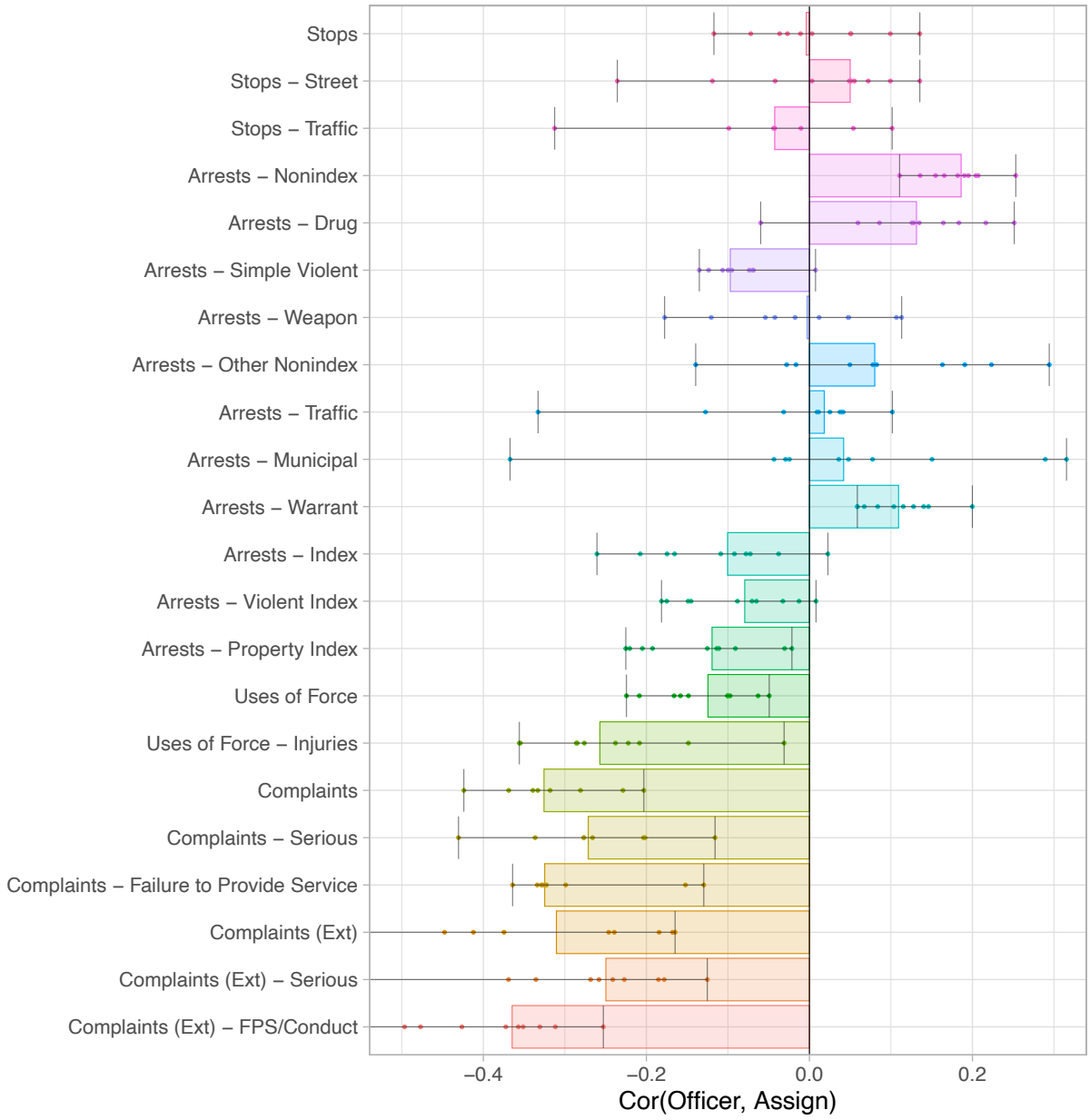
Note: Figure displays the KSS bias corrected correlations between officer and assignment effects, recovered from estimating the different specification and equation (1), with the x-axis denoting the time period of the analysis sample used. Main (2 Year) refers to the main specification with 10 2-year 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; 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.

Figure A.6: Degree of Officer-Assignment Sorting by Demographic Group



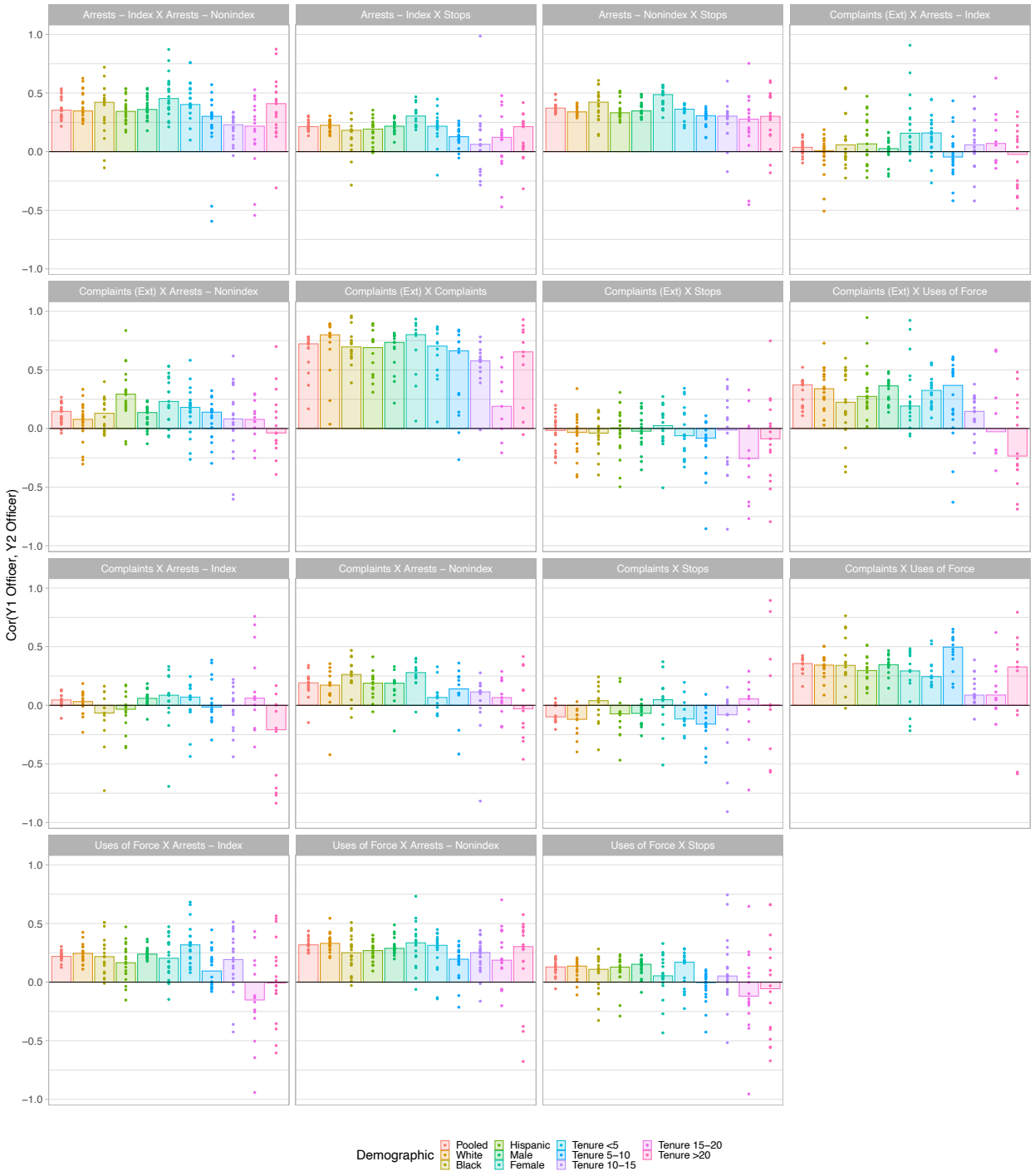
Note: Figure displays the KSS bias corrected correlations between officer and assignment effects, recovered from estimating equation (1) using subsamples of different demographic groups (y-axis), across 10 2-year periods with each point denoting an estimate for a specific time period and the bar denoting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

