The Boss in Blue: Supervisors and Police Behavior

Austin V. Smith*

January 21, 2025

Please click here for most recent version

Abstract

This paper studies the role of first-line police supervisors — sergeants — in the enforcement decisions of their officers. I leverage a unique institutional setting where officers switch sergeants frequently to estimate sergeant arrest effects, and document substantial variation in these effects between sergeants. Moving an officer from a sergeant in the 10th percentile of arrest effects to one in the 90th percentile would increase monthly arrests by 42% relative to the mean. I provide evidence that sergeants induce arrests for serious and low-level crimes through distinct policing strategies. Sergeants increase serious arrests by incentivizing their officers to respond to more 911 calls, while they increase low-level arrests through discretionary drug enforcement. Sergeant-induced low-level arrests disproportionately affect Black civilians and increase officer use of force. Connecting these estimates to pre-promotion characteristics, I find that sergeants who scored the lowest on their promotional exams are over-represented among those who encourage low-level arrests. My findings suggest that sergeant-focused policies may be particularly effective at reducing aggressive policing tactics without harming the enforcement of serious crimes.

*University of Arizona. email: austinvsmith@arizona.edu. I am deeply grateful to my committee members, Dan Herbst, Evan Taylor, Ashley Langer, and Juan Pantano, for their invaluable guidance, support, and mentorship. I would also like to thank Jesse Bruhn, Mallory Cash, Spencer Cooper, Price Fishback, Lance Gui, Ami Ichikawa, Andrew Johnston, Cynthia Lum, Daniel Nagin, Nayoung Rim, Matthew Ross, CarlyWill Sloan, Wint Thu, Mo Xiao, and seminar participants at the University of Arizona and SEA Annual Meeting for their insightful feedback. This research benefited greatly from conversations with Major Stephen Bishopp of the Dallas Police Department as well as numerous members of the DPD Public Information Office, whose expertise was essential to understanding many crucial institutional details. All remaining errors are my own.

1 Introduction

There is mounting evidence that aggressive policing tactics — such as the over-enforcement of low-level crimes and excessive use of force — erode public trust (Ang et al., 2024), undermine the mental and economic well-being of civilians (Geller et al., 2014; Ang, 2020; Mello, 2021), and result in legal sanctions that perpetuate cycles of criminality (Agan et al., 2023). This evidence has heightened demands for police reform that emerged following a decade marked by high-profile incidents of force used against minority civilians. Many of the most frequently proposed reforms focus on front-line officers, advocating for measures like increasing minority recruitment (Ba et al., 2021b) or improving officer training (Dube et al., 2023). However, far less attention has been given to policies targeting police management — in particular, sergeants, the immediate supervisors of front-line officers.

This oversight is striking, given economic evidence from other industries that highlights the substantial influence of managers on employee behavior (see Roberts and Shaw (2022) for a review). Managerial interventions could thus be particularly effective in policing, since sergeants oversee many officers simultaneously. Yet the unique nature of law enforcement—where officers exercise significant discretion with limited oversight—makes it unclear whether findings from other industries apply. Crucially, even if sergeants play a significant role, it is still not clear whether their supervision distinguishes between socially beneficial and harmful officer activities. In other words, can sergeants reduce aggressive low-level enforcement without harming their officers' motivation to enforce serious crimes?

In this paper, I provide novel evidence that sergeants influence the arrest outcomes of their subordinate officers. Such evidence has proven elusive because it requires disentangling the effects of sergeants from the discretion of their officers. I circumvent this obstacle using detailed data on patrol officers in the Dallas Police Department (DPD), who switch sergeants frequently throughout their careers. Officers cannot control the timing of switches, which are determined by vacancies and chief-initiated schedule realignments. With these switching events, I identify a 'sergeant effect' for each of the 347 sergeants in my data using the average change in arrests for officers who switch to and from each sergeant. I estimate sergeant effects net of location and shift characteristics using a two-way fixed effects framework (Abowd et al., 1999), and correct for sampling error using a standard empirical Bayes shrinkage procedure (Chetty et al., 2014). I show that sergeant effects are not driven by trends in officer behavior or location-specific crime rates, concurrent policy changes, or match quality, all of which would bias my estimates.

First, I demonstrate that sergeants have a substantial effect on the quantity of arrests made by their officers. Moving an officer from a sergeant in the 10th percentile of the sergeant effects distribution to one in the 90th percentile results in 1.6 additional arrests per month, representing a 42% increase relative to the mean. I estimate that sergeants account for 3.4% of the total variation in officer arrests. However, I show that sergeant variation is more impactful for *overall* arrests than officer variation, since sergeants manage an average of 6.33 officers at a time, which multiplies their effects.[1] Using an event-study design around switching events, I show that the effects of moving to a new sergeant are immediate and persistent. Combined, these results suggest that sergeant-focused policies can be more powerful and longer-lasting than officer-focused interventions. This is especially true for officer training programs, which have shown promise but tend to atrophy in their effects over time (e.g. Owens et al., 2018; Mello et al., 2023).

Next, I evaluate sergeant effects separately on serious and low-level arrests. Serious arrests consist of apprehensions for violent and property crimes, whereas low-level arrests are made primarily for victimless, quality-of-life offenses such as drug possession, disorderly conduct, and outstanding warrants. Given the diminished severity of the associated crimes, marginal low-level arrests may not produce public safety benefits that outweigh their costs to arrested individuals and their communities (Cho et al., 2023). Strikingly, I fail to find a strong relationship between serious and low-level sergeant effects.[2] This is especially true for the top half of the low-level effects distribution: for these sergeants, the correlation between low-level and serious effects is -0.02, and I can rule out a correlation larger than 0.13 and smaller than -0.16. To the extent that policies might try to reduce low-level arrests through sergeants, my findings suggest that these reductions are not likely to come at the expense of serious crime enforcement.

My results suggest that sergeants induce low-level and serious arrests through distinct channels of officer behavior. I investigate these behavioral channels directly by estimating how officer actions change when low-level and serious sergeant effects increase independently. I first show that low-level sergeant effects operate predominantly through drug arrests. When low-level effects increase by 1 standard deviation, a sergeant's officers make 54% more drug arrests relative to the mean. Ninety percent of the increased drug arrests are for simple possession, and I show that they disproportionately impact Black civilians. Consistent with such arrests being highly discretionary, I find that officer-initiated interactions are responsible for over half of sergeant-induced changes in low-level arrests. These findings suggest that some sergeants incentivize their officers to engage in "broken windows" policing strategies that target crimes indicative of broader social ills, such as addiction and poverty (Zhao et al., 2003). This is consistent with ethnographic evidence that documents sergeants explicitly asking their officers to make arrests for low-level crimes (Van Maanen, 1984). I find no evidence that civilian-reported crimes decrease when a sergeant who values low-level enforcement is assigned to an area, suggesting that their policing strategies do not improve public safety broadly. Moreover, I show that sergeant-induced disorder policing significantly amplifies officer use of force: a one standard deviation increase in low-level sergeant effects leads

[1]I arrive at this conclusion by estimating the impact of replacing sergeants at each percentile of the sergeant effects distribution with a fiftieth percentile sergeant for one month. Assuming each sergeant manages the average number of officers, I show that 90% of such replacements would lead to a larger (in magnitude) change in arrests than would replacing an officer at the same point in the distribution with a fiftieth percentile officer.

[2]The raw correlation between serious and low-level effects is small (0.11), but statistically significant at the 5% level. However, I find evidence that the true relationship between serious and low-level effects is non-linear, and the positive raw correlation estimate is driven by the bottom 5% of sergeants in the low-level effects distribution.

to a 15% increase in use of force incidents relative to the mean.

In contrast, I find that sergeants induce serious arrests through their officers' 911 response efforts. A one standard deviation increase in the serious sergeant effect leads officers to answer 3.6% more 911 calls per month relative to the mean. The calls that officers respond to are more severe than average; however there is no evidence that they are more likely to arrest conditional on the features of the call. These results suggest that serious sergeant effects make officers more active in their call-response duties without increasing their aggression. On the other hand, low-level sergeant effects make officers slightly more active in responding to calls, but significantly more likely to arrest conditional on call features. These changes are most pronounced for the least severe calls, such as mischief and vandalism. For serious sergeant effects, a greater focus on call activity translates into increased arrests for crimes that directly harm others, namely domestic violence, theft, and DWI. Because officers are involved in more civilian interactions, serious sergeant effects also increase use of force, but by far less than low-level sergeant effects, further reinforcing the notion that sergeant-induced low-level policing strategies lead to excessively violent police encounters.

Sergeant effects are strikingly large considering their management limitations. Most police interactions lack immediate supervision, and sergeants have fewer opportunities to monitor officers compared to supervisors in firms. I examine two potential mechanisms to explain sergeant influence on officer behavior. First, I ask whether sergeants "lead by example" by demonstrating their desired field activities to officers. Second, I examine variation in the direct monitoring of officers using a sergeant's presence at their own officers' assignments. I find evidence that sergeants who value low-level enforcement lead by example and are more likely to directly monitor their officers. A one standard deviation increase in low-level sergeant effects is associated with 0.078 more *sergeant* arrests per month (mean = 0.32), all of which are low-level, and 0.60 more calls answered with subordinate officers (mean = 7.87). However, serious sergeant effects cannot be attributed to either of these management mechanisms, as point estimates are economically small and statistically insignificant. Serious sergeant effects thus operate through mechanisms that cannot be observed in the data. Sergeants have a number of administrative duties that could be used to incentivize officers, including transfer recommendations, award nominations, and overtime approval. While the data limit my ability to comprehensively investigate these causal pathways, I find evidence that serious sergeant effects are associated with more calls and arrests outside an officer's regular shift hours, suggesting these sergeants are more likely to approve overtime for officers to respond to 911 calls.

I conclude by considering whether sergeant effects can be predicted using the information available to departments *before* someone is promoted to the position. Such predictions may be used to inform sergeant selection mechanisms or target training for newly promoted sergeants. I leverage multiple sources of detailed DPD personnel records, including newly obtained data on exams that determine promotion to sergeant. Such exams are ubiquitous across police agencies and are used to determine the order in which officers are promoted.

I find no evidence of sergeant differences across race, gender, or age at the time of promotion. However, the sergeant effect distributions differ significantly by exam performance. For serious effects, I find that above-average scorers have a *wider* distribution compared to below-average scorers (Kolmogorov-Smirnov p-value = 0.013), suggesting that top promotion candidates have more heterogeneous management styles — at least in terms of serious enforcement — than those at the bottom of the promotion list. Since serious effects are driven by call-response efforts, these findings indicate that high-scorers vary in their willingness to motivate officers in this dimension. On the other hand, the low-level effect distribution for high scorers is shifted left relative to that of low scorers (Kolmogorov-Smirnov p-value = 0.034), suggesting that sergeants who value low-level arrests tend to be the lowest-ranked promotion candidates. Insofar as policymakers want to reduce unnecessary low-level enforcement, my results indicate that policies that change the low-level enforcement preferences of marginally promoted sergeants — such as training targeted at the bottom of the promotion list or encouraging more officers to attempt the promotional exam — may be particularly effective.

This paper contributes to two strands of literature. The first strand studies how the incentive structures of police organizations contribute to enforcement outcomes (Owens and Ba, 2021). Studies have shown that arrests and use of force respond to union wage negotiations (Mas, 2006), local fiscal conditions (Makowsky and Stratmann, 2009), public access to complaint records (Rivera and Ba, 2022), field training officers (Adger et al., 2022), and police academy peers (Rivera, 2022). My paper is the first to demonstrate that sergeants affect these outcomes and are thus a crucial source of incentives within police departments. These findings speak more specifically to a burgeoning subsection of the literature that studies police management. By showing that first-line supervisors affect how officers distribute their enforcement efforts between serious and low-level crimes, I highlight a crucial distinction between the lowest levels of management and police executives, who have been the primary focus of this literature and are generally responsible for overarching tactical strategies, such as citywide stop and frisk policies or manpower location decisions (e.g. Mummolo, 2018; Bacher-Hicks and De La Campa, 2020; Kapustin et al., 2022). My study is most closely related to recent papers by Frake and Harmon (2023) and Gudgeon et al. (2023). These studies leverage clever natural experiments to show that enforcement outcomes are influenced by the prior misconduct exposure and race of first-line supervisors, respectively.[3] In contrast to these papers, which primarily study the causal pathways of supervisor effects, my findings demonstrate the full magni-

[3] Gudgeon et al. (2023) study lieutenants, who in their setting manage sergeants and directly influence officer behavior through the approval of arrests. One way to rationalize their findings with the lack of significant racial differences between sergeants in my study is that their setting, the Chicago Police Department, uses a unique merit-based promotion system that allows up to 30% of promotions to be based on recommendations rather than test scores (Chicago Police Department, 2024). This system was put in place for the explicit purpose of improving promotion chances for minority officers (Charles, 2021), so it is possible that these merit promotions are uniquely effective at selecting minority supervisors who avoid low-level enforcement.

tude of heterogeneity in supervisory preferences for their officers' enforcement activities. While Frake and Harmon (2023) and Gudgeon et al. (2023) show that observable supervisor features change their enforcement preferences, my findings suggest that these preferences are driven in large part by unobserved tastes that determine a sergeant's preferred policing strategies. Despite the differences between our studies, a shared conclusion is that first-line supervision crucially shapes police outcomes.[4]

Second, I contribute to work within labor economics on the importance of managers (Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Bloom et al., 2013; Lazear et al., 2015; Giorcelli, 2019; Adhvaryu et al., 2023). My primary contribution is showing that managers' subjective preferences can change employee behavior in work environments characterized by a high degree of employee discretion. While previous studies focus on unambiguous firm objectives, such as profits or productivity, I show that in cases where an organization's goals are not clearly defined, the preferences of management can fill this gap. I believe these findings generalize to other important settings where employee behaviors do not cleanly map into organizational objectives, such as teaching[5], medical residencies, or child protective services. Within this literature, my findings also contribute to more recent studies demonstrating the importance of managers for the functioning of public sector organizations (Bloom et al., 2015; Rasul and Rogger, 2018; Fenizia, 2022). By providing estimates of manager effects for street-level bureaucrats whose actions can impose substantial economic and personal costs to the civilians with whom they interact, I highlight an important setting in which public sector managers can directly affect the well-being of their constituents. In particular, my findings suggest that police supervisors who prioritize combating low-level crimes can have a substantial effect on both overall welfare and the distribution of policy enforcement across demographic groups.

The rest of the paper proceeds as follows. Section 2 describes the job functions of a police sergeant and how sergeant assignments are made within the DPD; Section 3 introduces the data; Section 4 describes the empirical strategy; Section 5 presents results and mechanisms; Section 6 uses pre-promotion observables to predict sergeant effects; and Section 7 concludes.

[4]My findings also contribute to a long-standing debate within criminology about the ability of sergeants to shape police behavior (Van Maanen, 1984; Brown, 1988). A substantial body of observational fieldwork has documented correlations between supervisor behaviors and officer decisions (Engel, 2000, 2001, 2002; Engel and Worden, 2003; Johnson, 2011, 2015a,b; Ingram et al., 2014). My paper is one of the first to establish a causal connection between sergeants and the behaviors of the officers they manage.

[5]While much of the teaching literature has focused on test score value-added measures of teacher productivity, recent work by Rose et al. (2022) demonstrates that teachers can also have large effects on their students' future criminality, and these effects are orthogonal to traditional value-added measures. Insofar as my findings generalize to this setting, managers in this context (i.e. principals) may be able to encourage teachers to prioritize different types of student engagement (focused on boosting test scores or reducing delinquent behavior) depending on their personal priorities for the school.

2 The Role of Police Sergeants

Sergeants are the first level of management within policing. Within each branch of a police department, officers are divided into units, each of which is led by a sergeant. In the patrol section, which is responsible for general crime control and 911 response activities, these units are divided according to location and time of day. I study patrol sergeants in the context of the Dallas Police Department (DPD), which assigns sergeants to one of 35 sectors within the city. There is at least one sergeant assigned to each sector on each of the three watches, or shifts.[6]

Sergeants are responsible for supervising the behavior of their assigned officers. The primary departmental objective of this supervision is to ensure that officers are acting in accordance with department rules and not neglecting their patrol duties. However, in practice, sergeants have discretion to command their officers as they see fit in order to satisfy their interpretation of these objectives. This discretion manifests in heterogeneous supervisory styles that have been documented both anecdotally and empirically (e.g. Engel, 2001). For example, one former DPD sergeant I spoke with told me that his primary job was to provide support to officers in the field rather than explicit instruction. However, as he explained, other sergeants may not hesitate to tell their officers to enforce specific criminal offenses more strictly. One police officer who was interviewed as part of Van Maanen (1984)'s ethnography of police sergeants described this phenomenon succinctly:

> "Now you take Sergeant Johnson. He was a drunk-hunter. That guy wanted all the drunks off the street, and you knew that if you brought in a couple of drunks a week, you and he would get along just fine. Sergeant Moss, now, is a different cat... What he wants are those vice pinches. Sergeant Gorden wanted tickets, and he'd hound [you] for a ticket a night. So you see, it all depends on who you're working for. Each guy's a little different."

Outside of asking their officers to enforce specific crimes, the former sergeant I spoke with expressed a desire to manage officers who "like to work" — suggesting that some sergeants may value a more general notion of officer productivity, typically through responding to calls and being visibly active in the field. On the other end of the spectrum, some sergeants can be relatively uninvolved in their officers' patrol work. They may commit to their administrative duties and only help when absolutely needed, an archetype was pointed out by both the former DPD sergeant and Van Maanen (1984), who uses the term "station house sergeants" to describe them.