Figure A.7: Degree of Officer-Assignment Sorting for Additional Outcomes



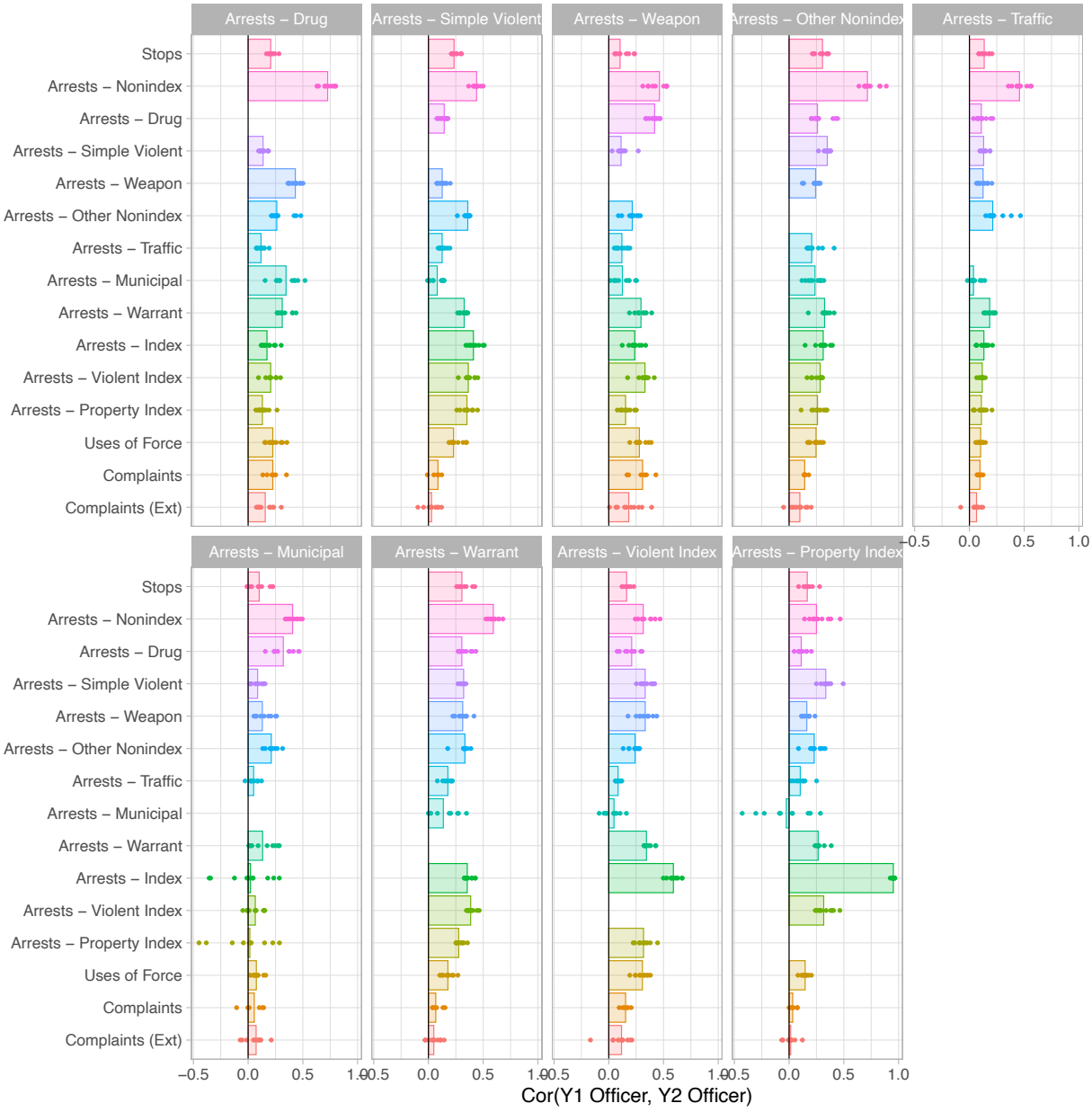
Note: Figure displays the KSS bias corrected correlations between officer and assignment effects, recovered from estimating equation (1) for additional outcomes, across 10 2-year periods with each point denoting an estimate for a specific time period and the bar denoting the median across these estimates.

Figure A.8: Cross-Outcome Correlation in Officer Effects by Demographic Group



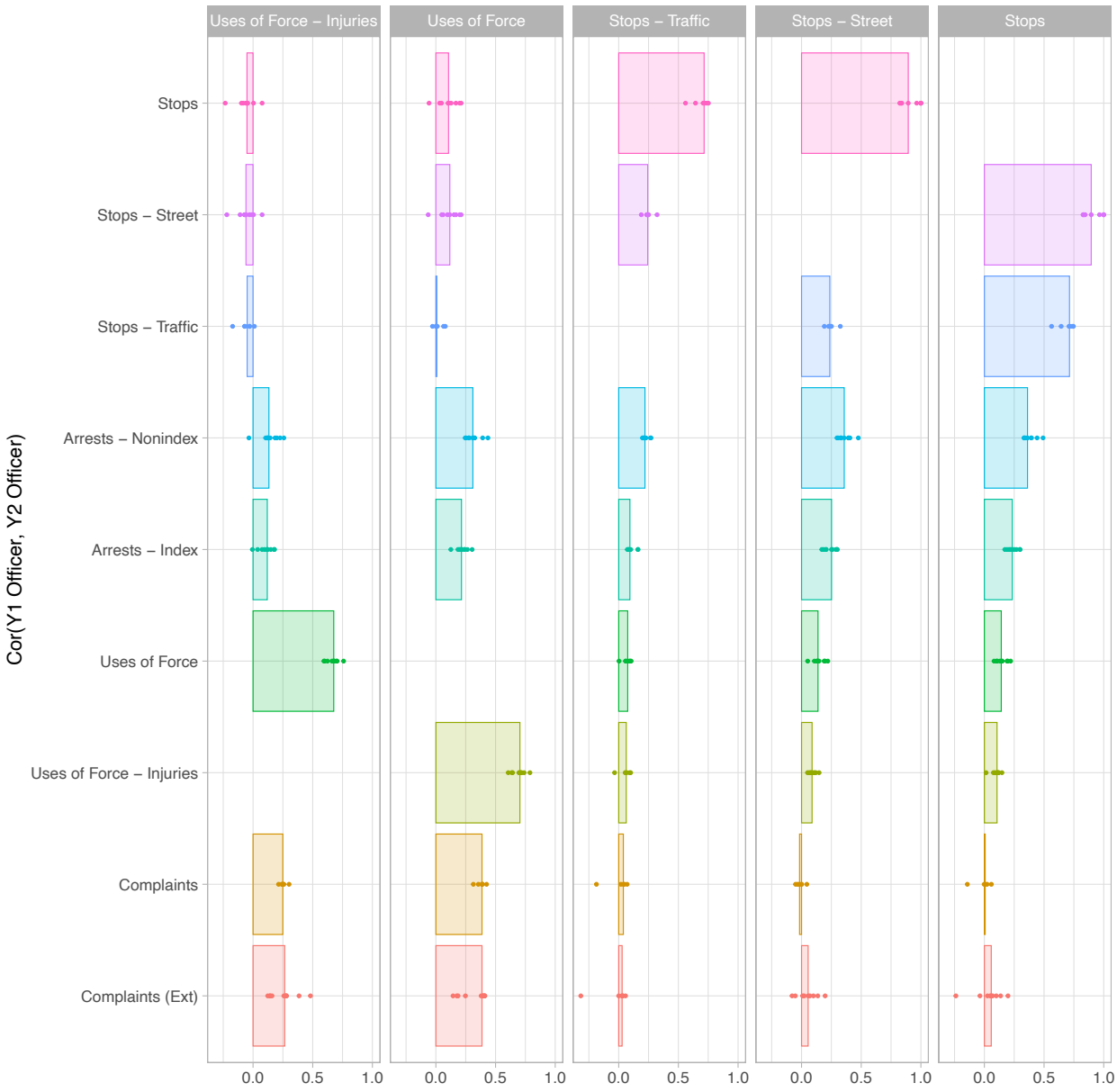
Note: Figure displays the KSS bias corrected correlations between officer effects for two different outcomes for different demographic groups, with officer effects recovered from estimating equation (1) using subsamples of different demographic groups, across 10 2-year periods with each point denoting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