While officers in the field largely handle civilian interactions without their sergeants present, sergeants have access to a number of administrative and informal mechanisms that enable them to incentivize officers to adhere to their preferences. Sergeants are responsible for writing recommendations for promotion

[6] For especially large or crime-ridden sectors, there may be more than one sergeant assigned. Each of these sergeants manages their own unit of officers.

or transfers into coveted interview positions, such as investigation or tactical teams. They conduct yearly performance evaluations, and approve overtime requests and schedule changes. They are expected to review use of force and arrest reports, and examine patterns in their officers' consensual searches and citations. They can document formal disciplinary action for violating department procedures or commendations for exemplary behavior (Rim et al., 2024). Additionally, police culture places a large emphasis on rank hierarchy, which provides strong incentives for officers to obey their superiors (King, 2005).

Sergeants also have the option to spend their shift actively patrolling the streets alongside their officers, where they can respond to calls and make their own arrests. In addition to the mechanisms described above, officers may model ideal patrol behavior based on how sergeants lead by example. That is, sergeants who want their officers to make more arrests for drug crimes may be more likely to make these arrests themselves. Prior research suggests that officers are receptive to this style of management, as they may believe that "street sergeants" understand the complexities of patrol work and, as a result, garner more respect (Van Maanen, 1984). Moreover, Engel and Worden (2003) show in a survey study that officers are more likely to believe specific patrol activities will be used to assess their job performance if they observe their sergeants engaging in those activities.

Through field activity, sergeants can also overcome limitations to their monitoring capabilities. They can assign themselves to calls being handled by their officers and advise them directly. While officers are expected to call their sergeants when there is uncertainty regarding how to handle a situation, officers may be more willing to call someone who they know will show up in person.

In the DPD, patrol officers change sergeants frequently and have limited discretion to determine when these switches happen or who they are assigned to. An officer's sergeant will change if their unit assignment changes or if their current unit receives a new supervisor. Reassignments occur due to officer or sergeant vacancies that are generated by promotion, retirement, death, or transfers into specialized units. Unlike most other large police departments, Dallas does not allow officers to select into these vacancies.[7] Instead, officer vacancies may be filled within a division and watch at the discretion of executive command staff. When sergeant vacancies occur, other sergeants can interview for the opening, but the final transfer decision is made by command staff.

In addition to filling vacancies, officers may also receive a new sergeant through department-wide schedule realignments. Once a year, executive commanders determine staffing needs within each of the patrol stations, shifts, and days of the week. If large-scale staffing changes are needed, then the Chief of Police can implement a Patrol Bid, which allows a designated set of sergeants and/or officers to choose their station, shift, and day-off groups in descending order according to time in rank. Since the bid may not occur every year,[8] eligibility for the bid is not known until 2 weeks prior, and it does not always occur

[7] See Ba et al. (2021a) for a discussion of the vacancy assignment procedure in the Chicago Police Department and its implications for officer sorting between high and low crime districts.

[8] In my sample, it happens 3 out of 5 years.

in the same month, officers are limited in their ability to sort on *trends* in crime or behavior, which is crucial to the identification strategy that I discuss in Section 4. Additionally, officers are not allowed to choose their sector or sergeant. These assignments are up to the discretion of command staff and, according to conversations with DPD officers, do not seem to follow a discernible pattern.

To become a sergeant, an officer must meet a tenure threshold and pass a promotional exam that tests their knowledge of department procedures and leadership potential. Exam-takers are ranked according to their performance and, with a few rare exceptions, promoted in order of their ranking when openings arise. The list is only valid for 18 months, so officers with lower exam scores are less likely to be promoted in that cycle. In the DPD, officers are required to spend at least 1 year as a senior corporal — the rank just above police officer — before qualifying for the sergeant's exam.[9] From 2010 to 2020, the exam was held three times: 2012, 2014, and 2018. It is divided into two parts. The first part is a multiple choice test that asks about department bylaws and readings on police leadership. Exam takers are required to meet a score threshold to qualify for the second part; in practice, the vast majority of exam-takers meet this threshold. The second portion is an oral exam in which testers are asked how they would handle various leadership scenarios that they could encounter as a sergeant. Oral exams are scored by observers at the rank of sergeant and above who are brought in from outside police agencies. The multiple choice and oral exams are then combined into a weighted average, with the multiple choice exam accounting for 40% and the oral exam accounting for 60% of a candidate's final promotional score.

3 Data

This project uses several administrative datasets obtained through FOIA requests from the Dallas Police Department and the Dallas County District Attorney's Office, covering the period from June 2014 to July 2019. I combine data on police incidents, personnel information, officer activity, and court outcomes to construct a monthly panel of sergeant assignments that links officer enforcement activity to their respective sergeants. I focus on sergeant assignments for patrol officers, whose primary duties include answering civilian-initiated calls for service, patrolling assigned beats, and responding to crimes observed 'on-view.' Patrol sergeants are assigned to a sector of the city and a watch.

Dallas maintains assignment data only at the patrol station level, a less granular geographic level, meaning they do not keep records of sergeant assignments. However, the Computer Aided Dispatch (CAD) system, which tracks the assignment of officers to police incidents, stores the daily sector and watch assignments of responding officers (and sergeants) working an incident (see Appendix C for details). I use these assignment data, along with station assignments and promotion histories, to construct monthly

[9]Officers are required to have at least 3 years of experience at the police officer rank before being promoted to senior corporal. Combining this experience requirement with the time that officers spend in the academy and field training, aspiring sergeants require a minimum of 5 years of department experience before taking the sergeant's exam.

sergeant assignments for patrol officers from June 2014 to July 2019. Specifically, I assign officers to the sector-watch where they are assigned on the most days within the month, and assign each sector-watch the sergeant observed with that assignment the most days. This assignment construction yields a panel of 2,067 officers, 388 sergeants, 15,355 officer-sergeant spells, and 61,166 officer-month observations.

I am interested in the effects of an officer's regularly assigned sergeant, who evaluates officer performance and works with the officer on most of their workdays. However, in practice, officers may not be assigned to the same sergeant each day. During a sergeant's off days, their duties will be given to a rotational fill-in sergeant. Moreover, officers may be temporarily reallocated to a different sector-watch based on manpower needs. In both cases, an officer's regularly-assigned sergeant still carries administrative responsibilities for the officer. To the extent that officers receive advice and instruction from multiple sergeants within a month, my assignment method captures the effects of the sergeant to whom they are most often exposed. The sergeant I assign to a sector-watch is observed as the assigned sergeant in CAD on 9.1 unique days per month on average, suggesting that it is unlikely my assignment method consistently selects fill-in sergeants who are more active than the regularly assigned one.

In order to ensure my estimates are consistent with effects driven by an officer's regularly-assigned sergeant, I subject the sample to two filters. First, I require that officer-sergeant spells last at least 2 consecutive sample months. This minimizes assignment errors that could result from officers working in a temporary assignment with more activity than their permanent one, in which case officer arrests may be erroneously credited to the wrong sergeant. If the underlying sergeant truly did not change but I assigned the officer a new sergeant, then these errors would attenuate the variance in sergeant effects since officer behavior would only change for idiosyncratic reasons. This eliminates 5,747 spells, 19% of which are single months with no assigned sergeant. Next, I filter out the remaining 866 spells in which a sergeant cannot be identified. I show in Figure A.6 that these sample restrictions do not meaningfully change my estimates.

To facilitate identification of sergeant and officer fixed effects, I remove any officers and sergeants who only appear together, any officer/sector-watch and sergeant/sector-watch pairs that only appear together, and any officers, sergeants, sector-watches, or day-off groups that only appear once in the data. I also require that officers appear in the data in at least 5 separate months. These restrictions eliminate 310 officer-supervisor spells, yielding an analysis sample of 1,805 officers, 347 supervisors, 8,432 officer-supervisor spells and 49,923 monthly officer observations.

In order to study trends around officer moves, I also construct a balanced event study sample. I define an event as two chronological spells involving the same officer but different sergeants. Within the event study sample, I require the duration of spells to be at least 5 months prior to the switch and at least 4 months after the switch. Since switches are determined at a monthly level, a switch occurs sometime in the final month in which the officer is assigned to the previous sergeant in the data. The switching month does not contribute to the 5 month pre-switch requirement.

I supplement the panel of sergeant assignments with data on officer activity from several sources. I use the universe of arrest reports to count the number of arrests made by each officer in every month of my sample. I match each arrest to all of the charges listed at the time of apprehension and partition arrests into two categories: serious and low-level. These categories are defined as in Rivera (2022). Serious arrests include index crimes (i.e., murder, rape, robbery, aggravated assault, theft, burglary, and arson), which are high-cost crimes that are tracked by the FBI. Serious arrests also include several non-index crimes with high social costs: simple assaults, any form of domestic violence, sexual assault, fraud, and DWI.[10] All other arrests are classified as low-level. Low-level crimes primarily consist of outstanding warrants,[11] disorderly conduct, and drug possession, which account for 81% of low-level arrests. Low-level arrests also encompass a range of public order offenses with no clear victim, such as vagrancy, liquor violations, and prostitution. Arrests may contain multiple charges, so I classify each arrest based on the most severe charge. In other words, any arrest with a serious charge is classified as serious. To ensure that my results are not driven by this classification decision, I also consider a more traditional partition of arrests into index and non-index categories.

I link each arrest to court outcomes using records obtained from the Dallas County District Attorney's office and classify its conviction status.[12] A conviction occurs if the arrest is matched to a court case that does not result in a dismissal. Convictions thus include plea bargains, as well as those administered by a judge or jury. If a charge does not match to court data, I consider it dismissed. Conviction is defined at the arrest level, so that an arrest results in a conviction if the arrestee was convicted on any of the charges related to the arrest.

I extract 911 calls from CAD data to separately evaluate civilian-initiated and proactive police encounters.[13] An arrest is considered officer-initiated if it does not originate from a 911 call. Additionally, I merge use of force reports and civilian complaints to the involved officers and the month of occurrence. I link officers and sergeants to internal personnel records that contain demographic information, tenure and promotion history, shift, day-off group, and bureau assignments. Finally, I link each sergeant promoted in 2012 or later to the promotional score that they achieved on the sergeant's exam for which they were promoted. I am unable to obtain scores for exams before 2012, however 58% of sergeants in my sample

[10] The only difference between my classification and Rivera (2022)'s is the inclusion of DWI in serious crimes.

[11] While I am unable to determine the crime associated with the warrant, national data suggest that the majority of outstanding warrants are for non-violent crimes and ordinance violations, such as unpaid traffic tickets (Slocum et al., 2021).

[12] Specifically, I use the name of the arrestee and the offense date to match an arrest to a case within the universe of cases disposed within Dallas County from 2014 to 2020. I first match arrests to all court cases with the same offense date. Then I use Jaro-Winkler distance to calculate the similarity of the first and last names of the matched defendants. If an arrest has a matching case with first and last names that perfectly match (i.e., Jaro-Winkler score equal to 1), I keep only that case. For all other arrests, I keep a match if it has a Jaro-Winkler score of 0.9 or higher. This matching technique is similar to the one used by Adger et al. (2022) and allows for some spelling errors in the arrest report while still being conservative about the name similarity required for a match.

[13] I use the cleaning procedure described by Online Appendix A4 in Weisburst (2024) to isolate 911 calls in CAD.

were promoted in or after this year.

Summary statistics for the full unrestricted data, the analysis sample, and the balanced event study sample are given in Table 1. The analysis sample is similar to the unrestricted data, suggesting that estimates of sergeant effects are unlikely to be biased by sample selection. Officers in the event study sample have slightly lower arrest activity when compared to the unrestricted data and analysis sample. One likely explanation is that the event study sample requires officers to have successive stable patrol assignments, which excludes officers who may prefer making arrests and are more likely to transfer to specialized teams where they can make a large number of arrests, such as gang or narcotics enforcement. Since all of my analyses include officer fixed effects, these sample differences should not significantly affect my findings.

Table 1 shows that officers are highly mobile and sergeants are exposed to a large number of officers within the sample. The average officer has just under four unique sergeants, and the average sergeant manages over 20 officers. This density within the managerial network is vital for my empirical strategy, since sergeant fixed effects can only be identified within groups of officers and sergeants connected by moves (Abowd et al., 2002). In my data, all of the observations are within one connected set.

On average, patrol officers in my sample make 3.8 arrests per month, three-fourths of which are for low-level crimes. The proportion of low-level arrests in my data is comparable to the national proportion of misdemeanor arrests, which account for 80% of all arrests according to estimates by Natapoff (2016).[14] There is substantial variation in arrests between officers. The standard deviation is 3.64, nearly the same size as the mean. Figure A.3a plots the distribution of average arrests per officer-month for sergeants. This figure suggests significant variation across sergeants in the arrests made by their officers. Officers working for a sergeant in the right tail of the distribution make over 6 arrests per month on average, while those in the left tail average 1 arrest or fewer. However, the average number of arrests made by officers working for sergeants cannot identify a sergeant's effect on arrests, as it cannot be disentangled from officer discretion. This discretion translates to even larger variation in arrests across officers (see Figure A.3b). Separating the effects of officers from those of sergeants requires observing changes in officer behavior under different sergeants. This is the crux of the empirical strategy, which I detail in the next section.

[14]Low-level crimes, as classified here, are not all misdemeanors and not all misdemeanors are low-level crimes. For example, possessing personal-use amounts of marijuana is a misdemeanor, while possessing personal-use amounts of cocaine is a felony. However, I classify both as low-level crimes.

4 Empirical Strategy

4.1 Estimating Sergeant Effects

First, I estimate sergeant effects on the quantity of officer arrests. I follow the two-way fixed effects approach pioneered by Abowd et al. (1999), which has been used to identify manager effects in a variety of settings (Benson et al., 2019; Frederiksen et al., 2020; Fenizia, 2022; Metcalfe et al., 2023). The model is specified as follows:

$$y_{it} = \theta_i + \psi_{J(i,t)} + x_{it}'\beta + \nu_{it}, \tag{1}$$

where y_{it} is the number of arrests made by officer i in year-month t, θ_i is an officer fixed effect, and $\psi_{J(i,t)}$ is a fixed effect for officer i's sergeant in month t. The time-varying control vector x_{it} sector-watch fixed effects to account for spatial and temporal variation in crime, which may affect arrests. Since sector-watch and sergeant assignments overlap, separate identification of sergeant and sector-watch fixed effects requires that each sector-watch in my sample is managed by multiple sergeants. The data cleaning procedure described in Section 3 ensures this is the case in my analysis sample. In practice, each sector-watch is managed by an average of 6.95 sergeants (see row 10, column 2 of Table 1). I also include fixed effects for the day-off group of the officer in order to control for changes in an officer's scheduled days that coincide with changes in their sergeant assignments. I also include a second-degree polynomial of officer tenure to account for time-varying changes in arrest behavior that may be correlated with the officer's sergeant due to their priority in schedule realignments.

Sergeant fixed effects are identified through officers who switch sergeants. Specifically, sergeants in this model are credited only for *changes* in the behavior of officers who switch to or from them. By including x_{it}, sergeant effects are measured using changes in arrests relative to the average within an officer's patrol location, shift, day-off, and tenure group. For ψ_j to identify the causal effect of supervisor j, I require that officer mobility between sergeants is as-good-as random, conditional on officer fixed effects and the controls. In other words, sergeant assignments must be uncorrelated with determinants of officer behavior that are not included in the model. However, the model allows officers to sort to sergeants based on the permanent components of officer effects θ_i and sergeants effects $\psi_{J(i,t)}$. Thus, if officers who have a preference for making arrests tend to work with sergeants who encourage officers to make them, the identifying assumptions would not be violated. Following Card et al. (2013), I examine three forms of endogenous mobility that could violate the identifying assumptions.

First, sergeant assignments must be uncorrelated with unobserved, time-varying determinants of officer arrests, including trends in both officer behavior and crime within the assigned sector. For example, if sergeants who are more lenient toward low-level arrests are more likely to be assigned officers whose preference for making arrests is increasing over time, the model would erroneously attribute any gains in arrests to the new sergeant. Moreover, if officers were systematically moved to neighborhoods with increasing demand for police enforcement, this could overstate the variation in sergeant effects. Such scenarios might arise if, for example, high-arrest sergeants were more likely to pressure command staff to fill vacancies in their unit when crime is rising in their assigned area.

Second, I require that changes in an officer's sergeant do not coincide with unobserved shocks to their enforcement behavior. In this context, one may be particularly worried about departmental policy changes that coincide with an officer's move. For example, hot-spot policing is a common strategy in which resources are concentrated in areas with a high crime concentration (Weisburd and Eck, 2004). If high-arrest sergeants are better at identifying crime hot-spots and request new officers for hot-spot policing, I would observe increased arrests for officers moving to these sergeants, but for reasons unrelated to the sergeant's management style.

Finally, my identification assumes that officers do not sort to sergeants based on their idiosyncratic match quality. If, for example, command staff is able to match officers to sergeants with whom they have a comparative advantage in making arrests — a case of positive assortative matching — then there would be match-specific effects, η_{ij}, that are correlated with $\psi_{J(i,t)}$ and missing from the model.

Sergeant assignments in the DPD, as described in Section 2, limit officers' ability to sort based on these endogenous factors. While officers and sergeants can select their station and watch through the patrol bid, they cannot control when the bid occurs, which options are available during their selection, or their eventual sector assignments. Moreover, they cannot control the timing or destination of moves generated by vacancies, limiting their ability to adjust their behavior in anticipation of a switch or sort based on crime trends in specific locations. To the extent that there are contemporaneous changes in the unobserved determinants of arrests that are correlated with sergeant switches, the sergeant fixed effects will identify a combination of sergeant effects and effects from other sources. However, policy changes are unlikely to drive the patrol bid or vacancies. New policing initiatives are often carried out by teams distinct from regular patrol officers, as was the case for hot-spot initiatives in Dallas during the sample period (Jang et al., 2012).

I rigorously test the validity of the identification assumptions in Section 5. However, in the spirit of Card et al. (2013), one can also conduct simple event studies around officer moves in order to assess the variation in the data that is leveraged for identification of the sergeant fixed effects. In Figure 1, I present event studies by splitting sergeants into terciles based on the average number of arrests made by the officers they manage during the sample. I then plot officer arrest paths separately by the sergeant terciles that they transition to and from. Arrests are residualized by officer fixed effects and the control vector using within-supervisor variation, as in Chetty et al. (2014). In practice, this means that I estimate $\hat{\theta}_i$ and $\hat{\beta}$ by estimating equation 1. I then calculate $y_{it} - \hat{\theta}_i - x_{it}'\hat{\beta}$ using these estimates. This is necessary since any sorting pattern of sergeants would lead estimates of $\hat{\theta}_i$ and $\hat{\beta}$ to be contaminated by sergeant effects if the sergeant fixed effects were not included. In order to evaluate a reasonable pre- and post-switch window while maintaining most of the switches in the data, I use the sample of switches that are balanced 2 months prior to the move and 2 months after the move.

Figure 1 exhibits a few notable patterns. First, when an officer changes sergeants, their arrest behavior shifts abruptly and remains persistent. This is consistent with a fixed effects specification, in which the sergeant's effect activates once the officer moves and does not degrade over time. Second, while there is evidence of fluctuations in officer behavior prior to a switch, these fluctuations do not seem to be systematically related to the direction of the switch. Third, Figure 1 suggests that officer-sergeant match quality is not a key determinant of moves. Sorting based on match quality implies that officers tend to move to supervisors with whom they have a comparative advantage (or, in the case of negative sorting, disadvantage) in making arrests.One implication of sorting based on match quality is the asymmetry in the effects of upward and downward moves. As shown by Card et al. (2013), in the presence of an endogenous match effect η_{ij}, the expected difference in arrests as a result of the move to a high-arrest sergeant ($j = 2$) in period t from a low-arrest sergeant ($j = 1$) in period $t - 1$ is given by:

$$E[y_{it} - y_{it-1} | J(i,t) = 2, J(i,t-1) = 1] = \psi_2 - \psi_1 + E[\eta_{i2} - \eta_{i1} | J(i,t) = 2, J(i,t-1) = 1]. \tag{2}$$

Similarly, the expectation for a move in the opposite direction is given by:

$$E[y_{it} - y_{it-1} | J(i,t) = 1, J(i,t-1) = 2] = \psi_1 - \psi_2 + E[\eta_{i1} - \eta_{i2} | J(i,t) = 1, J(i,t-1) = 2]. \tag{3}$$

Under positive (negative) assortative matching, both expected difference in match quality terms will be positive (negative). Thus, there would be an average mover premium (cost), regardless of the arrest propensity of an officer's new sergeant. In general, a lack of match-based sorting implies that the arrest changes generated by moving from $j = 1$ to $j = 2$ are equal and opposite the changes caused by moving from $j = 2$ to $j = 1$. Upon visual inspection of Figure 1, it is clear that officers who move to a sergeant in a higher tercile experience an increase in arrests, while those who move to a lower tercile sergeant experience a decrease in arrests. Moreover, moves in opposite directions seem to be symmetric. For example, a move from the 3rd tercile to the 1st seems to be equal and opposite in magnitude to a move from the 1st tercile to the 3rd. Reassuringly, moves within the same tercile do not seem to produce average changes in either direction, suggesting no average premium or cost to moving. I further verify the symmetry across moves in Appendix Figure A.1, which plots the average change in residual arrests for upward moves against the average change for downward moves in the opposite direction. The points line up roughly along the -45 degree line, providing support for symmetry in my data.[15]

[15] Splitting sergeants into terciles results in relatively few cases of symmetric moves to evaluate. In Appendix Figure A.2, I perform the same event study analysis by splitting sergeants into quartiles. The conclusions of Figure 1 and Appendix Figure A.1 hold in a similar way in this case.

Under the identifying assumptions, the fixed effects will be unbiased. However, consistency requires that the number of observations tends to infinity *within each officer-sergeant pairing*. Thus the raw fixed effects are likely to be estimated with error even if the identification assumptions are satisfied. This error will be more severe for sergeants with few in-sample observations. To reduce estimation error in the fixed effects, I adopt Empirical Bayes shrinkage procedures commonly used in the teacher value-added literature.(e.g. Kane and Staiger, 2008; Chetty et al., 2014). Specifically, I bootstrap the estimation of equation 1 in order to obtain estimates of the variance in sergeant fixed effects that can be attributed to the true signal variance, σ_ψ, and the variance attributable to sampling error, σ_ϵ.[16]. I then multiply each raw fixed effect by the Empirical Bayes shrinkage factor, defined as the ratio of signal variance to total variance, $\frac{\hat{\sigma}_\psi}{\hat{\sigma}_\psi + \hat{\sigma}_\epsilon}$. As the contribution of the error variance grows, the Empirical Bayes factor shrinks a sergeant's effect toward the mean of the sergeant effect distribution, which is 0 by construction (see Appendix D for further details). I perform the same procedure for the officer fixed effects.

4.2 Variance Decomposition

In addition to individual sergeant fixed effects, I am also interested in the contribution of sergeants and officers to variation in enforcement outcomes.

$$Var(y^*_{it}) = Var(\theta_i) + Var(\psi_{J(i,t)}) + 2Cov(\theta_i, \psi_{J(i,t)}) + Var(\nu_{it}), \tag{4}$$

$$y^*_{it} = y_{it} - x_{it}\hat{\beta}. \tag{5}$$

I focus on variation in pair-level average residualized arrests because variation within an officer-sergeant pairing is uninformative for estimates of the sergeant fixed effects. Arrests are residualized by the controls, with $\hat{\beta}$ estimated using within-sergeant and within-officer variation from the full model in equation 1 (Chetty et al., 2014).

While the Empirical Bayes procedure reduces measurement error in the estimated fixed effects, the variance components may still be biased if the number of officer movers is too low relative to the number of sergeants — the well-known *limited mobility bias* problem (Andrews et al., 2008). This would overinflate the variance of sergeant fixed effects, leading us to conclude that sergeants have a greater impact than they actually do, and it would bias the covariance negatively due to the inverse correlation of measurement errors between officer and sergeant effects. This bias can be severe, as has been demonstrated in the context of firm-worker networks (Bonhomme et al., 2023). However, compared to other contexts, the

[16]For the bootstrap, I follow the procedure outlined by Best et al. (2023). I obtain residuals $\hat{\nu}_{it}$ and randomly resample them, stratifying by sergeant-officer pair in order to preserve the match structure of the data. I then re-estimate the sergeant fixed effects. I repeat this process 1000 times and use the distribution of fixed effect estimates for each sergeant to calculate $\hat{\sigma}_\psi$ and $\hat{\sigma}_\epsilon$

officer-sergeant mobility network in my data is particularly dense. Over 85% of officers in my sample switch sergeants and the entire sample is connected by officer moves.

Nonetheless, I apply the bias-correction strategies developed by Andrews et al. (2008) and Kline et al. (2020) to the variance component estimates. The Andrews et al. (2008) method derives the bias term under the assumption of homoskedastic errors, while Kline et al. (2020) (KSS hereafter) derive the bias term under unrestricted heteroskedasticity. The KSS bias term is a linear combination of each observation's variance, weighted by its influence on the plug-in variance estimator. The KSS bias-corrected variance terms take the form of leave-one-out estimators that rely on model parameters computed when leaving out the i-th observation. The KSS estimator can only be used on the leave-one-out connected set, which includes officers and sergeants who remain connected when any single officer is removed. The leave-one-out connected set removes only 3 sergeants and 3 officers from my data. In Section 5, I show that the bias-correction methods provide similar estimates of the variance and covariance components.

4.3 Disaggregating Sergeant Effects

Officers may allocate varying levels of effort to enforcing low-level versus serious crimes. Officers can increase low-level arrests by actively searching for evidence of such crimes. For instance, officers may stop more civilians on the street or conduct more frequent searches for contraband during interactions. While such arrests may be legally justified, there is limited evidence that they enhance public safety (Cho et al., 2023). On the other hand, an officer may make more serious arrests if they actively volunteer for 911 calls when other officers are not available and respond quickly to calls when they are assigned. It is not immediately clear whether a larger sergeant effect results from officers being encouraged to make more serious arrests, more low-level arrests, or both. The sergeant effect is a composite of both serious and low-level effects:

$$\psi_j = \psi_j^L + \psi_j^S.$$

The correlation between ψ_j^L and ψ_j^S reveals whether there is complementarity between low-level and serious enforcement as induced by sergeants. If sergeants who induce more arrests achieve this through uniform increases in officer effort, we would expect ψ_j^L and ψ_j^S to be positively correlated. In that case, it would suggest that officers' time constraints are not binding, meaning they may spend significant portions of their shift inactive or extend their shifts with overtime. In contrast, a negative correlation suggests that sergeant effects involve a trade-off between low-level and serious enforcement. In this case, officers may have binding time constraints that force them to choose one type of enforcement activity at the expense of the other. A lack of correlation between low-level and serious sergeant effects suggests that sergeants induce each type of arrest through independent behaviors, and officers are unlikely to face binding time constraints.

Estimates for ψ_j^L and ψ_j^S are obtained using two-way fixed effects models with low-level and serious arrests as outcomes, respectively. I shrink the raw estimates of low-level and serious sergeant effects using the Empirical Bayes procedure described in Section 4.1. Since low-level and serious arrests occur at different rates in the data, I focus on standardized versions of each to allow for interpretable comparisons:

$$\hat{\psi}_j^{L*} = \hat{\psi}_j^L / SD(\hat{\psi}_j^L),$$

$$\hat{\psi}_j^{S*} = \hat{\psi}_j^S / SD(\hat{\psi}_j^S).$$

5 Results

5.1 Sergeant Effect Estimates

Figure 2 plots the density of the raw and shrunken sergeant effects. By construction, the mean of the sergeant effects distribution is 0, so that each sergeant effect is interpreted as the number of arrests induced per month relative to the average sergeant. As expected, the shrinkage procedure reduces variation and increases the concentration of effects around 0. However, even after shrinkage, a one standard deviation increase in sergeant effects is associated with an increase of 0.66 arrests per month, or 17% relative to the mean. The distribution of fixed effects is roughly symmetric around the average sergeant but has a heavy left tail, suggesting a significant number of sergeants with low enforcement levels. The estimates suggest that moving from a high arrest sergeant to a low arrest sergeant makes a sizeable difference in the enforcement behavior of officers: moving an officer from a 10th percentile sergeant to a sergeant in the 90th percentile would lead to 1.6 more arrests per month (42% relative to the mean).

To contextualize these magnitudes, we can calculate the impact of replacing a sergeant with a new one.[17] Replacing all sergeants above the 90th percentile with sergeants at the 50th percentile would lead to 2,380 fewer arrests over the sample period, a 2.26% reduction in total arrests. This suggests that replacing a small number of sergeants — in this case, just 35 — could have a substantial impact. While there is greater variation in officer effects than in sergeant effects (see Figure A.4), the fact that sergeants manage many officers suggests that variation in sergeants may play a larger role in explaining the total number of arrests. To illustrate this point, consider the model-implied effect of replacing a sergeant for one month compared to replacing an officer at the same point in the distribution. For example, replacing a 90th percentile sergeant who manages the average number of officers (6.33 officers per month) with a

[17] To avoid double-counting arrests attributed to multiple officers, I adjust the arrest measure so that each officer receives credit for only half an arrest when multiple officers are listed on the arrest report. This does not substantially alter the results compared to the original arrest measure.

50th percentile sergeant would lead to 2.8 fewer arrests, holding other factors constant.[18] On the other hand, an equivalent replacement of a 90th percentile officer would only lead to 2.6 fewer arrests.

To generalize this result, I calculate the change in arrests that would occur from replacing a sergeant at each percentile of the effects distribution with a 50th percentile sergeant for one month, assuming that they manage the average number of officers. I then calculate this equivalently for each officer. In Figure A.7, I compare these changes at each percentile of the arrest effect distributions. In 90% of cases, replacing a sergeant leads to changes that are at least as large (in absolute value) as those from replacing an equivalent officer.[19] The largest differences between sergeant and officer replacements occur in the left tail of the distribution, where a concentration of sergeants induce far fewer arrests than the average. Replacing sergeants from the 60th to 90th percentiles also leads to larger changes than equivalent officer replacements, but the gap narrows beyond the 90th percentile, consistent with the long right tail of the officer effects distribution. Because a small number of officers make many arrests, replacing those officers can have as large an effect as replacing an equivalent sergeant. However, across most of the distribution, replacing a sergeant leads to changes that are at least as large as replacing an equivalent officer. I interpret these findings as evidence that variation in sergeants is more impactful for arrests overall than variation in officers.

I present the results of the variance decomposition in Table 2, which include decompositions using raw fixed effects, Bayes-shrunken fixed effects, and both variance component bias-correction methods. As expected, the raw fixed effects overestimate the contribution of sergeants to variation in officer arrests. All three of the bias-adjustment methods produce smaller estimates of the variance in sergeant and officer effects. With the preferred KSS specification (columns (7) and (8)), I find that variation in sergeants can explain 3.39% of the total variation in officer arrests. In contrast, officers account for nearly three-fourths of the variation in arrests. Given that I estimate arrests at the officer level, the larger contribution of officers relative to sergeants is not surprising. However, as demonstrated previously, sergeant effects are magnified by the breadth of officers whom they manage, which is not quantified in the variance decomposition.

In the fourth row of Table 4, I report the covariance between sergeant and officer effects. I find evidence that high-arrest officers tend to sort to low arrest sergeants across all specifications. One interpretation of this result is that officers who make a lot of arrests prefer to work under relatively uninvolved supervisors, which may lead to less scrutiny over their behavior. However, the magnitude of sorting is small and accounts for no less than -1.51% of the total variation across each specification. Consistent with institutional practices that constrain an officer's ability to select specific sergeants, sorting—even on fixed

[18]The calculation is performed by taking the difference between the 90th percentile of the sergeant effects distribution and the 50th percentile, and multiplying that by 6.33.

[19]Instead of assuming each sergeant manages the average number of officers, one can perform these calculations using the actual number of officers managed by each sergeant. This approach shows that 72% of sergeant replacements produce larger changes than equivalent officer replacements. Calculations are available from the author upon request.

characteristics—appears to be limited.

My estimates can be indirectly compared to manager effects on productivity in other settings. My estimate of the variance in arrests attributable to sergeants is roughly half the size of Adhvaryu et al. (2020)'s estimate of the variance in worker productivity attributed to line managers in Indian garment factories (7.3% vs. 3.4%). However, officers in my context explain a significantly larger portion of the variation in arrests than factory line workers explain in their own productivity (5.4% vs. 72.3%). These results demonstrate the substantial amount of discretion that police officers have relative to workers in other contexts. In Lazear et al. (2015)'s study of managers within a technology-based services firm, the authors estimate that a 1 standard deviation change in manager effects increases productivity 2.6 times more than a standard deviation change in worker effects, assuming the manager has an average-sized team. The same calculations in my context imply that increasing sergeant effects by 1 standard deviation produces an effect 1.44 times larger than a standard deviation increase in officer effects, again suggesting that sergeants matter more than officers for variation in total arrests, though they may matter less than managers in previously studied industries.

Since front-line workers in policing have significant discretion in making arrests, it is not surprising that police management has a smaller effect on their behavior. The fact that managerial changes still outperform worker changes in policing is particularly noteworthy. One of the central questions posed by this paper is whether front-line managers can change police behavior, given the limitations of their position. The results presented in this section suggest that they indeed can.

5.2 Diagnostic Checks

This section conducts diagnostic checks to address concerns regarding the validity of the fixed effects estimates from equation 1. To begin, I provide evidence in support of the identifying assumption that sergeant mobility is uncorrelated with unobserved determinants of arrests. Section 4 discusses three forms of endogenous officer mobility that could bias my estimates of the sergeant effects. I begin by assessing each of these identification threats in turn.

First, I consider endogenous mobility based on trends in officer behavior or crime. There may be concern that officers are assigned to sergeants based on recent changes in their arrests. For example, an officer's arrests may decline after attending a mandated training program. If they were then more likely to be assigned to a sergeant who demands fewer arrests, part of the reduction in arrests attributed to the new sergeant would stem from the declining trend before the switch. On the other hand, if an opening within a low-enforcement sergeant's unit arises following a retirement, one might be worried that the vacancy is likely to be assigned to an officer whose arrests are trending upwards if the department is receiving complaints from their aggressive behavior. In either case, sergeant effects would be biased by trends in officer behavior.