Figure A.9: Cross-Outcome Correlation in Officer Effects for Arrest Sub-Outcomes



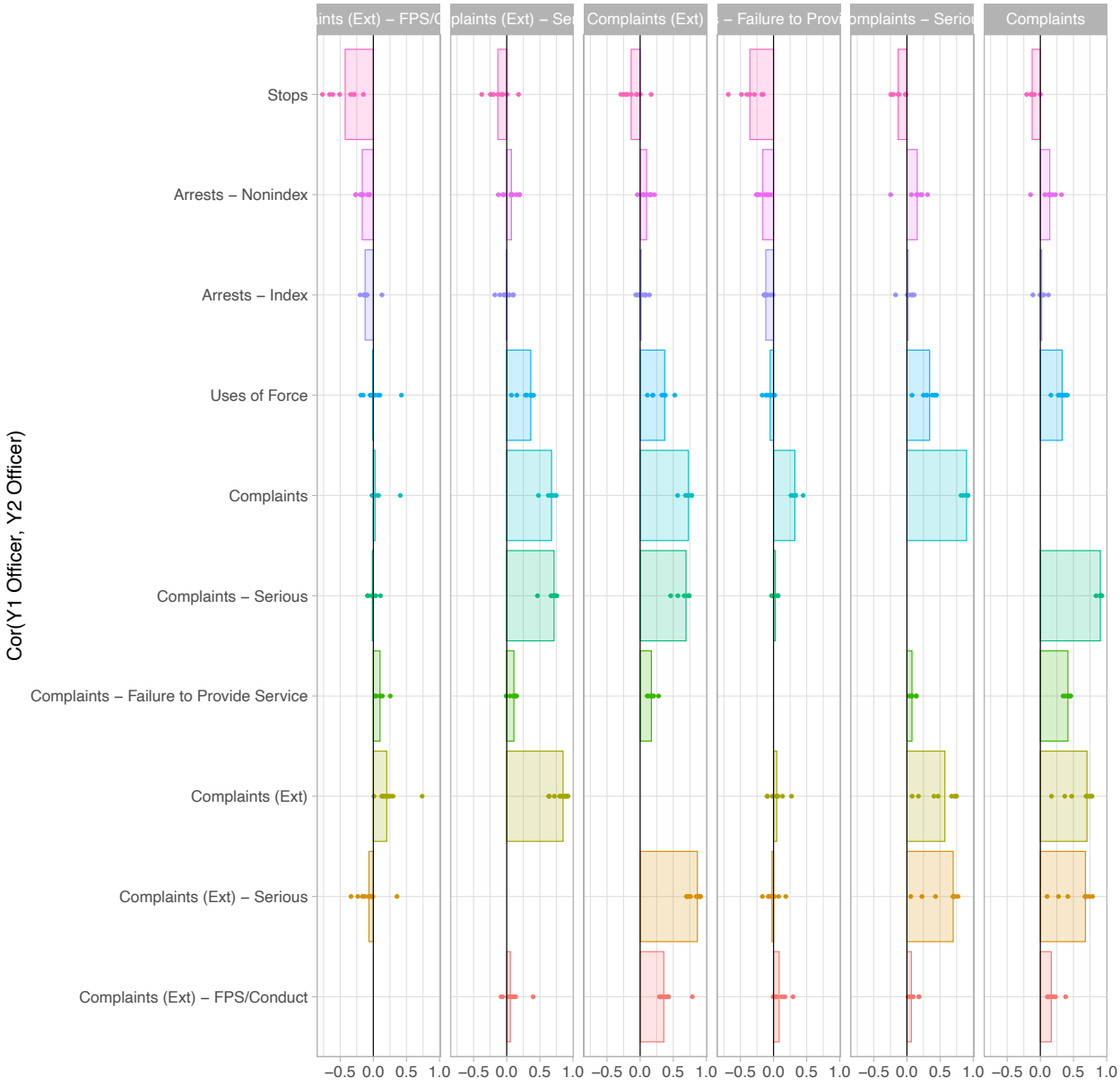
Note: Figure displays the KSS bias corrected correlations between officer effects for two different outcomes focusing on the main outcomes and arrest sub-outcomes (both in y-axis), with officer effects recovered from estimating equation (1) across 10 2-year periods with each point denoting a correlation for a specific time period and the bar denoting the median across these correlations.

Figure A.10: Cross-Outcome Correlation in Officer Effects for Stops and Force Sub-Outcomes



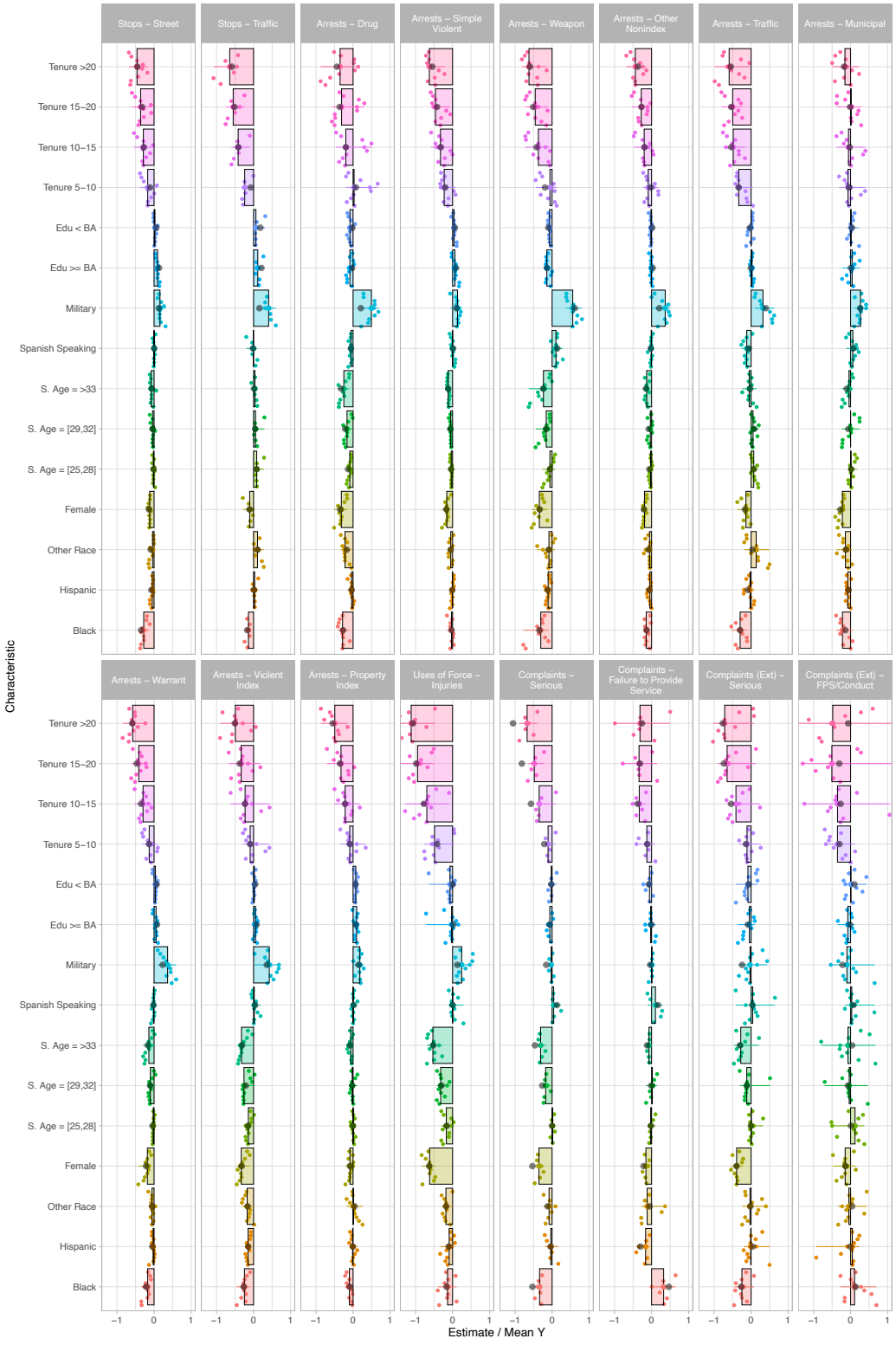
Note: Figure displays the KSS bias corrected correlations between officer effects for two different outcomes focusing on the main outcomes and stop and force sub-outcomes (all in y-axis), with officer effects recovered from estimating equation (1) across 10 2-year periods with each point denoting a correlation for a specific time period and the bar denoting the median across these correlations.