In order to test for this form of endogenous sorting, I examine heterogeneity in trends prior to a sergeant move using an event study. Specifically, I estimate a model of arrest behavior around the time of a move:

$$y_{et} = \alpha_e + \sum_{k \neq -1} [\pi_0^k D_{et}^k + \pi_1^k D_{et}^k (\Delta \hat{\psi}_e)] + x'_{et}\beta + \epsilon_{et}. \tag{6}$$

Here, e indexes a switching event, uniquely determined by officer i and the switch month T, and k indexes months relative to T. The variable D_{et}^k is an indicator for an observation being k months relative to the switch. The coefficients π_0^k capture dynamics related to a change in sergeants which are common across all switches. I include baseline model controls (tenure, sector-watch fixed effects, and day-off group fixed effects) to adjust for time trends and an event fixed effect, α_e, in order to control for differences in baseline arrest rates prior to the switch.

The parameters of interest are the π_1^k's, which capture period-specific heterogeneity depending on the size of the change in sergeant effect. I test for endogenous reassignment by examining the pre-move event study coefficients. The event study model also nests a test for general misspecification of the sergeant effects, as equation 1 implies that a sergeant switch results in an instantaneous and non-degrading change in arrests. I estimate equation 6 using the event study sample, so that $k \in [-5, 4]$.

I plot the event study coefficients in Figure 3. Reassuringly, there is no evidence of heterogeneous trends in arrest behavior prior to officers changing sergeants.. An F-test of joint significance for the pre-move coefficients yields a p-value of 0.8473 (see column 1 of Table B.2). Moreover, following a switch to a high-arrest sergeant, an officer's arrests immediately increase and remain elevated throughout the panel, in line with the insights from the nonparametric event study in Figure 1.[20]

The previous test cannot capture whether officers sort to sergeants based on crime rate trends. One may be concerned that sergeants who prefer aggressive enforcement are more sensitive to changes in crime and are more likely to ask command staff to fill vacancies when crime is rising in their sector. This would result in officers making more arrests after a move because the demand for arrests in their new location is increasing, not because of their new sergeant. I test for endogenous crime trends by estimating the correlation between changes in sergeant effects and crime prior to the switch, measured by 911 calls.[21] I report these estimates in Table B.4, separately for officers moving into and out of each sergeant's unit.

[20]The size of the effect after moving is also close to 1, which is reassuring, as the π_1^k's are interpreted as the change in arrests following a move to a sergeant who induces one more arrest per month than the previous sergeant. To the extent that the estimates are below 1, this reflects measurement error that arises due to using estimated objects and a smaller subset of the data in which officers make slightly fewer arrests relative to the full sample.

[21]I use 911 calls as a measure of crime rather than crime reports, as crime reporting is endogenous to police activity (Weisburd, 2021). To the extent that aggressive policing can affect the public's willingness to contact the police, 911 calls may also be endogenous to police activity (Ang et al., 2024). Nonetheless, I take the stance that, because crime reports will be generated by proactive policing and 911 calls are civilian-initiated, the latter is a more appropriate indicator for crime in my context, as the reporting biases are likely to be smaller.

Based on joint F-statistics for the pre-period crime coefficients, I find no evidence that trends in crime are correlated with changes in the sergeant effect for incoming or outgoing officers.[22]

A second source of endogenous mobility arises from unobserved shocks correlated with sergeant assignment. For example, reassignment to high-arrest sergeants may occur at the same time as a policy that asks officers to make more arrests. I test for the presence of correlated contemporaneous shocks using a placebo test on 'incumbent' officers. Incumbents are officers who already work with the new sergeant of a switching officer when the switch occurs. If sergeant effects are systematically driven by unobserved policy shocks happening at the same time an officer switches, then one would expect the policies' arrest effects to be reflected in the incumbents' arrests as well.

I use an event study to test whether incumbent arrests are affected when a new officer joins the team. For each switching event e in which officer i changes from sergeant j to sergeant j', I model the number of arrests made by officers $l \neq i$ managed by sergeant j' 5 months before the switch and 4 months after:

$$Arrests_{let} = \alpha_{le} + \sum_{k \neq -1} [\pi_0^k D_{et}^k + \pi_1^k D_{et}^k (\Delta\hat{\psi}_e)] + x_{let}'\beta + \epsilon_{let}. \tag{7}$$

The model takes a form similar to equation 6. Once again, I am interested in the π_1^k terms, which describe how arrests made by incumbent officers in month k change when the difference in the effects of sergeants j' and j increases by 1. This model also provides a secondary test for endogenous crime trends, as we would expect arrests to increase for incumbent officers prior to a positive switch in sergeant effects if larger sergeant effects are driven by growing demand for arrests. In Figure A.10, I present the estimates and 95% confidence intervals for the event study coefficients. The estimates are close to 0 and insignificant across all months relative to the new officer's switch date. These results indicate that sergeant effects are unlikely to be contaminated by contemporaneous changes in enforcement policy.

The third identification concern pertains to match-specific error components. If officers sort to sergeants with whom they have a comparative advantage in making arrests, then the model will be misspecified and the fixed effects biased. Using the event studies in Figure 1, I showed it is unlikely that officers and supervisors sort based on match quality.

A related issue is that the model assumes sergeant and officer effects are additively separable. If sergeant effects were officer-specific, then the separate officer and sergeant fixed effects would not be informative and may be a product of statistical noise. I conduct two tests of the additive separability assumption. First, following Card et al. (2013), I examine the average residuals from equation 1 separately by groups of officer and sergeant effects. Specifically, I divide each officer-month observation into quintiles of officer and sergeant effects. If the additive separability assumption did not hold, then I would expect the

[22] This test looks at officers who switch sector-watches. One may still be concerned that sergeants with a large arrest effect are more likely to switch into sector-watches with increasing crime. Table B.5 presents results from a regression that predicts the sergeant effect of a sector-watch in each month using crimes in the last 5 months. I do not find evidence that trends in crime are predictive of sergeant effects.

model to systematically under- or over-estimate arrests for certain officer-sergeant groups. For example, if some aggressive officers felt more comfortable making arrests when working under an uninvolved station house sergeant, then we would expect large positive average residuals for top quintile officers matched with bottom quintile supervisors. Appendix Figure A.9 demonstrates that the mean residuals do not exhibit any clear patterns that would indicate a violation of the additive separability assumption. Across all officer-sergeant groups, the residuals are relatively small, ranging from -0.1 to 0.18, suggesting that the threat of misspecification is minimal in this setting.

A second test of additive separability compares the explanatory power of the baseline specification to a fully saturated model that contains a fixed effect for each officer-sergeant pair. I report the R^2 and Adjusted R^2 for these models in columns 3 and 5 of Appendix Table B.3. The fully saturated model fits better than the baseline, though the increase in Adjusted R^2 of 0.054 suggests that match components play a limited role in this setting. To the extent that match effects are present in the model, the evidence presented up to this point is most consistent with them being uncorrelated random effects.

In practice, each sergeant effect is identified using a relatively small number of officer switches — 33.4, on average. Even after using the Empirical Bayes and KSS bias-correction methods, one may still be concerned that the estimated sergeant effects are driven by noise. To show that sergeant effects capture meaningful variation in arrests, I estimate a set of 'placebo' sergeant effects by randomly reallocating sergeants to officers, preserving the number of unique officers for each sergeant. I then calculate the variance in arrests attributable to the placebo sergeants. I do this exercise 100 times and plot the distribution of variance estimates in Figure A.8, along with the KSS variance estimate from Table 2. To be conservative, I do not perform bias correction for the placebo. Reassuringly, the placebo estimates are close to 0 and my model variance estimate lies well outside a 95% confidence interval of the sergeant effect variance that would be obtained by chance.

In total, the findings from this section indicate that the sergeant effects identify meaningful changes in officer behavior that are attributable to their sergeant.

5.3 Disaggregating Sergeant Effects by Crime Type

In this section, I disaggregate the sergeant effects to evaluate sergeant effects separately for serious and low-level arrests. Figure 4 shows a binned scatterplot with low-level sergeant effects on the horizontal axis and serious sergeant effects on the vertical axis, accompanied by a linear fit and a nonparametric 95% confidence band (Cattaneo et al., 2024). The linear fit suggests a positive relationship between low-level and serious effects, and the estimated correlation being small but statistically significant at 0.11. However, the confidence band does not allow me to rule out a non-linear relationship between the two dimensions of sergeant effects. Visual inspection of the plot suggests that sergeants who induce significantly fewer low-level arrests than average also tend to induce fewer serious arrests; however, for the rest of the distribution, the relationship is flat.

Indeed, if I remove the bottom 5% of sergeants in the low-level effect distribution, the estimated correlation decreases to 0.07 and becomes statistically insignificant. For sergeants in the upper half of the low-level effects distribution, the estimated correlation with serious effects is -0.02 and I can rule out a correlation larger than 0.13 or smaller than -0.16. While a concentration of sergeants appears to reduce overall enforcement effort, the lack of a strong correlation between serious and low-level enforcement across the distribution suggests that managers influence low-level and serious arrests independently. Insofar as policies are interested in reducing low-level arrests through sergeants, my findings suggest that such policies are unlikely to reduce the enforcement of serious crimes.[23]

To illustrate this point, I calculate the change in low-level and serious arrests that would result from replacing the top 5% of low-level effect sergeants with those in the 50th percentile for both serious and low-level effects. This hypothetical change is coarse: I target only low-level sergeant effects, but, as a consequence, it also impacts the distribution of serious effects. I estimate that this change in the sergeant effects distributions would result in 1,018 fewer low-level arrests over the 5-year period, but *increase* serious arrests by 9.7. Because a roughly even proportion of sergeants with large low-level effects have positive and negative serious effects (see Figure A.12), replacing all of these sergeants would significantly reduce low-level enforcement without affecting arrests for serious crimes on-net. Although this example is likely unrealistic, as it assumes departments can perfectly target replacement sergeants, it demonstrates the potential of sergeant-focused personnel policies to affect enforcement practices that may be especially costly without negatively impacting socially beneficial police behaviors.

The lack of a strong relationship between low-level and serious sergeant effects is particularly striking, as these effects for *officers* are strongly and positively correlated (see Figure A.14). The complementarity of low-level and serious arrests for officers suggests that variation in arrests across officers can be attributed to differences in overall effort, since officers who make more arrests tend to do so for both serious and low-level crimes. This may occur if some officers derive larger intrinsic benefits or have lower costs from the act of arresting, but the difference between these costs and benefits are relatively constant between serious and low-level arrests. In other words, officers appear to be differentiated on their willingness to make any arrest, rather than their willingness to make arrests of a certain type. In contrast, sergeants appear to affect low-level and serious arrests through distinct actions. My findings suggest that sergeants may be more concerned than officers with how enforcement is distributed across different crimes. Insofar as policymakers aim to reduce overly aggressive low-level enforcement, targeting sergeants may be more effective than targeting individual police officers.

[23] The lack of correlation between sergeant effects on arrests for crimes of differing severity levels is robust to instead dividing arrests according to the traditional FBI index classification. See Appendix Figure A.11.

5.4 What Drives Sergeant Effects?

What behaviors do sergeants influence to induce more low-level or serious arrests? Since the broad categorizations of serious and low-level arrests mask richer within-category heterogeneity, I investigate this question by estimating how serious and low-level sergeant effects change officer behavior for different subcategories of arrests. I estimate regressions of the following form:

$$y_{it}^{c} = \alpha_{L}^{c}\hat{\psi}_{J(i,t)}^{L*} + \alpha_{S}^{c}\hat{\psi}_{J(i,t)}^{S*} + \theta_{i}^{c} + x_{it}^{\prime}\beta^{c} + \nu_{it}^{c}, \tag{8}$$

where y_{it}^{c} is the number of arrests of type c made by officer i in year-month t. As in the baseline specification, I control for officer, sector-watch, and day-off group fixed effects as well as officer tenure. The α^{c} coefficients capture the change in type c arrests associated with a one standard deviation increase in the low-level (serious) effect of an officer's sergeant.

I first estimate this equation for the number of arrests within more granular categories of crime. In Table 3, I report estimates for the 3 most frequent serious and low-level charges. Increasing a sergeant's serious effect leads to statistically significant increases across the three most frequent serious crimes: domestic violence (column 1), theft (column 2), and DWI (column 3). However, the effects are largest—both nominally and relative to the mean—for domestic violence arrests, which account for 0.45 arrests each month (12% of the total average) but increase by 37% when officers are assigned to a sergeant with a one standard deviation larger serious effect. Additionally, sergeants with larger low-level effects actually *reduce* domestic violence arrests by 0.016 per month (3.5% relative to the mean). Sergeants make crime-specific trade-offs, even though the two dimensions of sergeant effects are uncorrelated on aggregate. However, I find no evidence that larger low-level sergeant effects change enforcement of theft or DWI.

Among low-level crimes, the largest discrepancy between behavior induced by serious and low-level sergeant effects occurs for drug arrests. A one standard deviation increase in a sergeant's low-level effect increases an officer's drug arrests by 0.17 per month, which is over 50% of the mean. Roughly 90% of this effect is driven by increased arrests for simple possession, rather than drug distribution (see Table B.6). The same size increase for a sergeant's serious effect reduces drug arrests by 0.036 per month (10% of the mean). Conversely, both dimensions of sergeant effects are positively associated with warrant and disorderly conduct arrests, although the magnitudes are larger for the low-level effect.

These findings have important implications for racial disparities. Black civilians represent a disproportionate share of drug arrests (55% compared to 50%), suggesting that aggressive low-level enforcement may exacerbate existing disparities. Indeed, I find that Black arrests increase at a faster rate than Hispanic or white arrests in response to positive changes in low-level sergeant effects, even when accounting for the higher baseline arrest rates for Black civilians (Table B.8). Conversely, changes in serious sergeant effects increase arrests across all races in proportion with their baseline arrest rates, which is sensible given the lack of crime composition changes for serious sergeant effects.

Patrol officers can make arrests through two primary channels: self-initiated interactions — such as traffic stops, investigating abandoned buildings, or stopping citizens on the street — or 911 calls. I examine how sergeant effects impact arrests through each channel by estimating equation 8 separately for officer-initiated and call-initiated arrests. The results in Table 4 show that sergeant-induced serious arrests are entirely initiated through calls (column 2), while low-level arrests come from both calls and officer-initiated interactions. However, for low-level sergeant effects, officer-initiated interactions account for over 60% of the total increase in arrests, which is particularly striking given that officer-initiated arrests are less common than those from 911 calls. These findings suggest that low-level sergeant effects are linked to a broken windows style of policing, where sergeants direct officers to proactively seek out and enforce low-level crimes to reduce crime more broadly. However, such arrests may be particularly costly to society overall, as these behaviors have not generated sufficient costs to justify a civilian complaint. Moreover, in a separate event study analysis, I find no evidence that low-level sergeant effects are associated with reductions in overall crime (see Appendix E) — suggesting that these enforcement strategies are unlikely to justify their costs to society.

While officer-initiated arrests are highly discretionary, it is less clear whether call-initiated arrests would change due to increased exposure to 911 calls or a lower threshold for criminal behavior, conditional on the call response. In Dallas, patrol officers can volunteer for calls if the unit assigned to that area is unavailable. Additionally, officers may work overtime, enabling them to answer more calls than they typically would during their shift. This suggests that some portion of sergeant effects may operate through increased call activity, rather than through a higher likelihood of arresting at each call. I investigate the mechanisms behind call-initiated arrests by regressing monthly call-specific outcomes on low-level and serious sergeant effects, with results presented in Table 5. I find that the number of calls answered per month increases with positive changes in both the serious and low-level sergeant effects (column 1). A one standard deviation increase in the serious sergeant effect leads subordinates to answer 2.2 more calls per month, compared to an average of 61.4. Increasing the low-level sergeant effect results in less than half the additional calls. Since I find no evidence that a sergeant's serious or low-level effect changes the number of 911 calls received in their sector (see Appendix E), these results are consistent with officers choosing to answer additional calls when working for sergeants with larger arrest effects.

While serious effects are associated with larger changes in the number of calls answered, low-level effects lead to more substantial changes in the likelihood that officers make an arrest at the calls they respond to (column 2). A one standard deviation increase in low-level effect leads to arrests at .3% more calls, which is nearly twice the size of changes caused by increasing the serious effect. Unsurprisingly, sergeants who increase low-level arrests primarily induce low-level arrests at calls. However, higher serious effects are associated mainly with increases in serious call arrests, along with a small increase in low-level call arrests (columns 3 and 4). These findings reinforce the earlier observation that management behaviors that increase low-level arrests in isolation do not seem to meaningfully impact enforcement for serious crimes. Moreover, it appears that serious sergeant effects are associated with more arrests for low-level crimes, but only those that are reported to the police by a civilian rather than detected by an officer on patrol.[24]

Figure A.15 contextualizes the findings from Table 5 by reporting similar results for call outcomes disaggregated by call type. For sergeants with large low-level effects, their officers arrest at a higher rate, even conditional on the severity-level of the call. Strikingly, the effect on arrest probability is largest for the least severe call type — mischief. However, the figure shows that the positive overall effect on the call arrest percentage for sergeants with high serious effects is driven entirely by officers responding to more serious calls. There is no evidence that increasing the serious sergeant effects raises arrest probability after conditioning on call type. In other words, serious sergeant effects operate entirely through call-response effort, while low-level sergeant effects drive discretionary enforcement, even during civilian-initiated 911 calls.

The results thus far suggest that sergeants incentivize their officers to change the quantity of serious and low-level arrests. However, do sergeants value the *quality* of arrests? To answer this question, I estimate the impact of serious and low-level sergeant effects on conviction rates. Doing so requires overcoming an empirical challenge. It is not uncommon for an officer to make zero arrests in a month, which prevents me from studying conviction rates directly. I adopt the approach used by Gudgeon et al. (2023) and estimate the impact of serious and low-level sergeant effects on the number of convicted arrests and total arrests separately. I use these estimates to calculate how changes in the serious and low-level sergeant effects affect the ratio of convicted arrests to total arrests, and then compare the new ratios to the ratio of averages. I plot the estimated changes in the conviction rate along with bootstrapped 95% confidence intervals in Figure 5. I also plot estimated changes in the conviction rate for serious and low-level arrests separately.

Both dimensions of sergeant effects increase overall conviction rates. This suggests that sergeants who induce arrests of either kind do not do so by encouraging officers to make low-quality arrests that will ultimately be thrown out in court. However, the results for both serious and low-level conviction rates indicate that sergeant-induced arrests are not necessarily of higher quality. Serious sergeant effects are not associated with changes in the serious conviction rate and are actually associated with *lower* conviction rates for low-level arrests. On the other hand, low-level sergeant effects are associated with higher conviction rates for both types of arrests. These patterns are driven by compositional shifts in the types of arrests that officers make. Serious arrests have a higher conviction rate than low-level arrests, so conviction rates increase when working for sergeants with high serious effects, since serious effects make up a larger share of your arrests. On the other hand, drug arrests have a conviction rate 4.75 times higher

[24] Figure A.13 depicts the results from Tables 3, 4, and B.8 graphically, with estimates normalized to the mean of each arrest category to allow for easier comparison of effects across categories.

than the average low-level arrest. As a result, the disproportionate impact of low-level sergeant effects on drug arrests increases low-level conviction rates for their officers, while the negative impact of serious sergeant effects on drug arrests decreases low-level conviction rates for officers working with sergeants who prioritize the enforcement of serious crimes.

Finally, I consider how sergeant effects interact with two other measures of costly police behavior: use of force and complaints. In Table 6, I report results from regressions that estimate each of these outcomes as a function of the low-level and serious sergeant effects. In column 1, I find that increases in both serious and low-level sergeant effects lead to more uses of force. However, the change is significantly larger for low-level sergeant effects. A one standard deviation increase in the low-level effect leads to 0.02 more uses of force per month, a 14% increase relative to the mean, compared to a change of 0.006 for an equivalent increase in serious effects. In column 2, I find that low-level sergeant effects increase complaints, while serious sergeant effects decrease complaints; however, both coefficients are imprecisely estimated. Both serious and low-level sergeant effects lead to more officer activity and more formal interactions with civilians, likely contributing to increased use of force. However, the stark difference in effect sizes suggests that targeted low-level enforcement may lead to violent escalation, likely disproportionate to the costs of the crimes it aims to address.

Overall, the results in this section reveal that, although sergeants have substantial influence over their officers' arrest decisions, they do so in heterogeneous ways. In particular, sergeants may induce more serious or more low-level arrests, and I find evidence that sergeants can change officer behavior along one dimension without significantly altering behavior along the other. Indeed, inducing low-level arrests and serious arrests leads to very different downstream officer behaviors. Officers make more serious arrests through call activity. In contrast, to induce more low-level arrests, sergeants incentivize their officers to proactively detect low-level crimes during patrols, particularly those involving drug possession.

5.5 Mechanisms: Leading By Example and Monitoring Officers

I now turn to the question of *how* sergeants are able to change the actions of their officers. Since police officers not only have discretion but also operate largely outside the direct view of their sergeants, there has been a long-standing debate over whether supervision *can* change officer behavior, let alone whether it actually does (Brown, 1988). In this section, I evaluate two measurable sergeant behaviors that could affect officer decisions: leading by example and direct monitoring. As explained in Section 2, leading by example may be particularly effective, since officers are more likely to respect the "street sergeants" who understand firsthand the complexities of working in patrol (Van Maanen, 1984). Accordingly, I use two proxies for leading by example: a sergeant's *own* arrest activity and the number of calls to which sergeants respond as first-responders. To distinguish between lead by example mechanisms for low-level and serious sergeant effects, I separately consider a sergeant's serious and low-level arrests.

Sergeants can also influence their officers' behavior through increased monitoring, which may involve assigning themselves to their officers' calls more frequently. In addition to overcoming the inherent limitations of police supervision, this approach may also increase the likelihood that officers will seek advice from their sergeants in future uncertain situations, knowing that their sergeant is more likely to respond in person. This would enable sergeants to give more direct advice in line with their enforcement preferences. I test for sergeant monitoring using a sergeant's presence at their officers' calls. I measure sergeant presence using CAD call assignments.

I estimate the importance of these mechanisms by regressing the monthly behaviors of a sergeant against their serious and low-level arrest effects. In a given month, sergeant behaviors may be influenced by their sector-watch assignment or the composition of their subordinates. I estimate the impact of low-level and serious sergeant effects on sergeant activities by leveraging within-assignment variation. Specifically, for unit u (i.e. a sector-watch) managed by sergeant j in month t, I estimate models of the following form:

$$y_{jut} = \alpha_L \hat{\psi}_j^L + \alpha_S \hat{\psi}_j^S + \alpha_1 \bar{\hat{\theta}}_{ut}^L + \alpha_2 \bar{\hat{\theta}}_{ut}^S + x_u + \epsilon_{jut}, \tag{9}$$

where y is an action of supervisor j in unit u during year-month t. I include sector-watch fixed effects (x_u) to control for variation in sergeant behaviors driven by the time and location of their assignments. Since some sergeant behaviors occur explicitly in response to the needs of their subordinates, I control for the average low-level and serious arrest effects of officers within their unit.

I report results for these regressions in Table 7. I find that low-level sergeant effects are associated with leading by example (columns 1-4). Sergeants with large low-level effects make significantly more arrests (column 1) and these arrests are exclusively low-level (column 3). I also find that low-level sergeant effects are associated with being the first-responder at more calls (column 4). Additionally, low-level sergeant effects are associated with greater officer monitoring, as sergeants respond to 7.6% more subordinate calls per month relative to the mean for every one standard deviation increase in low-level effects (column 5).

In contrast, I find no evidence that serious sergeant effects are associated with leading by example or enhanced officer monitoring in the field. The point estimates for serious sergeant effects in each column are small and imprecisely estimated. It is likely that sergeants induce serious arrests through other forms of behavior not captured by the data. These sergeants may provide better transfer recommendations for their officers contingent on greater call activity. Alternatively, sergeants with a high serious effect may be more willing to communicate their preferences to officers directly through radio assistance.

One other possibility that can be indirectly tested in the data is granting officers more overtime contingent on their call activity. Police have very few overtime restrictions, which allows them to significantly increase their income if their sergeant is willing to grant overtime requests (Chalfin and Goncalves, 2023). I do not have access to overtime data, so I use officer shift data to measure the number of calls and arrests made outside of their regular hours. Table B.7 shows that nearly half of the total increase in calls answered as a result of changes in serious sergeant effects is driven by calls outside of an officer's regular hours. Moreover, overtime *arrests* increase, primarily for serious crimes. I also find that low-level effects are associated with significant increases in calls and arrests outside an officer's regular shift. Although these measures are imperfect proxies for overtime, they suggest that sergeants may use their administrative control over overtime approval to shape officer behavior.

6 Predicting Sergeant Effects

To what extent are sergeant effects mediated by observable characteristics that are determined *before* someone has been promoted to sergeant? Answering this question is important for two reasons. First, it has significant implications for structuring sergeant-focused policies, which, as my results suggest, may serve as an effective tool in police reform. If sergeant effects can be predicted based on pre-promotion performance, departments could use these insights to inform promotion decisions or provide targeted management training before officers assume full-time sergeant duties. Second, to the extent that sergeant effects provide some insight into the preferences of these first-line managers, knowing how these preferences vary according to pre-promotion characteristics is of independent interest. For example, it is not clear whether one's preferences as a worker carry over to their preferences as a manager to their former job position. One aspect of learning how to be a manager may operate through seeing one's old job through a new perspective; or, promoted workers may simply use their new job powers to impose the work environment they always wanted. In the context of policing, it is well-established that racial minorities and older police officers make fewer arrests (Ba et al., 2021a,b). The extent to which this variation carries over into managerial preferences provides insight into this question.

I examine variations in sergeant effect distributions across four observable pre-promotion characteristics: race, gender, age at the time of the promotional exam, and the score achieved on the exam. Since I only observe exam scores beginning with the 2012 round of tests, I limit my sample to the 202 sergeants who were promoted from these exams. I call sergeants "older" if they were above the average across all exams and I call them "above-average scorers" if they were above the average score for their particular exam.

In Figure 6a, I present densities separately by exam score. I observe a notable difference between high and low scorers in the distributions of both serious and low-level effects. For low-level effects, the distribution of above-average scorers is shifted to the left of that for below-average scorers, with a Kolmogorov-Smirnov test indicating that this difference is statistically significant (p-value = 0.034). Thus, individuals who score below average on the promotional exams tend to induce more low-level arrests than those who score above average. These differences are meaningful considering exam score is the primary determinant of promotion. The exams test for knowledge of department procedures and aptitude within relevant supervisory situations. Thus marginal promotees — who are barely promoted by virtue of their low exam score — are more likely to value low-level arrests. This result is particularly interesting in light of the fact that, anecdotally, only a small portion of qualified officers take the sergeant promotional exam. For the 2018 exam, only 24% of eligible officers took the test. If officers who chose not to take the exam were deterred for reasons other than fear of weak performance on the test, departments may be able to reduce sergeant-induced low-level enforcement by encouraging more qualified officers to take the exam.

In Figure 6b, I also find evidence of a statistically significant difference in the distribution of serious effects between high and low exam scorers (Kolmogorv-Smirnov p-value = 0.013). In contrast to low-level effects, the differences in serious effects are not driven by a monotonic shift in one direction. Instead, high scorers exhibit a wider distribution of serious effects than low scorers. The distribution for low scorers is more concentrated around 0. These findings suggest that high scorers are significantly more heterogeneous than low-scorers, at least in terms of their effects on serious enforcement. One potential explanation is that high scorers may tend to be more bookish and administratively-inclined, as was suggested by Van Maanen (1984). It is possible that this group of test takers is more heterogeneous in their preferences for involvement with their officers, as some may be more focused on station house duties and others more intent on motivating their officers.

In the appendix, I present the empirical density of low-level and serious effects, separately by race (Figures A.18, A.19), gender (Figures A.16, A.17), and age groups (Figures A.20, A.21). I do not find evidence of significant differences in the distribution of either sergeant effect along each of these three observable dimensions. Kolmogorov-Smirnov tests for the equality of the distributions generate p-values that are well-above standard significance thresholds. However, I find evidence that low-level officer effects differ by race (Figure A.22), consistent with the previous literature (e.g. Ba et al., 2021b). Thus, it appears that variation in officer enforcement effects do not translate into similar variation in sergeant effects. The exam-based promotion mechanism may filter out officers from different racial categories, which could explain the variation in officer behavior across racial groups. For example, officers who are uniformly less likely to make low-level arrests may select into the promotion process at a higher rate. This issue warrants further, in-depth analysis, which is beyond the scope of this paper. However, my findings suggest that more work should be done to unpack the implications of exam-based promotion mechanisms.

7 Conclusion

This paper demonstrates that supervision plays a crucial role in shaping police enforcement decisions. Importantly, supervisors' effects manifest through distinct enforcement behaviors. My findings have several important implications for police reform policy.

First, training programs or personnel realignments targeting first-line supervisors may be particularly effective in altering how police use discretion. Other police reforms, such as training programs, have shown to be effective but degrade in impact over time (Owens et al., 2018). My findings suggest that sergeants provide persistent incentives for their officers to police in specific ways, indicating that policies aimed at altering sergeants' preferences could have a lasting impact. Furthermore, since sergeants represent a smaller portion of police agencies, policies requiring regular retraining of sergeants could be more cost-effective than broader programs that require periodic retraining of all frontline officers While I cannot evaluate the costs of such policies, my findings suggest that more focus ought to be given to designing sergeant-based reforms.

Second, my results suggest that management styles focused on aggressively targeting low-level crimes are disconnected from efforts aimed at reducing more serious offenses. The independence of supervisory effects across these dimensions supports a growing body of literature suggesting that reducing violent and property crimes does not necessarily require harsh enforcement of low-level offenses, which are often more related to public health and civilians' overall quality of life (Cho et al., 2023). Additionally, I provide evidence that targeting low-level crimes results in more collateral damage through the use of force than strategies focused on increasing arrests for serious crimes.

Third, this paper underscores the importance of future research into promotion mechanisms, both in the public sector more broadly and in policing specifically. By showing that managerial effects vary systematically based on performance in a standardized promotion process, my findings suggest that even 'objective' promotion tools can result in unexpected trade-offs for public organizations when selecting staff for supervisory roles.

Ultimately, my paper suggests that policy interventions focused on first-line police management could offer a fruitful direction for future research.

References

Abowd, J. M., Creecy, R. H., and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. *Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.*

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

Adger, C., Ross, M., and Sloan, C. (2022). The effect of field training officers on police use of force. *Working Paper*.

Adhvaryu, A., Bassi, V., Nyshadham, A., and Tamayo, J. A. (2020). No line left behind: Assortative matching inside the firm. *NBER Working Paper no. 25852*.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2023). Managerial quality and productivity dynamics. *The Review of Economic Studies*, 90:1569–1607.

Agan, A., Doleac, J. L., and Harvey, A. (2023). Misdemeanor prosecution. *Quarterly Journal of Economics*, 138:1453–1505.

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

Ang, D. (2020). The effects of police violence on inner-city students*. *The Quarterly Journal of Economics*, 136:115–168.

Ang, D., Bencsik, P., Bruhn, J., and Derenoncourt, E. (2024). Community engagement with law enforcement after high-profile acts of police violence. *NBER Working Paper 32243*.

Ba, B., Bayer, P., Rim, N., Rivera, R., and Sidibé, M. (2021a). Police officer assignment and neighborhood crime. *NBER Working Paper 29243*.

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

Bacher-Hicks, A. and De La Campa, E. (2020). The impact of new york city's stop and frisk program on crime: The case of police commanders. *Working Paper*.

Benson, A., Li, D., and Shue, K. (2019). Promotions and the peter principle. *Quarterly Journal of Economics*, 134:2085–2134.

Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, 118:1169–1208.

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

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? evidence from india. *Quarterly Journal of Economics*, 128:1–51.

Bloom, N., Lemos, R., Sadun, R., and Reenen, J. V. (2015). Does management matter in schools? *Economic Journal*, 125:647–674.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122:1351–1408.

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41:291–322.

Brown, M. K. (1988). *Working the Street: Police Discretion and the Dilemmas of Reform*. Russell Sage Foundation.

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

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2024). On binscatter. *American Economic Review*, 114:1488–1514.

Chalfin, A. and Goncalves, F. (2023). Professional motivations in the public sector: Evidence from police officers. *NBER Working Paper 31985*.

Charles, S. (2021). Cpd revives controversial 'merit promotions' system.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104:2633–2679.

Chicago Police Department (2024). Promotions_process_and_timeline-iap-07-03.

Cho, S., Gonçalves, F., and Weisburst, E. (2023). The impact of fear on police behavior and public safety. *NBER Working Paper No. 31392*.

Dube, O., Macarthur, S. J., and Shah, A. K. (2023). A cognitive view of policing. *NBER Working Paper 31651*.

Engel, R. S. (2000). The effects of supervisory styles on patrol officer behavior. *Police Quarterly*, 3:262–293.

Engel, R. S. (2001). Supervisory styles of patrol sergeants and lieutenants. *Journal of Criminal Justice*, 29:341–355.

Engel, R. S. (2002). Patrol officer supervision in the community policing era. *Journal of Criminal Justice*, 30:51–64.

Engel, R. S. and Worden, R. E. (2003). Police officers' attitudes, behavior, and supervisory influences: An analysis of problem solving. *Criminology*, 41:131–166.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica*, 90:1063–1084.

Frake, J. and Harmon, D. (2023). Intergenerational transmission of organizational misconduct: Evidence from the chicago police department. *Working Paper*.

Frederiksen, A., Kahn, L. B., and Lange, F. (2020). Supervisors and performance management systems. *Journal of Political Economy*, 128:2123–2187.

Gaure, S. (2013). lfe: Linear Group Fixed Effects. *The R Journal*, 5(2):104–116.

Gaure, S. (2014). Correlation bias correction in two-way fixed-effects linear regression. *Stat*, 3:379–390.

Geller, A., Fagan, J., Tyler, T., and Link, B. G. (2014). Aggressive policing and the mental health of young urban men.

Giorcelli, M. (2019). The long-term effects of management and technology transfers. *American Economic Review*, 109:121–152.

Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P. N., and Wooldridge, J. M. (2015). An evaluation of empirical bayes's estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics*, 40:190–222.

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

Ingram, J. R., Weidner, R. R., III, E. A., and Terrill, W. (2014). Supervisory influences on officers' perceptions of less lethal force policy: A multilevel analysis. *Policing*, 37:355–372.

Jang, H., Lee, C. B., and Hoover, L. T. (2012). Dallas' disruption unit: Efficacy of hot spots deployment. *Policing*, 35:593–614.

Johnson, R. R. (2011). Officer attitudes and management influences on police work productivity. *American Journal of Criminal Justice*, 36:293–306.

Johnson, R. R. (2015a). Leading by example: Supervisor modeling and officer-initiated activities. *Police Quarterly*, 18:223–243.

Johnson, R. R. (2015b). Police organizational commitment: The influence of supervisor feedback and support. *Crime and Delinquency*, 61:1155–1180.