Figure A.11: Cross-Outcome Correlation in Officer Effects for Complaint Sub-Outcomes



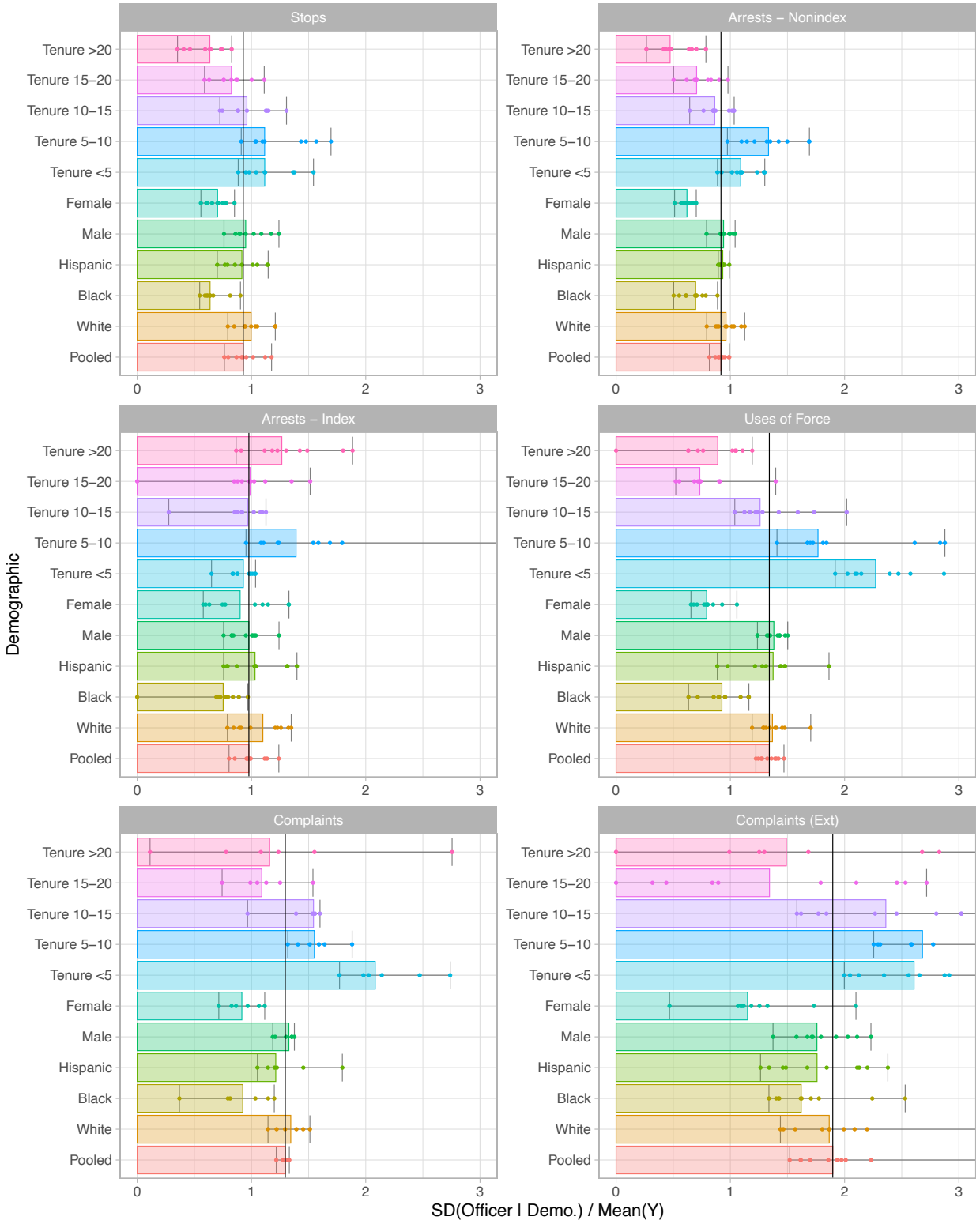
Note: Figure displays the KSS bias corrected correlations between officer effects for two different outcomes focusing on the main outcomes and complaint sub-outcomes (all in y-axis), with officer effects recovered from estimating equation (1) across 10 2-year periods with each point denoting a correlation for a specific time period and the bar denoting the median across these correlations.

Figure A.12: Effect of Officer Characteristics on Officer Effects for Additional Outcomes



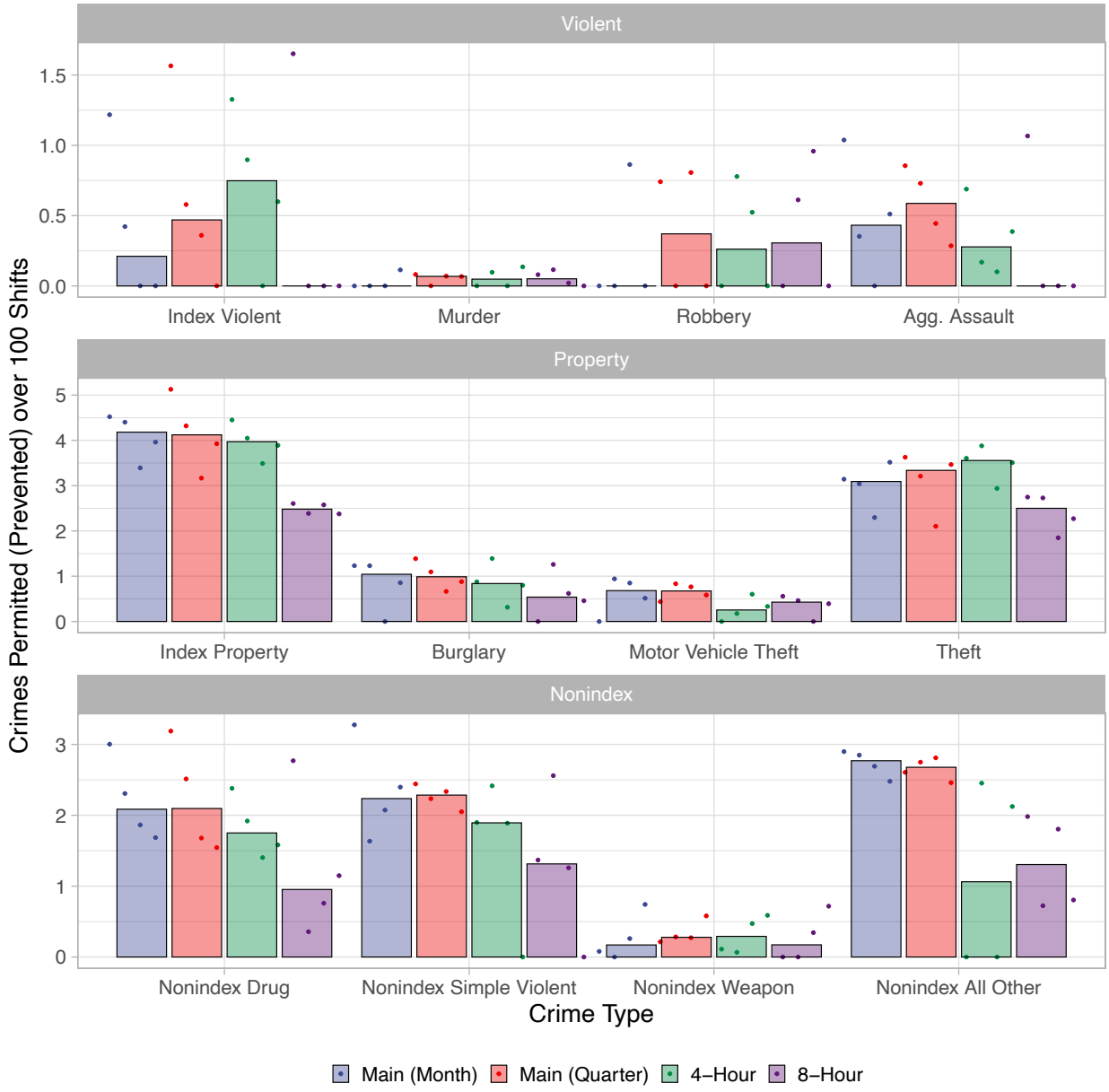
Note: Figure displays the coefficients of the effect of officer characteristics on officer effects (recovered by estimating equation (1)) for additional outcomes, which are the results of estimating equation (4). All estimates are scaled by the mean of the outcome. Each smaller point is an estimate from the 10 2-year period samples, the bars are the median of these estimates, and the lines span the minimum and maximum of the estimates, while the larger points denote the estimate of the stacked regressions using all years.