Kane, T. and Staiger, D. (2008). Estimating teacher impacts on student achievement: An experimental evaluation.

Kapustin, M., Neumann, T., and Ludwig, J. (2022). Policing and management. *NBER Working Paper 29851*.

King, W. R. (2005). Toward a better understanding of the hierarchical nature of police organizations: Conception and measurement. *Journal of Criminal Justice*, 33:97–109.

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

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics*, 33.

Makowsky, M. D. and Stratmann, T. (2009). Political economy at any speed: What determines traffic citations? *American Economic Review*, 99:509–527.

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

Mello, S. (2021). Fines and financial wellbeing.

Mello, S., Ross, M., Ross, S., and Johnson, H. (2023). Diversity training and employee behavior: Evidence from the police. *Working Paper*.

Metcalfe, R. D., Sollaci, A. B., and Syverson, C. (2023). Managers and productivity in retail. *NBER Working Paper 31192*.

Morris, C. N. (1983). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78:47.

Mummolo, J. (2018). Modern police tactics, police-citizen interactions, and the prospects for reform. *Journal of Politics*, 80:1–15.

Natapoff, A. (2016). *Oxford Handbook Topics in Law*. Oxford University Press.

Owens, E. and Ba, B. A. (2021). The economics of policing and public safety. *Journal of Economic Perspectives*, 35:3–28.

Owens, E., Weisburd, D., Amendola, K. L., and Alpert, G. P. (2018). Can you build a better cop?: Experimental evidence on supervision, training, and policing in the community. *Criminology and Public Policy*, 17:41–87.

Rasul, I. and Rogger, D. (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *Economic Journal*, 128:413–446.

Rim, N., Rivera, R., Kiss, A., and Ba, B. (2024). The black-white recognition gap in award nominations. *Journal of Labor Economics*, 42:1–23.

Rivera, R. (2022). The effect of minority peers on future arrest quantity and quality. *Working Paper*.

Rivera, R. G. and Ba, B. (2022). The effect of police oversight on crime and allegations of misconduct: Evidence from chicago. *Working Paper*.

Roberts, J. and Shaw, K. L. (2022). Managers and the management of organizations. *NBER Working Paper 30730*.

Rose, E. K., Schellenberg, J., and Shem-Tov, Y. (2022). The effects of teacher quality on adult criminal justice contact. *NBER Working Paper 30274*.

Slocum, L. A., Schaefer, B. P., Torres, L., Huebner, M. M. B., and Hughes, T. (2021). Warrant enforcement in louisville metro and the city of st. louis from 2006 – 2019: A cross-site analysis.

Van Maanen, J. (1984). Making rank: Becoming an american police sergeant. *Urban Life*, 13:155–177.

Weisburd, D. and Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *Annals of the American Academy of Political and Social Science*, 593:42–65.

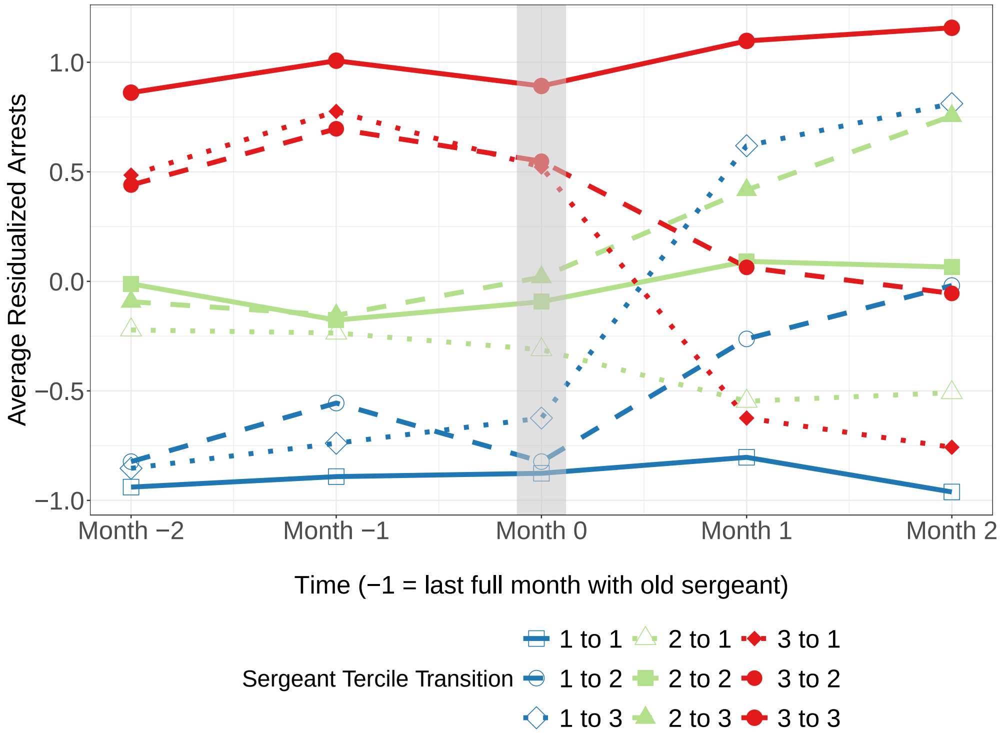
Weisburd, S. (2021). Police presence, rapid response rates, and crime prevention. *Review of Economics and Statistics*, 103:280–293.

Weisburst, E. K. (2024). Whose help is on the way? *Journal of Human Resources*, 59:1122–1149.

Zhao, J. Scheider, M. C., and Thurman, Q. (2003). A national evaluation of the effect of cops grants on police productivity (arrests) 1995-1999. *Police Quarterly*, 6:387–409.

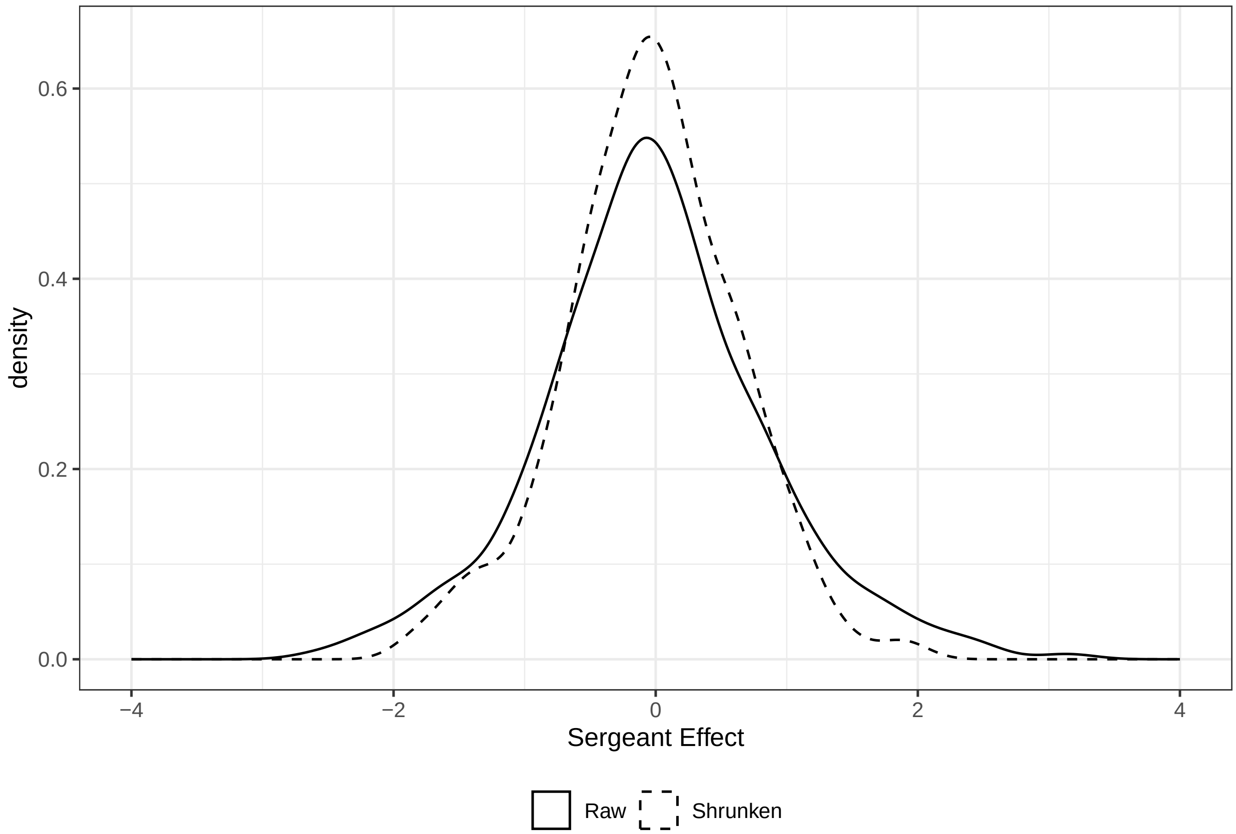
8 Tables and Figures

Figure 1: Event Study Around Sergeant Switch



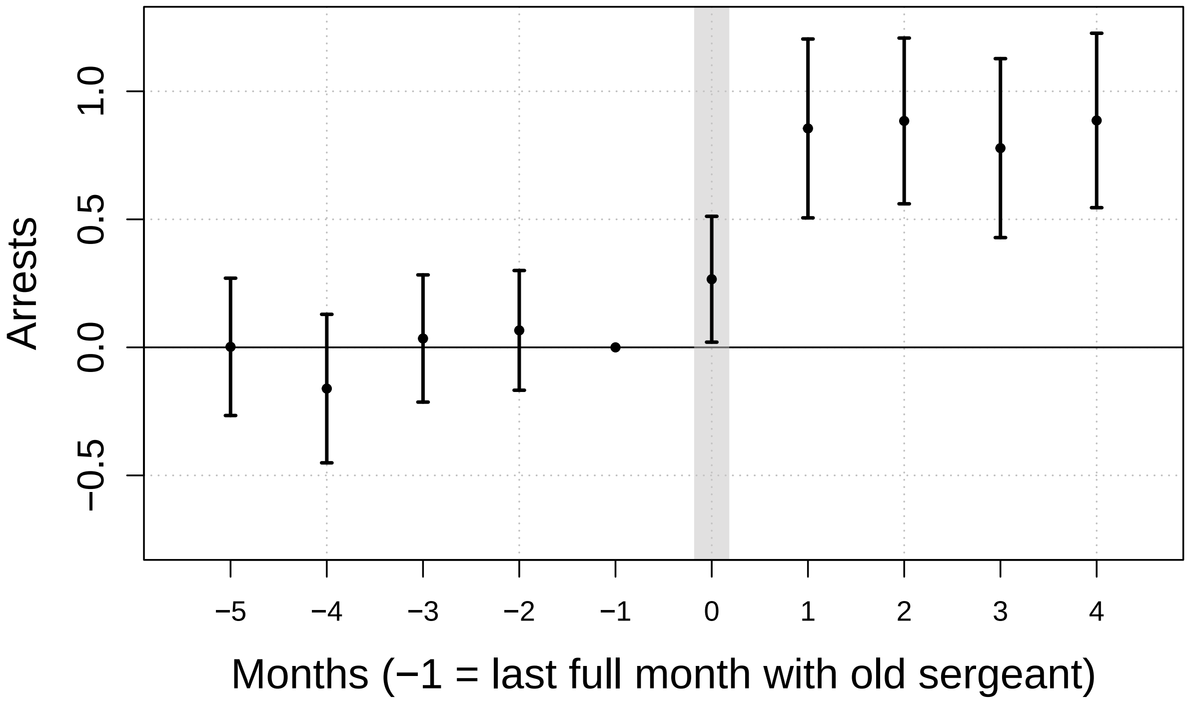
Notes: This figure plots the average number of arrests made by officers in the months around receiving a new sergeant by the magnitude of the sergeant change. In particular, I group sergeants into terciles according to the average number of residual arrests made by their officers throughout the sample. Each line then plots the average residualized arrests made by officers who transition between terciles, where the terciles of the previous and subsequent sergeant are described by "Sergeant Tercile Transition." Arrests are residualized by a second-degree polynomial of officer tenure and officer, sector-watch, and day-off group fixed effects.

Figure 2: Distribution of Sergeant Effects



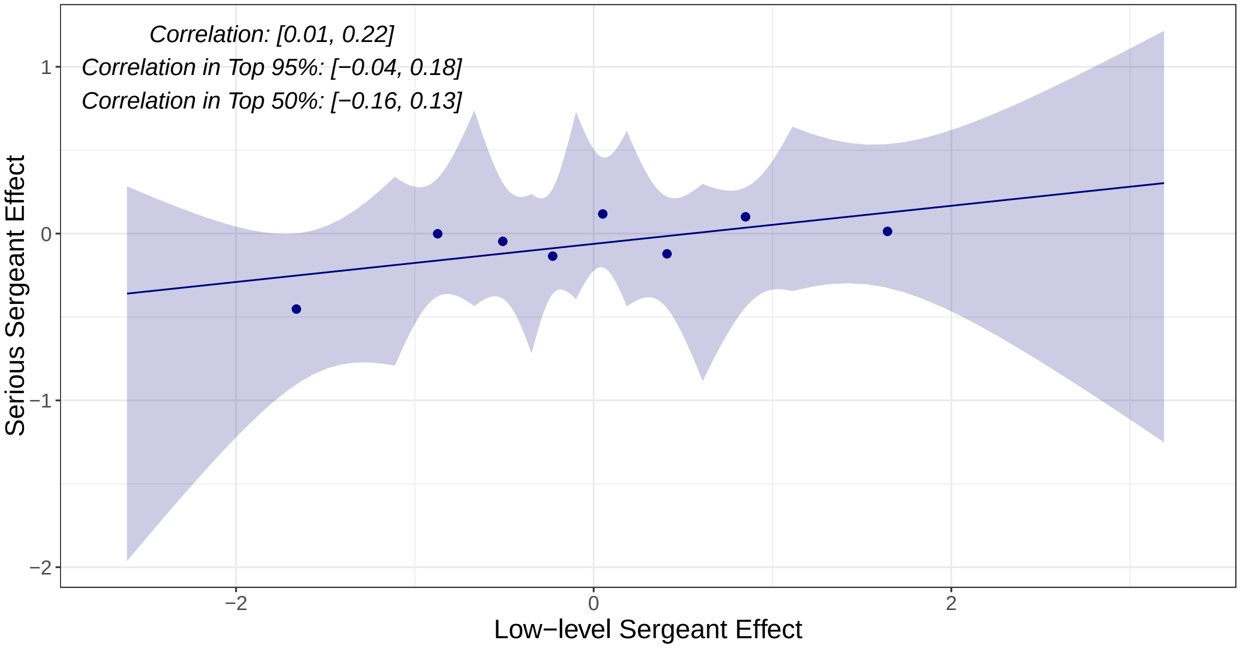
Notes: This figure plots the sergeant effects estimated using the sergeant fixed effects in equation 1. Sergeant effects are interpreted as the number of monthly arrests that an officer makes working under a sergeant, relative to the average sergeant. The solid line represents the raw effects obtained from estimating equation 1 using OLS. The dotted line represents the shrunken effects, which are the raw fixed effects multiplied by the Bayesian shrinkage factor as described in Section 4.

Figure 3: Event Study Coefficients



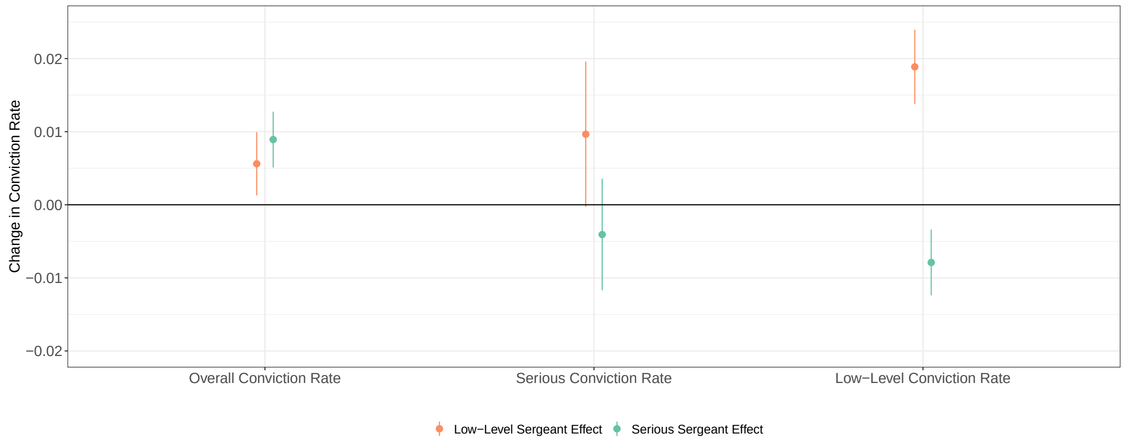
Notes: This figure presents estimates of $\hat{\pi}_1^k$ from the sergeant switching event study model described by equation 6, where k denotes the months around a sergeant switching event. The switch occurs in month 0. Month -1, the last full month an officer spends with their old sergeant, is used as the reference month. The model is estimated using the event study data that are balanced on [-5, 4]. Standard errors are clustered at the officer level.

Figure 4: Relationship between low-level and serious sergeant effects



Notes: This figure displays a binned scatterplot of the relationship between the standardized low-level sergeant effects and standardized serious sergeant effects. Low-level (serious) effects describe sergeant effects on arrests for low-level (serious) crimes, defined as described in Section 3. Low-level and serious sergeant effects are standardized, so that magnitudes are interpreted as standard deviations away from the mean corresponding to each axis. The bins are chosen according to the procedure described by Cattaneo et al. (2024). The blue line is a linear fit and the purple field is a 95% nonparametric confidence band.