Figure A.13: Standard Deviations of Officer Effects by Demographics



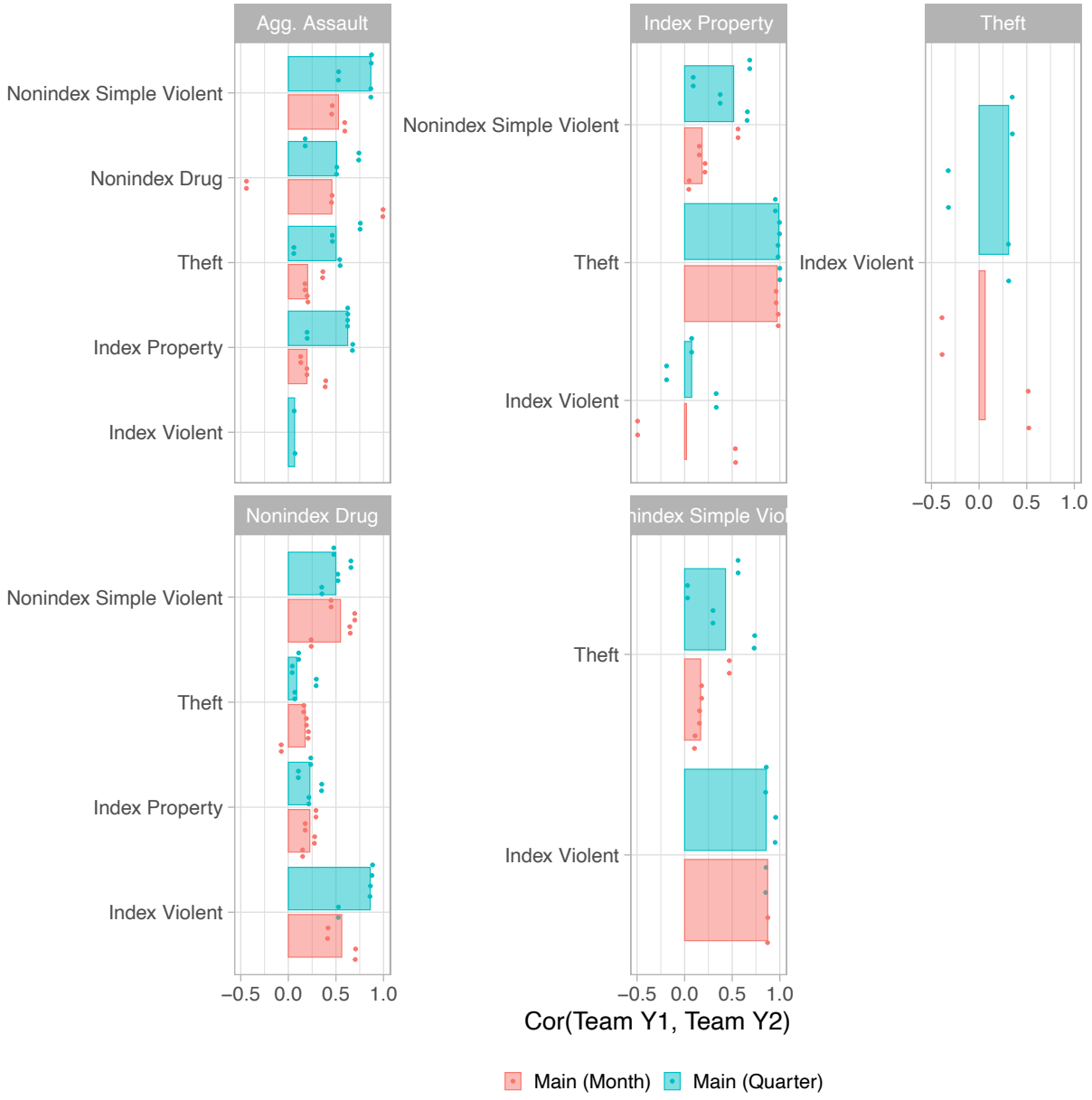
Note: Figure displays the KSS bias corrected standard deviations of officer effects estimated by equation (1) on subsamples of the data which restrict to the relevant demographic group. All estimates are scaled by the mean of the outcome. Each smaller point is an estimate from the 10 2-year period samples, the bars are the median of these estimates, and the lines span the minimum and maximum of the estimates. The vertical lines correspond to the median standard deviation of the officer effects in the pooled sample.

Figure A.14: Approximation of Number of Crimes Permitted (Prevented) by 1 SD Increase (Decrease) in Single Officer Effect over 100 Shifts



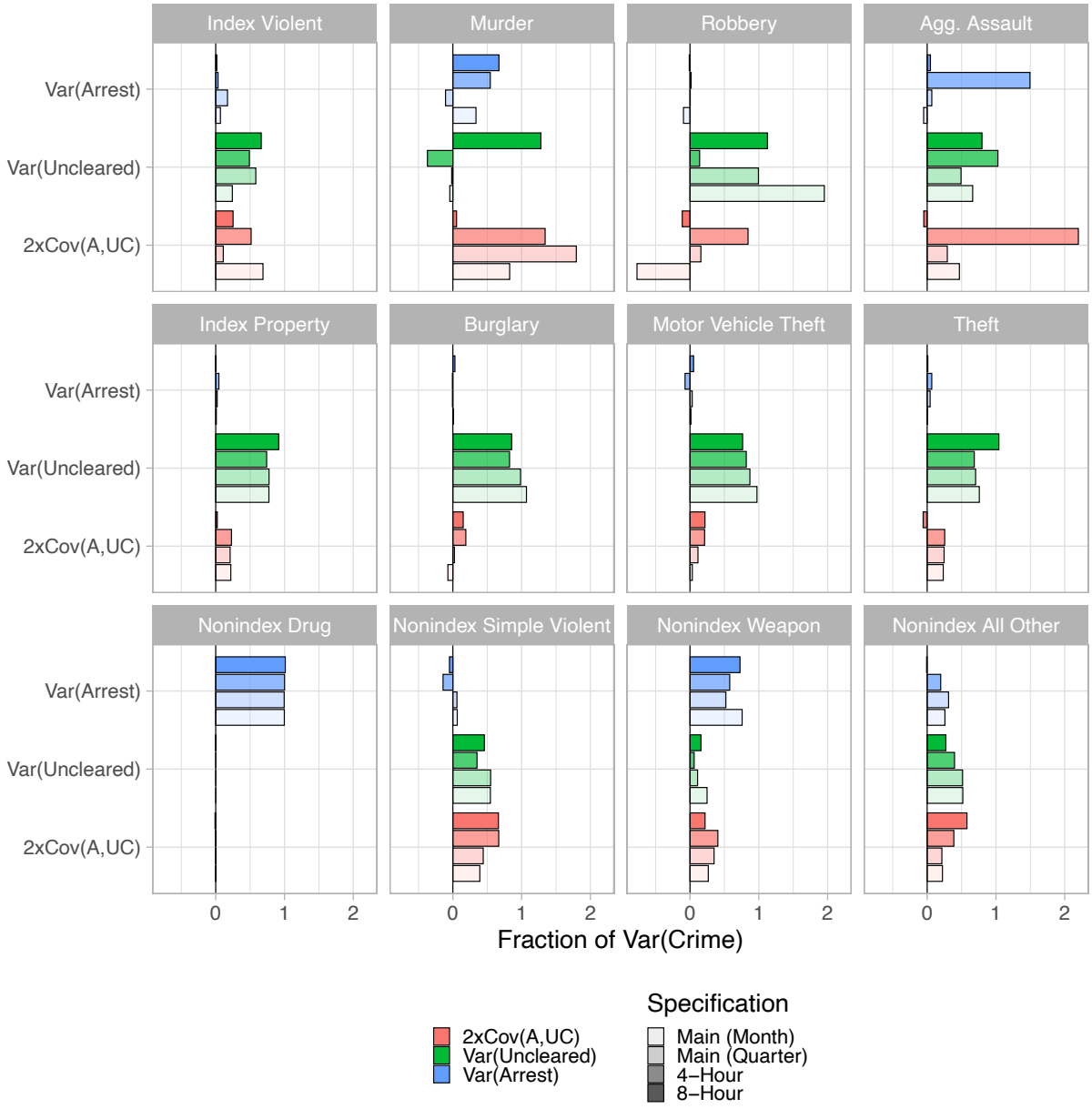
Note: Figure displays the KSS bias corrected standard deviations of team effects (Ω), recovered by estimating equation (6) for each crime type and across different specifications, with each estimate being divided by 59 (the average team size) and multiplied by 100 and by 9 (2 for 4-hour and 1 for 8-hour) to provide an approximation of how many more crimes would be reported (prevented) if one officer's crime effect were increased (decreased) by 1 SD over the course of 100 shifts. 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, while 4-Hour and 8-Hour aggregate the data to 4 and 8-hour periods using the same level of controls interacted with quarter. Each point represents an estimate recovered from one of four sample periods each 2.75 years long, and bars denote the median of these estimates. 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.

Figure A.15: Cross-Crime Correlation in Team Effects



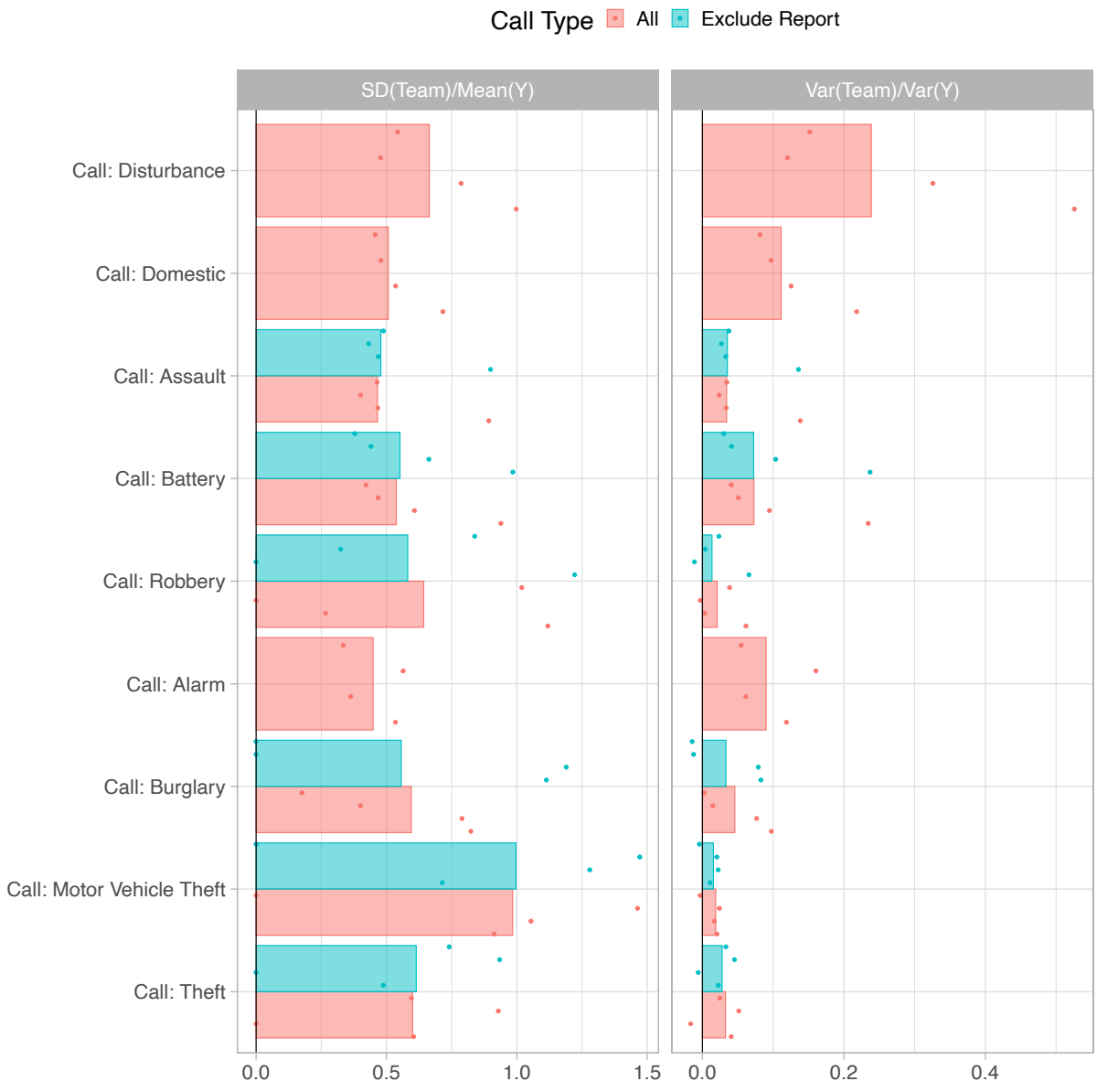
Note: Figure displays the KSS bias corrected correlations between team effects (Ω) on crime for a subset of crime types. Estimates of team effects recovered from estimating equation (6) on four 2.75 year samples, and the bars denoting the median of these estimates. All outcomes used were scaled at per 100,000 population. 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.

Figure A.16: Decomposing Team Effects on Crime into Arrests and Uncleared Crimes



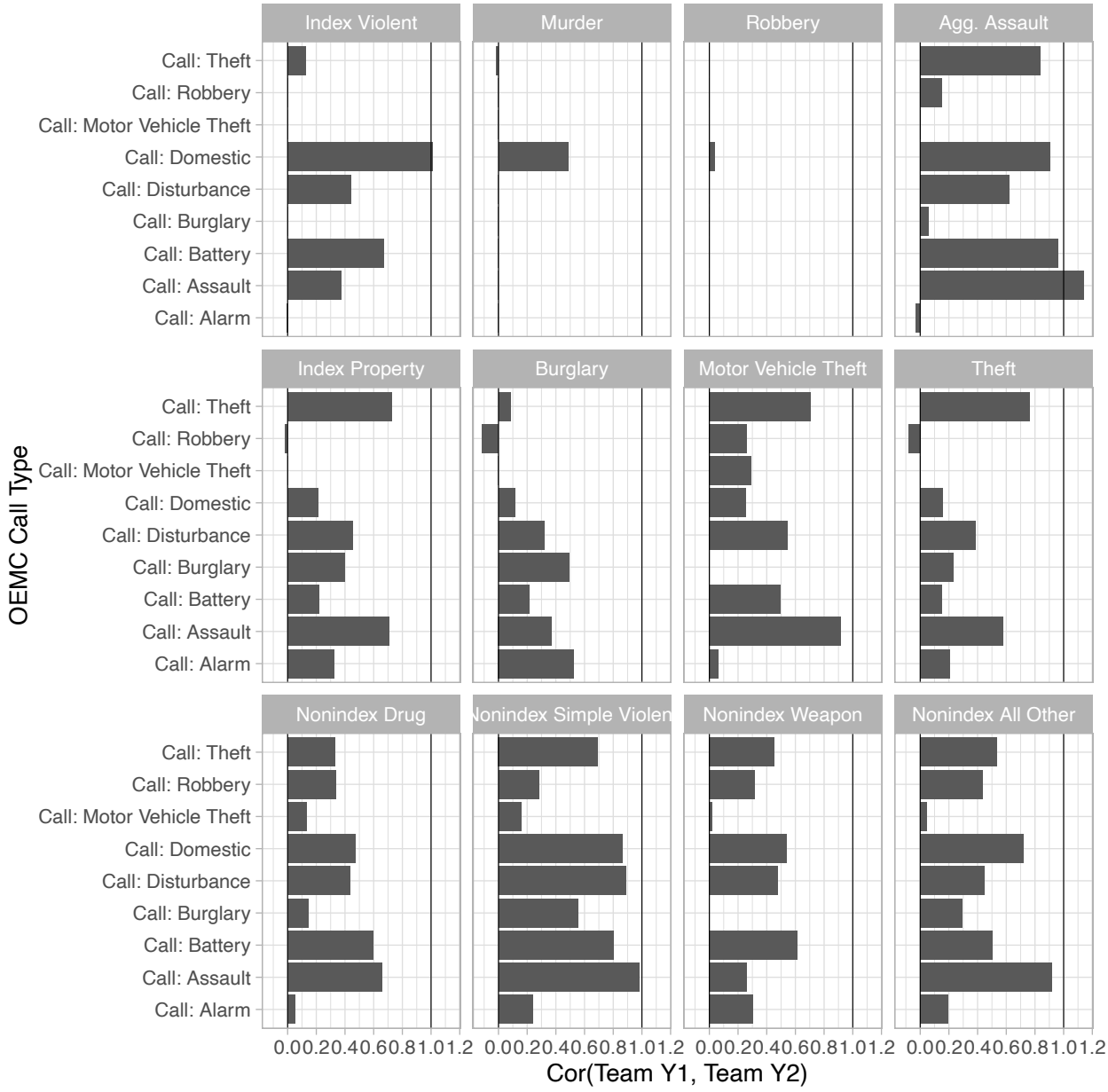
Note: Figure displays the KSS bias corrected variance and covariances for team effects (Ω) for the components of total crime (arrests and uncleared crimes) scaled by the overall variance in team effects on crime. Estimates of team effects recovered from estimating equation (6) on four 2.75 year samples, and the bars denoting the median of these estimates. All outcomes used were scaled at per 100,000 population. 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, while 4-Hour and 8-Hour aggregate the data to 4 and 8-hour periods using the same level of controls interacted with quarter. Variance components can be negative due to bias correction.