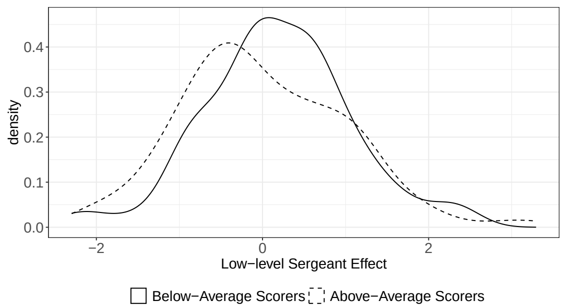
Figure 5: Impact of Sergeant Effects on Conviction Rates



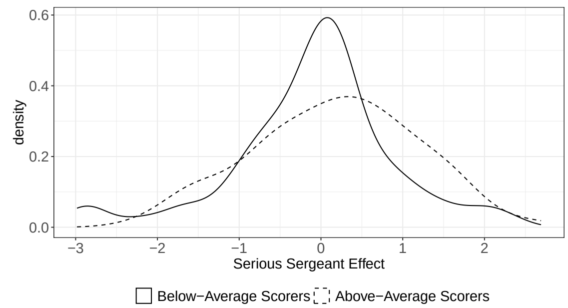
Notes: This figure plots the change in officer conviction rates that results from increasing the low-level and serious sergeant effects by one standard deviation. I calculate the change in conviction rate in two steps. I first regress the number of total and convicted arrests separately on the standardized low-level and serious sergeant effects, along with the model controls as in equation 8. Then, for each of the two sergeant effects, I add the regression coefficients for convicted and total arrests to their respective sample means and take the difference between this ratio and the ratio of the means. For each estimate, I calculate a 95% confidence interval using a bootstrap with 100 resamples. I do this procedure separately for all arrests ('Overall Conviction Rate), serious arrests ('Serious Conviction Rate), and low-level arrests ('Low-Level Conviction Rate').

Figure 6: Sergeant Effects Distributions by Promotional Exam Performance

(a) Low-level effect distribution



(b) Serious effect distribution



Notes: These figures plot distributions of standardized sergeant effects separately by the sergeant's performance on the promotional exam they took to attain the rank of sergeant. I define above-average scorers as sergeants who score above the average for their exam and below-average scorers as those who score below their exam-specific average. Kolmogorov-Smirnov p-value for low-level effects = 0.0343. Kolmogorov-Smirnov p-value for serious effects = 0.0128.

Table 1: Summary Statistics

	Full Sample	Analysis Sample	Event Study Sample
	(1)	(2)	(3)
1. Number of officers	2,067	1,805	833
2. Number of sergeants	387	347	287
3. Number of officers with >1 sgt.	1,856	1,623	833
4. Number of sergeants with >1 off.	384	344	270
5. Mean number of sergeants per off.	5.21	3.97	2.67
6. Mean number of officers per sgt.	27.7	20.6	7.74
7. Total officer-sergeant spells	15,355	8,432	2,247
8. Total switching events	13,288	5,798	1,277
9. Number of sector-watches	105	102	102
10. Mean number of sergeants per sector-watch	8.48	6.95	4.61
11. Arrests mean	3.81	3.80	3.65
SD	3.65	3.64	3.46
12. Low-level arrests mean	2.88	2.87	2.75
SD	3.03	3.02	2.88
13. Serious arrests mean	0.925	0.923	0.897
SD	1.29	1.29	1.26
14. Drug arrests mean	0.315	0.311	0.276
SD	0.931	0.928	0.885
15. Warrant arrests mean	0.771	0.766	0.739
SD	1.35	1.35	1.29
16. Disorderly conduct arrests mean	0.416	0.409	0.370
SD	0.941	0.921	0.839
17. Proactive arrests mean	1.72	1.71	1.61
SD	2.26	2.24	2.13
18. Convicted arrests mean	0.776	0.773	0.712
SD	1.30	1.30	1.21
19. Use of force mean	0.119	0.118	0.114
SD	0.324	0.322	0.318
20. Complaint mean	0.0139	0.0141	0.0151
SD	0.117	0.118	0.122
Number of observations	61,166	49,923	12,770

Notes: The table reports summary statistics for three samples. The Full Sample is the unrestricted sample of all patrol officers. The Analysis Sample contains all patrol officer months that satisfy the restrictions described in Section 3. The Event Study sample contains all officer-sergeant switching events in which the focal officer is observed with the pre-switch sergeant at least 5 months prior to the switch and the post-switch sergeant at least 4 months after the switch. Serious arrests are defined as index arrests as well as domestic violence, fraud, simple assault, and DUI. All other arrests are considered low-level. Drug (warrant/disorderly conduct) arrests are any arrests which contain a drug (warrant/disorderly conduct) charge and do not contain any other higher-level (i.e. serious) charges. An arrest is considered to be convicted if the arrest is matched to a court disposition and not dismissed; this includes guilty findings by judge, jury, or plea. Use of force (complaint) is a binary indicator for any use of force (complaint) taking place in a month.

Table 2: Variance Decomposition

	Raw		Shrinkage		Homosk. Bias-Correction		Heterosk. Bias-Correction	
	Component	% Share	Component	% Share	Component	% Share	Component	% Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Var(y^*)$	11.153	100.00%	11.153	100.00%	11.153	100.00%	11.159	100.00%
$Var(\psi)$	0.559	5.01%	0.382	3.43%	0.379	3.40%	0.378	3.39%
$Var(\theta)$	8.906	79.86%	8.002	71.75%	8.097	72.60%	8.068	72.31%
$Cov(\psi, \theta)$	-0.168	-1.51%	-0.157	-1.41%	-0.0592	-0.53%	-0.0599	-0.54%
$Var(\psi + \theta)$	9.129	81.86%	8.071	72.37%	8.357	74.93%	8.33	74.62%
N sergeants	347		347		347		344	
N officers	1805		1805		1805		1802	

Notes: This table presents the variance decompositions described in equation 4. As described in Section 4, y^* is the number of monthly arrests, residualized on sector-watch, day-off group, and a second-degree polynomial of tenure; ψ is the sergeant fixed effect; θ is the officer fixed effect. All statistics are calculated on data aggregated to the officer-supervisor pair. Columns (1) and (2) report results for the raw fixed effects estimates. Columns (3) and (4) use fixed effects that are multiplied by the Bayesian shrinkage factor, constructed as described in Section 4. Columns (5) and (6) use the bias correction method proposed by Andrews et al. (2008) that assumes homoskedastic error terms. This bias correction is implemented using the 'lfe' package in R (Gaure, 2013) and uses simulation methods to calculate the trace of large matrices, as described in Gaure (2014). As such, I report the average of 100 iterations. Columns (7) and (8) implement the Kline et al. (2020) bias correction method that allows for unrestricted heteroskedasticity in the error terms. This method can only be conducted on the leave-out connected set, which is why the number of sergeants and officers decrease. This implementation adapts the Julia package provided by Kline et al. (2020) and developed by Paul Courcera. The Julia code can be found at https://github.com/HighDimensionalEconLab/VarianceComponentsHDFE.jl.

Table 3: Sergeant Effects by Crime Type

	Serious Crimes			Low-Level Crimes		
	Domestic Violence	Theft	DWI	Drugs	Warrants	Disorderly Conduct
	(1)	(2)	(3)	(4)	(5)	(6)
Low-level Sergeant Effect	-0.0161**	-0.0025	-0.0013	0.1713***	0.1772***	0.1076***
	(0.0080)	(0.0053)	(0.0061)	(0.0238)	(0.0203)	(0.0134)
Serious Sergeant Effect	0.1687***	0.0382***	0.0352***	-0.0360***	0.0755***	0.0350***
	(0.0082)	(0.0064)	(0.0063)	(0.0139)	(0.0148)	(0.0093)
Baseline Controls	✓	✓	✓	✓	✓	✓
Observations	49,923	49,923	49,923	49,923	49,923	49,923
Y mean	0.45308	0.13927	0.10885	0.31138	0.76578	0.40903

Notes: This table reports the estimated coefficients for the sergeant effects in equation 8, using the three most frequent serious and low-level crimes. Serious and low-level crimes are mutually exclusive categories. However, within serious and low-level crimes, an arrest may fall under multiple different criminal charges. The low-level (serious) sergeant effect is given by the standardized Bayes-shrunken sergeant effect on low-level (serious) arrests. The baseline controls include officer fixed effects and the full set of controls used in equation 1. Standard errors are clustered at the officer level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Sergeant Effects on Arrests by Interaction Source

	Officer Initiated Arrests	Call Initiated Arrests
	(1)	(2)
Low-level Sergeant Effect	0.4495***	0.2542***
	(0.0372)	(0.0239)
Serious Sergeant Effect	0.0277	0.2442***
	(0.0248)	(0.0201)
Baseline Controls	✓	✓
Observations	49,923	49,923
Y mean	1.7008	2.0967

Notes: This table reports the estimated coefficients for the sergeant effects in equation 8 using officer-initiated and call-initiated arrests. An arrest is call-initiated if it can be linked to a 911 call in CAD. Otherwise, it is considered officer-initiated. The low-level (serious) sergeant effect is given by the standardized Bayes-shrunken sergeant effect on low-level (serious) arrests. The baseline controls include officer fixed effects and the full set of controls used in equation 1. Standard errors are clustered at the officer level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Sergeant Effects on Call Activity

	Calls Answered	Arrest Probability at Calls	Low-Level Call Arrests	Serious Call Arrests
	(1)	(2)	(3)	(4)
Low-level Sergeant Effect	1.069^{***}	0.0032^{***}	0.2637^{***}	-0.0119
	(0.3701)	(0.0004)	(0.0192)	(0.0101)
Serious Sergeant Effect	2.202^{***}	0.0017^{***}	0.0281^{*}	0.2154^{***}
	(0.3084)	(0.0003)	(0.0153)	(0.0101)
Baseline Controls	✓	✓	✓	✓
Observations	49,923	49,923	49,923	49,923
Y mean	61.415	0.02741	1.4101	0.67718

Notes: This table reports the estimated coefficients for the sergeant effects in equation 8 using various measures of call activity as outcome variables. In column 1, calls answered refer to calls in which the officer is among the first units dispatched to the scene. In column 2, arrest probability at calls is the proportion of calls that an officer answers that result in arrest. I only count the arrest for an officer if they are on the arrest report, since only these arrests will be counted toward their monthly total in the analysis sample. In columns 3 and 4, low-level and serious call arrests are measured in the same way, only counting arrests for which the officer is present on the arrest report. The low-level (serious) sergeant effect is given by the standardized Bayes-shrunken sergeant effect on low-level (serious) arrests. The baseline controls include officer fixed effects and the full set of controls used in equation 1. Standard errors are clustered at the officer level. $^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$.

Table 6: Sergeant Effects and Other Activity

	Use of Force Incidents	Complaints
	(1)	(2)
Low-level Sergeant Effect	0.0203***	0.0040
	(0.0043)	(0.0029)
Serious Sergeant Effect	0.0059*	-0.0019
	(0.0034)	(0.0025)
Baseline Controls	✓	✓
Observations	49,923	49,923
Y mean	0.13813	0.02398

Notes: This table reports the estimated coefficients for the sergeant effects in equation 8 using use of force incidents and complaints as outcome variables. The low-level (serious) sergeant effect is given by the standardized Bayes-shrunken sergeant effect on low-level (serious) arrests. The baseline controls include officer fixed effects and the full set of controls used in equation 1. Standard errors are clustered at the officer level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 7: Sergeant Effect Mechanisms

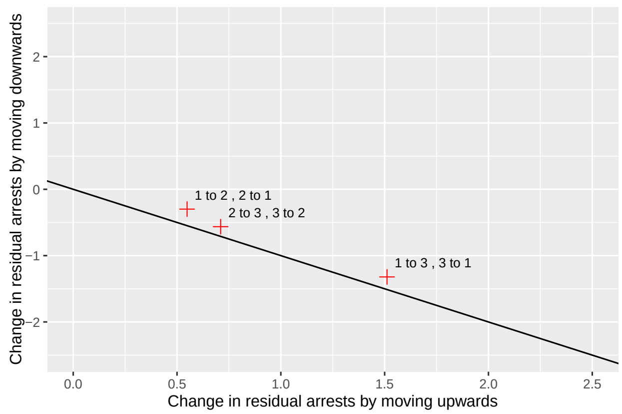
	Leading by Example				Monitoring
	Total Arrests	Serious Arrests	Low-Level Arrests	First-Responder Calls	Subordinate Calls
	(1)	(2)	(3)	(4)	(5)
Low-level Sergeant Effect	0.0777***	0.0062	0.0715***	0.5387*	0.6004**
	(0.0249)	(0.0082)	(0.0192)	(0.2832)	(0.2393)
Serious Sergeant Effect	-0.0161	-0.0047	-0.0114	0.0937	0.2676
	(0.0210)	(0.0068)	(0.0160)	(0.2776)	(0.2230)
Controls	✓	✓	✓	✓	✓
Observations	7,983	7,983	7,983	7,983	7,983
R^2	0.08395	0.03647	0.07874	0.13804	0.19850
Y mean	0.31605	0.08130	0.23475	4.2619	7.8373

Notes: This table presents results from regressing measures of sergeant behavior on the estimated low-level and serious sergeant effects, as described by equation 9. Data are at the sector-watch by month level. Controls include the average estimated low-level and serious officer arrest effects for officers within the unit and sector-watch fixed effects. The outcome variables in each column are: (1) the number of arrests that the unit's supervisor makes in the month, (2) the number of those arrests which are serious, (3) the number of those arrests which are low-level, (4) the number of calls for service that the sergeant is first to respond to, and (5) the number of calls for service that a sergeant responds to in which their subordinates are also present. Standard errors are clustered at the sergeant level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Appendices

A Figures

Figure A.1: Symmetry in moves

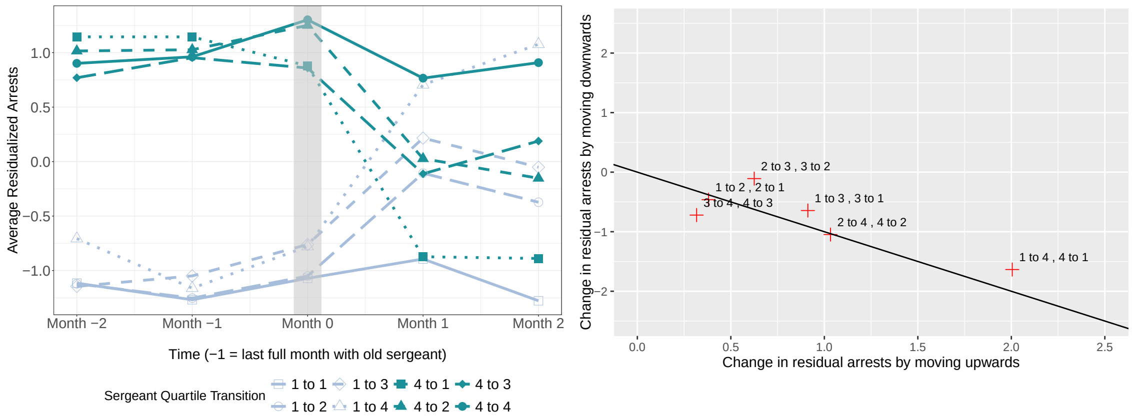


Notes: Each crosshair represents a pair of symmetric moves between sergeants in different terciles of average residualized arrests. Changes in residual arrests are calculated as the average difference between the average number of arrests 2 months after a move and 2 months before a move. Arrests are residualized by officer, sector-watch, and day-off group fixed effects and a second degree polynomial of tenure, as described in Section 4.

Figure A.2: Event study using sergeant quartiles

(a) Event Study Figure

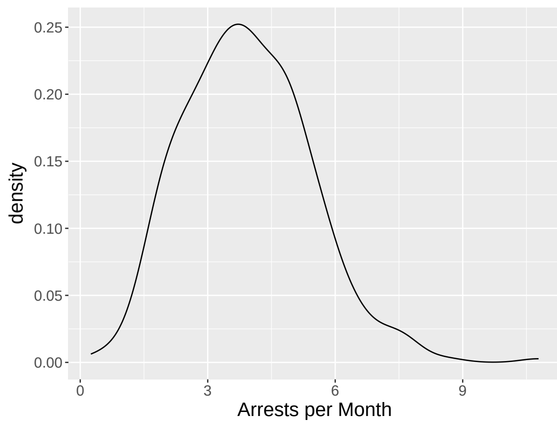
(b) Symmetry Test



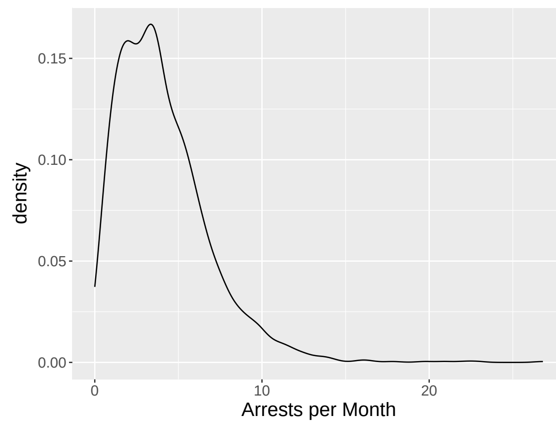
Notes: These figures present the same information as Figures 1 and A.1, instead splitting supervisors into quartiles rather than terciles. To limit the amount of lines in A.2a, I only plot transitions from supervisors in the highest and lowest quartiles. The symmetry test uses all quartile transitions. Note that in A.2a. In A.2b, the crosshairs align roughly with the -45 degree line, suggesting the presence of symmetry across moves in equal and opposite directions.