Figure A.17: 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 equation (6), and each point corresponds to one of four 2.75 year samples, 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 A.18: Correlation between Team Effects on 911 Calls and Crime



Note: Figure displays KSS bias corrected correlations between team effects (Ω) for crime outcomes and 911 call outcomes (y-axis), with estimates of team effects recovered from estimating equation (6) on four 2.75 year samples, and the bars denoting the median of these estimates. Values can exceed -1 or 1 due to bias correction.

Table A.1: Pruning and KSS Statistics for Crime Effects

Spec Type	Date Controls	Sample Period	Initial Obs	Final Obs	Share Obs Remaining
(1)	(2)	(3)	(4)	(5)	(6)
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	1	577766	577347	0.9993
Officer + Sgt	Year x Quarter x District x Watch x DoW	1	577766	577726	0.9999
Pooled Officer	Year x Month x District x Watch x DoW	1	577766	577726	0.9999
Pooled Officer	Year x Quarter x District x Watch x DoW	1	577766	577726	0.9999
Regular + Tactical Officer	Year x Quarter x District x Watch x DoW	1	577766	576234	0.9973
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2	492900	491650	0.9975
Officer + Sgt	Year x Quarter x District x Watch x DoW	2	492900	492149	0.9985
Pooled Officer	Year x Month x District x Watch x DoW	2	492900	492436	0.9991
Pooled Officer	Year x Quarter x District x Watch x DoW	2	492900	492436	0.9991
Regular + Tactical Officer	Year x Quarter x District x Watch x DoW	2	492900	491419	0.9970
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	3	508397	507628	0.9985
Officer + Sgt	Year x Quarter x District x Watch x DoW	3	508397	508129	0.9995
Pooled Officer	Year x Month x District x Watch x DoW	3	508397	508129	0.9995
Pooled Officer	Year x Quarter x District x Watch x DoW	3	508397	508129	0.9995
Regular + Tactical Officer	Year x Quarter x District x Watch x DoW	3	508397	507713	0.9987
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	4	506415	501596	0.9905
Officer + Sgt	Year x Quarter x District x Watch x DoW	4	506415	501860	0.9910
Pooled Officer	Year x Month x District x Watch x DoW	4	506415	501860	0.9910
Pooled Officer	Year x Quarter x District x Watch x DoW	4	506415	501860	0.9910
Regular + Tactical Officer	Year x Quarter x District x Watch x DoW	4	506415	498134	0.9836

Note: Table displays the results of the Bonhomme (2021) pruning algorithm on the crime data using different date controls: year x district x day of week-8 hour time period (watch) x month or quarter. Observations are at the district-day-hour level.

B Appendix B: Additional Analyses

B.1 Testing for Violations of Internal Validity

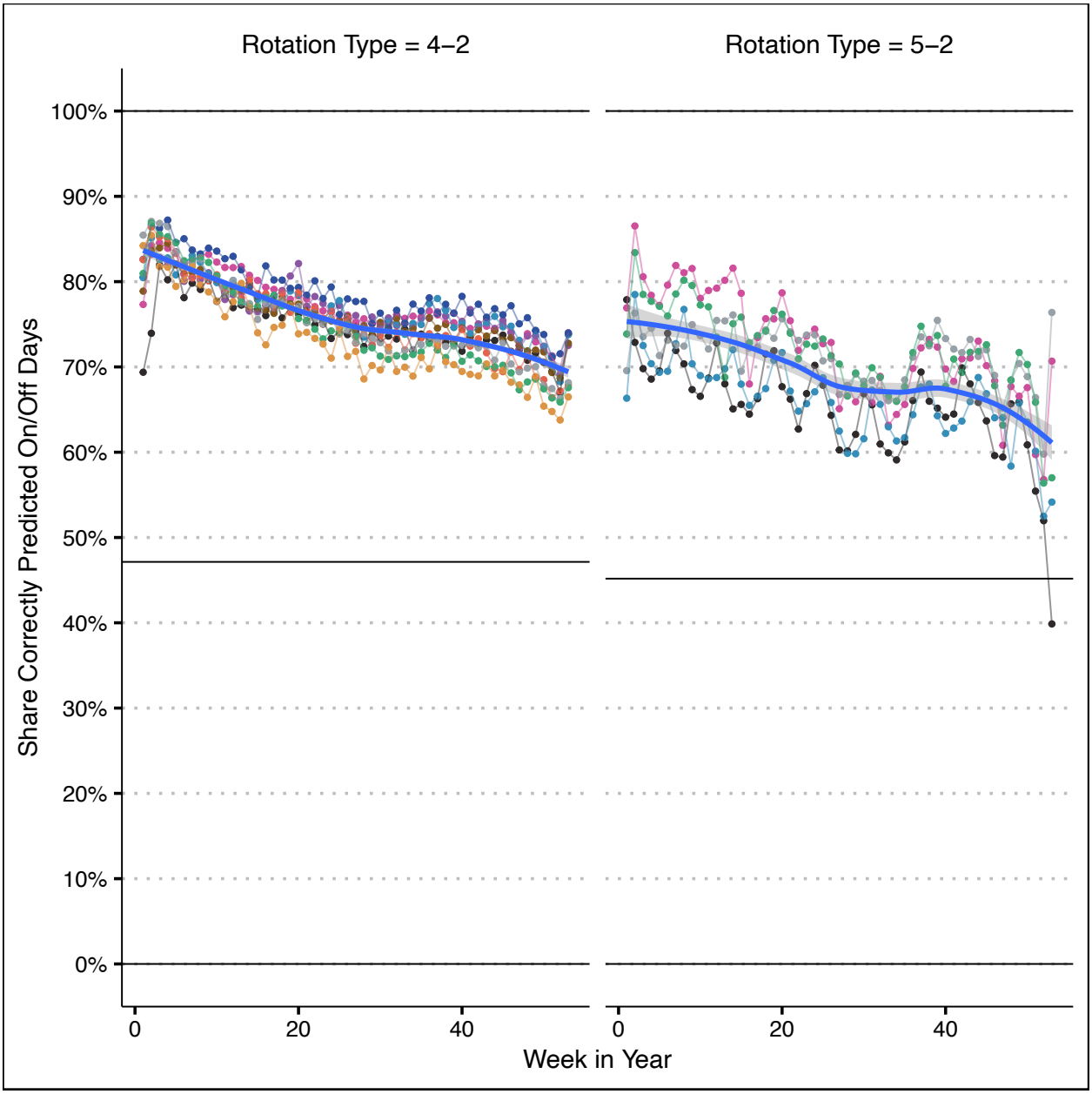
To test for systematic violations of the calendar, we perform the following tests. First, if officers cannot easily manipulate the calendar system, their outcomes should not relate to calendar-related statistics, such as the number of days they work in the year, the share of days they work in their main 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.2, no outcome has a 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. 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, Jordan, and Kim (2023))— notably Gudgeon, Jordan, and Kim (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.

Next we turn to concerns over assignment selection based on endogenous or match-specific factors. While the main models will allow for officers to select into different assignments based on their fixed characteristics, threats to identification include movement between assignments being a result of an exogenous shock in a systematic manner (e.g., a random

positive shock to an officers' complaints leads to them being more likely to move higher complaint assignments) and match-specific effects (e.g., officers selecting specific assignments based on officer-assignment benefits) (Card, Heining, and Kline (2013)). Focusing on the latter issue, Panel D in Figure B.2 displays the relationship between officer outcomes and the likelihood of working in the same GAB in the following shift, with a mean of around 70%. As shown, compared with the overall likelihood of moving GABs, a specific outcome on a given day has virtually zero relationship both with and without unit-year-watch fixed effects.

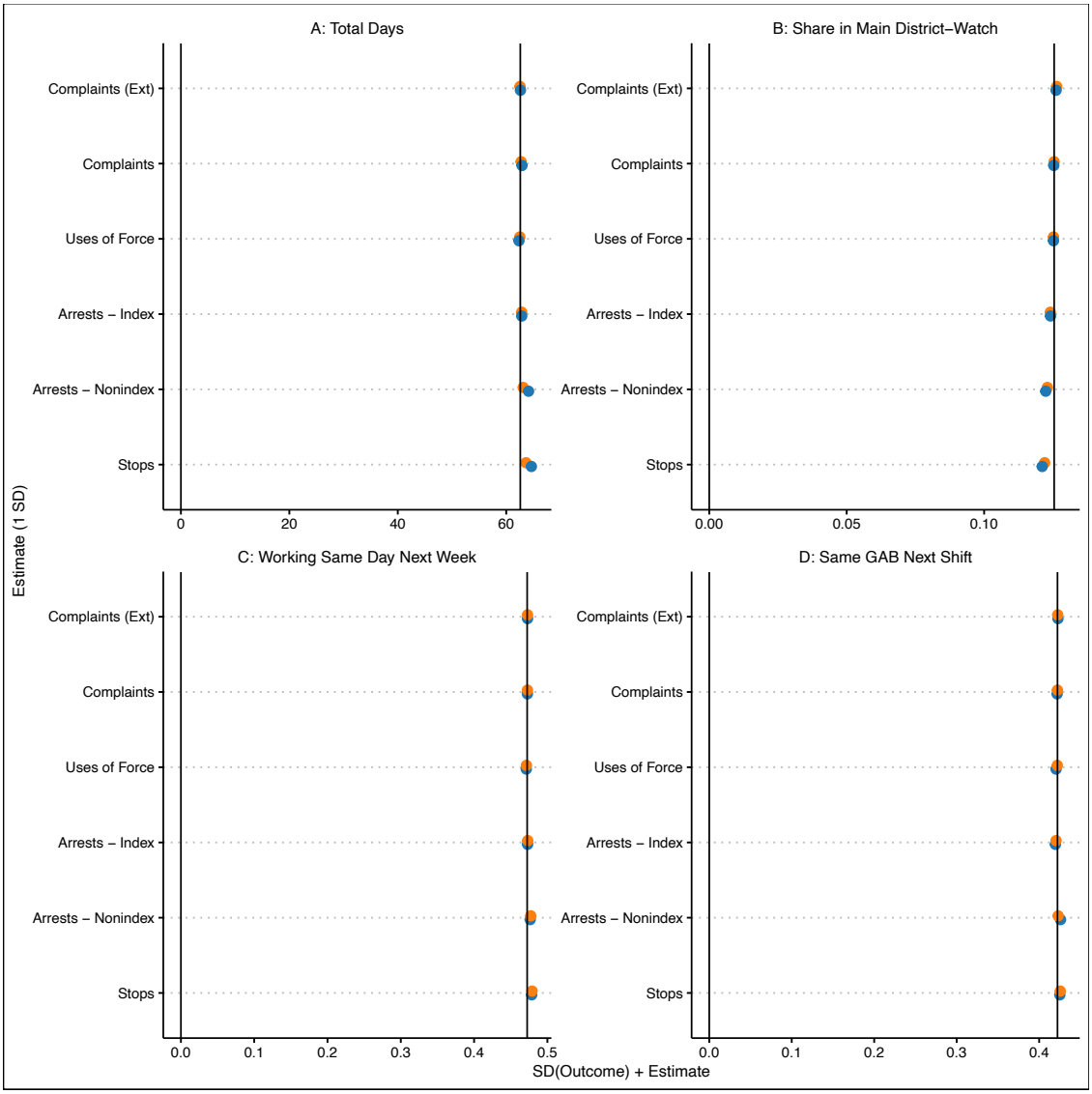
To test for match effects, I adapt the event study plot of Card, Heining, and Kline (2013) to study the differences in outcomes of 'movers' between different types of assignments. Figure B.3 displays this event study for 3 years of the data and for nonindex arrests – a highly discretionary but also frequent outcome. The plot displays the average changes in nonindex arrests per shift experienced by officers moving between GABs as classified by the quantile (4 being the highest) of their nonindex arrests. If match effects are significant, we expect that officers moving from a high arrest assignment to a low arrest assignment should experience the a reduction in arrests as the gain in arrests experienced by an officer moving from a low arrest assignment to a high arrest assignment. As shown in the figure, the reductions of 4-1 officers are similar to the gains of the 1-4 officers. Furthermore, the losses of going 4-2 is larger than the gains of going 1-2, and opposite for 4-3 and 1-3. Second, we expect lateral movements (to different assignments of similar ranks in arrests) should result in no significant changes in average arrests – and again, we see this is the case. Overall, these results suggest that match effects are not of a first order concern.

Figure B.1: Accuracy of Predicted Days On and Off



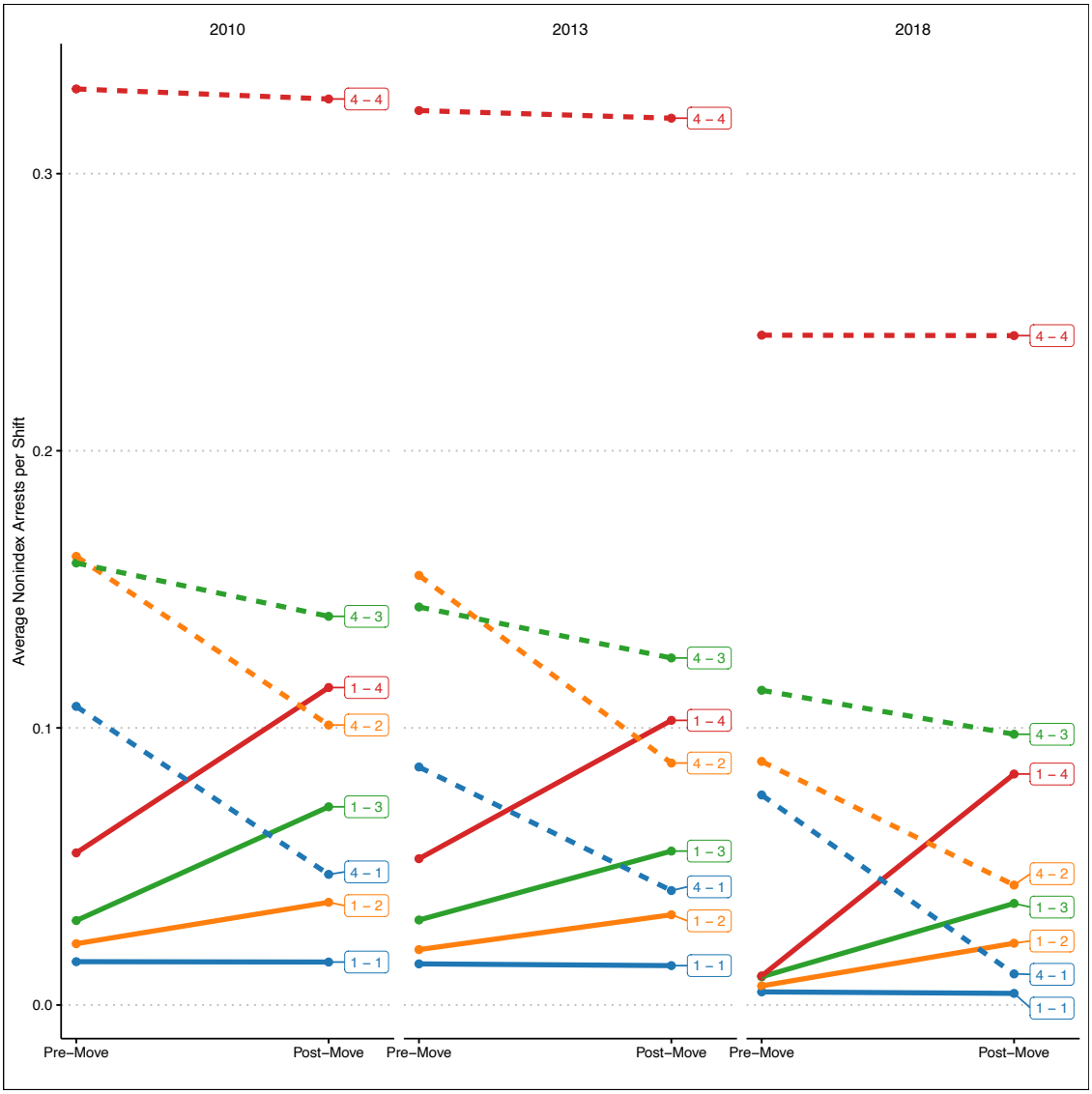
Note: Figure displays the share of correctly predicted days on and days off using based on projecting backward and forward the first 4 days on and 2 days off (4-2) and 5-2 spells, with average correct predictions aggregated at the year-week level and each color denoting a different year of data. The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

Figure B.2: 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 outcome at the over level. Points denote estimates with and without fixed effects and vertical lines denote 1 SD of the outcome.

Figure B.3: Event Study of Changing Between GABs by GAB Quantile of Nonindex Arrests on Officer's Nonindex Arrests



Note: Figure displays an event-study plot of the change in officer's average nonindex arrests among those who switch between assignments for 3 sample years. Assignments are classified into quantiles in terms of their average nonindex arrests per officer per shift, denoted by 1 through 4 (4 being the highest), and each line denotes a movement from either a 1st quantile or 4th quantile assignment to a 1st, 2nd, 3rd, or 4th quantile assignment.