Figure A.3: Distribution of Arrests

(a) Across Sergeants



(b) Across Officers



Notes: These figures present empirical distributions for monthly arrests. For a given sergeant, I calculate the average number of monthly arrests made by officers who work for them. A.3a then plots the distribution of this average. Then, for each officer, I calculate the average number of arrests they make in a month across all months they are in the sample. I plot the distribution of this average in A.3b.

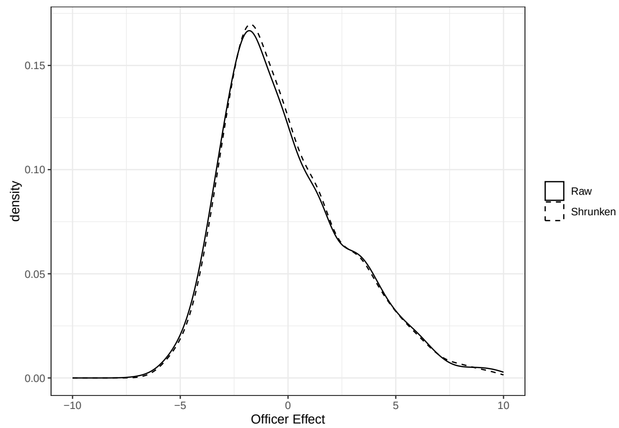


Figure A.4: Distribution of Officer Effects

Notes: This figure plots the officer fixed effects estimated using equation 1. The solid line presents the raw fixed effects, while the dotted line presents the raw effects multiplied by the Bayesian shrinkage factor as described in Section 4.

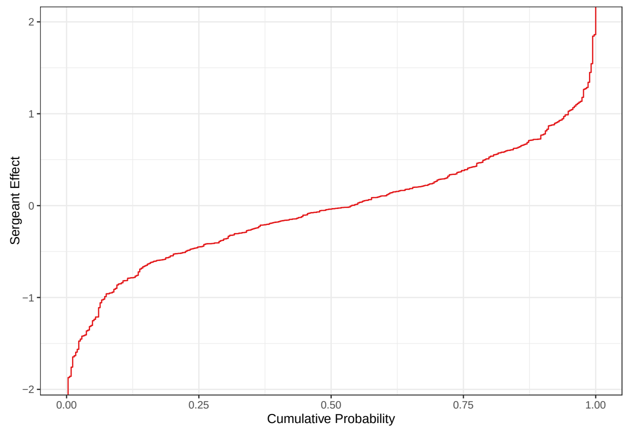


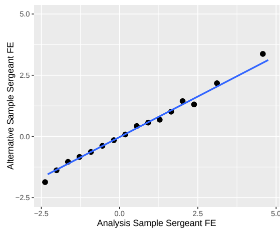
Figure A.5: CDF of Supervisor Fixed Effects

Notes: This figure displays the CDF of the sergeant effects, estimated using the sergeant fixed effects in equation 1 that are multipled by the Bayesian shrinkage factor described in Section 4.

Figure A.6: Robustness to Alternative Sampling Decisions

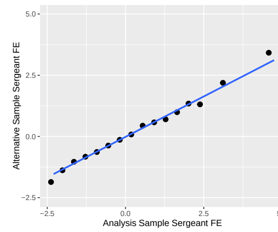
(a) Unrestricted

Correlation = 0.8720



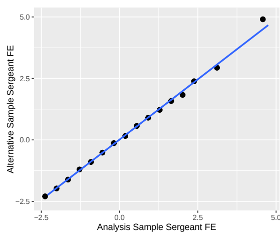
(b) Impute Missing Within Spell

Correlation = 0.8666



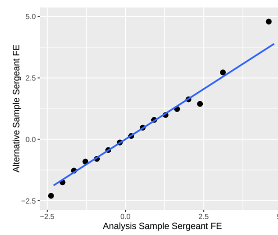
(c) Impute Missing Within Spell, remove everything else

Correlation = 0.9873



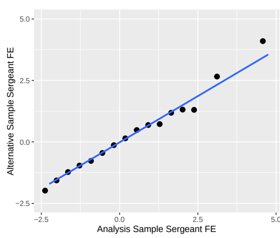
(d) Impute all temporary assignments

Correlation = 0.9362



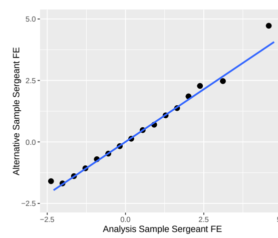
(e) Keep temporary, remove missing

Correlation = 0.9153



(f) Keep missing, remove temporary

Correlation = 0.9311



Notes: This figure presents the correlation between sergeant fixed effects under different sampling restrictions. (a) makes no sample restrictions, (b) imputes missing observations within a continuous sergeant spell and keeps any other missing sergeant observations, (c) is the same as (b) but other observations with unknown sergeants are removed, (d) imputes temporary one-off assignments with different sergeants using an officer's permanent sergeant, (e) keeps all of the temporary assignments but removes all months with an unknown sergeant, and (f) keeps months with an unknown sergeant but removes the temporary assignments.

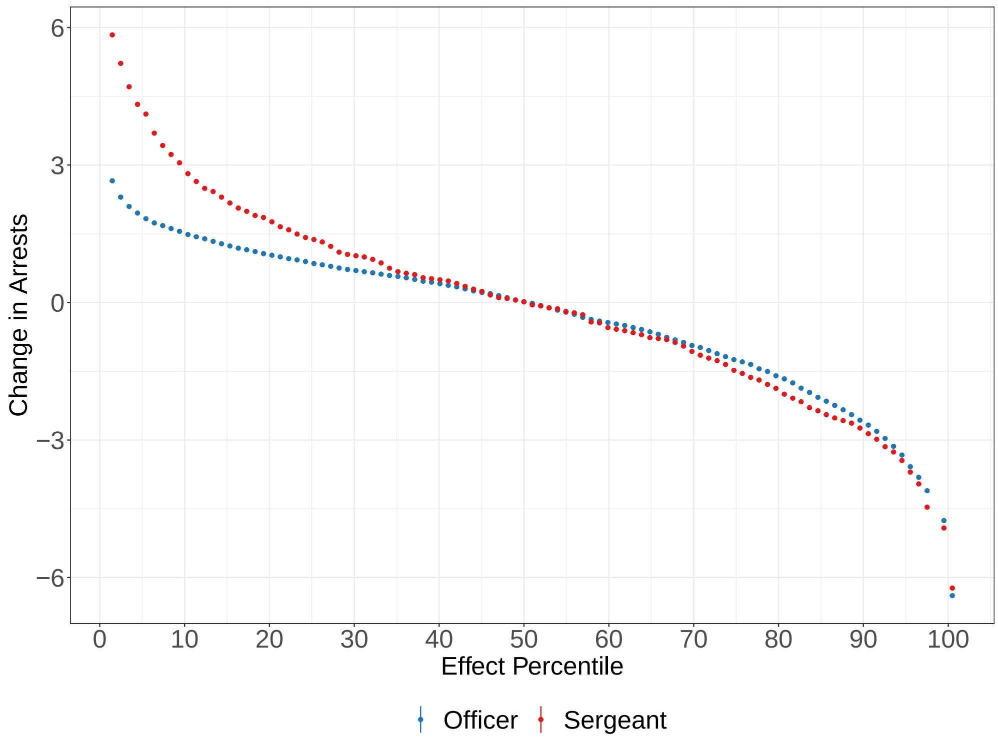


Figure A.7: Effect of a one-month replacement with the fiftieth percentile

Notes: This figure plots the calculated effect of one-month replacements of sergeants (in red) and officers (in blue) with a fiftieth percentile employee from the effects distribution. Each sergeant is placed into their percentile in the effects distribution and the change in arrests that would be produced from replacing them with a fiftieth sergeant is calculated by subtracting each sergeant's effect from the fiftieth percentile sergeant effect, and multiplying by multiplying by the average number of officers managed in a month (6.33). I then plot the change in arrests against each percentile by averaging over all sergeants within that percentile. The change in arrests for officers is calculated identically, except I do not multiply by 6.33. Neither of the empirical effect distributions are mean 0 because the Empirical Bayes shrinkage procedure introduces a small degree of bias in order to reduce the mean square error of the fixed effects.

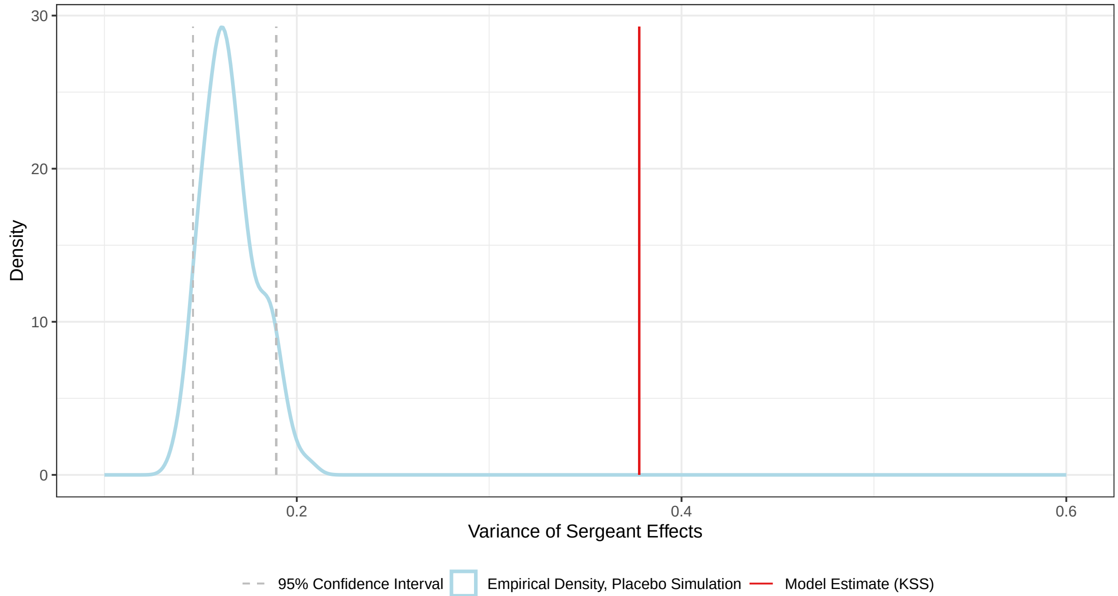


Figure A.8: Placebo Test, Sergeants randomly reassigned to officers

Notes: This figure depicts the results of placebo tests that randomly reallocate sergeants to officers, preserving the number of unique officers managed for each sergeant. For every reallocation, I estimate equation 1 and report the resulting (unadjusted) variance of sergeant effects. The empirical density, in light blue, plots the density of variance estimates for 100 reallocations. The dashed lines denote the 95% confidence interval of the placebo variance estimates. The red vertical line denotes my main estimate of the variance in sergeant effects, adjusted for measurement error using the KSS method described in Section 4.2

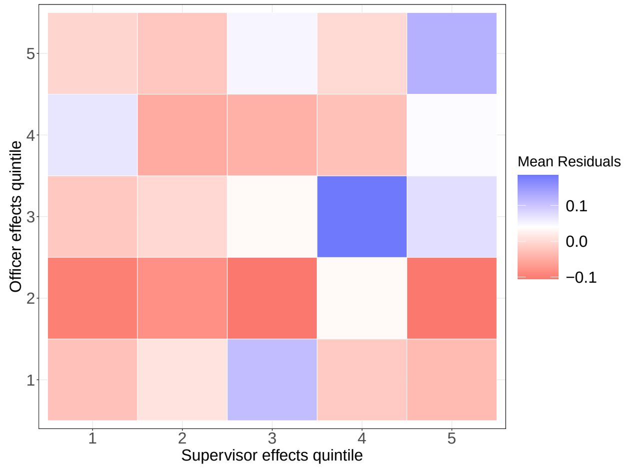
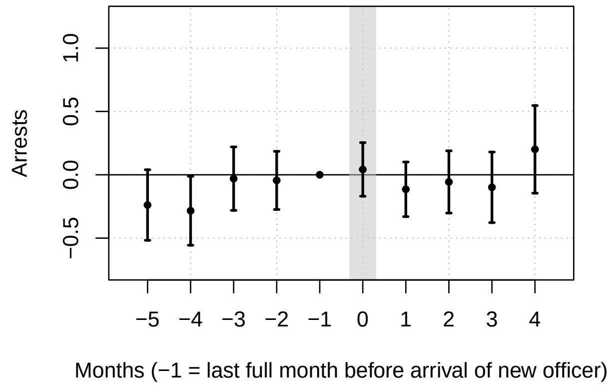


Figure A.9: Residuals by quintile of officer and sergeant arrest effects

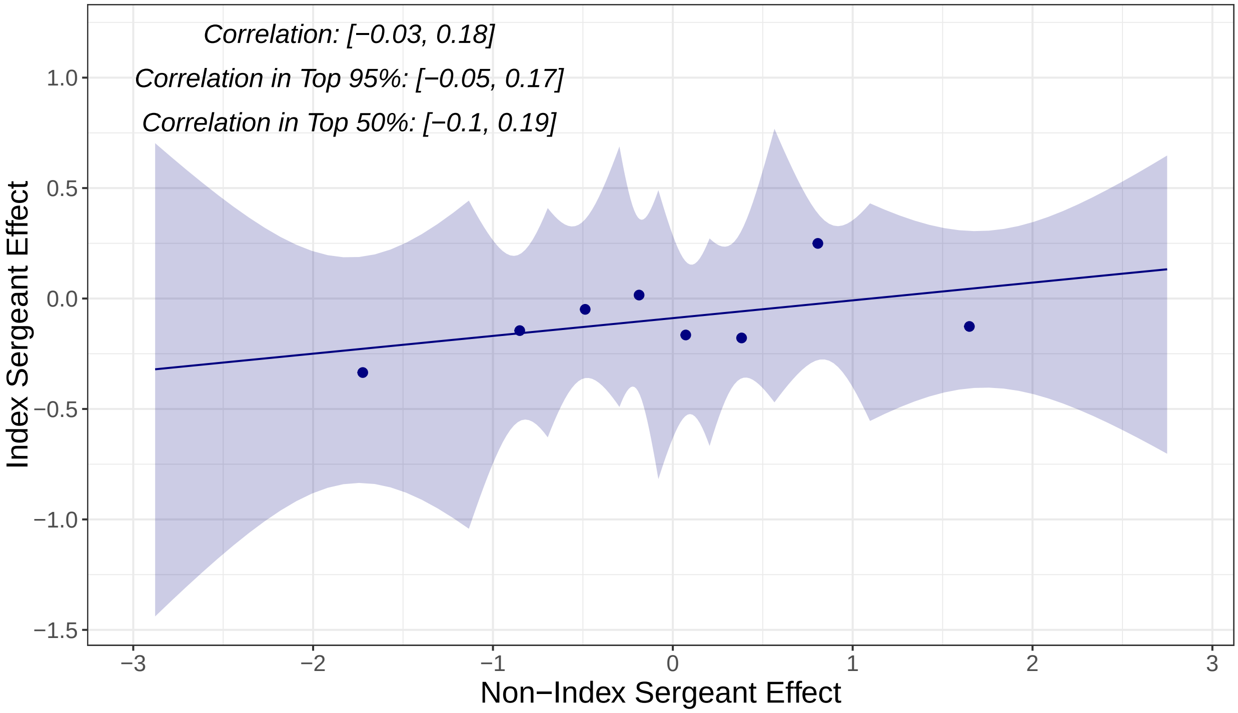
Notes: This figure reports the average residuals by quintiles of officer and sergeant (supervisor) arrest effects. Darker blue indicates more positive residuals and darker red indicates more negative residuals.

Figure A.10: Arrests Made by Incumbent Officers



Notes: This figure plots the event study coefficients from equation 7. For an officer switching event e in which officer i switches from supervisor $\underline{j}$ to supervisor $\bar{j}$, incumbent officers are those who work with supervisor $\bar{j}$ for 5 months before the event and and 4 months after the event. The x-axis indicates months relative to officer i's switch. Standard errors are clustered at the level of the switching officer.

Figure A.11: Relationship between non-index and index sergeant effects



Notes: This figure displays a binned scatterplot of the relationship between the standardized non-index sergeant effects and standardized index sergeant effects. Non-index (index) sergeant effects capture the effects of a sergeant on non-index (index) arrests. Index arrests are made for the FBI's Part I Index Crimes: murder, aggravated assault, rape, robbery, motor vehicle theft, burglary, theft, and arson. Non-index arrests are arrests for all other crimes. The bins are chosen according to the procedure described by Cattaneo et al. (2024). The blue line is a linear fit and the purple field is a 95% nonparametric confidence band.

Figure A.12: Distribution of Sergeants between Serious and Low-level effects

Tercile of Serious Effect			
3	11.2%	11.2%	11.0%
2	7.8%	12.7%	12.7%
1	14.4%	9.2%	9.8%
	1	2	3

Tercile of Low-level Effect

Notes: This figure displays the percentage of sergeants within each tercile grouping of (Bayes-shrunken) low-level and serious sergeant effects.

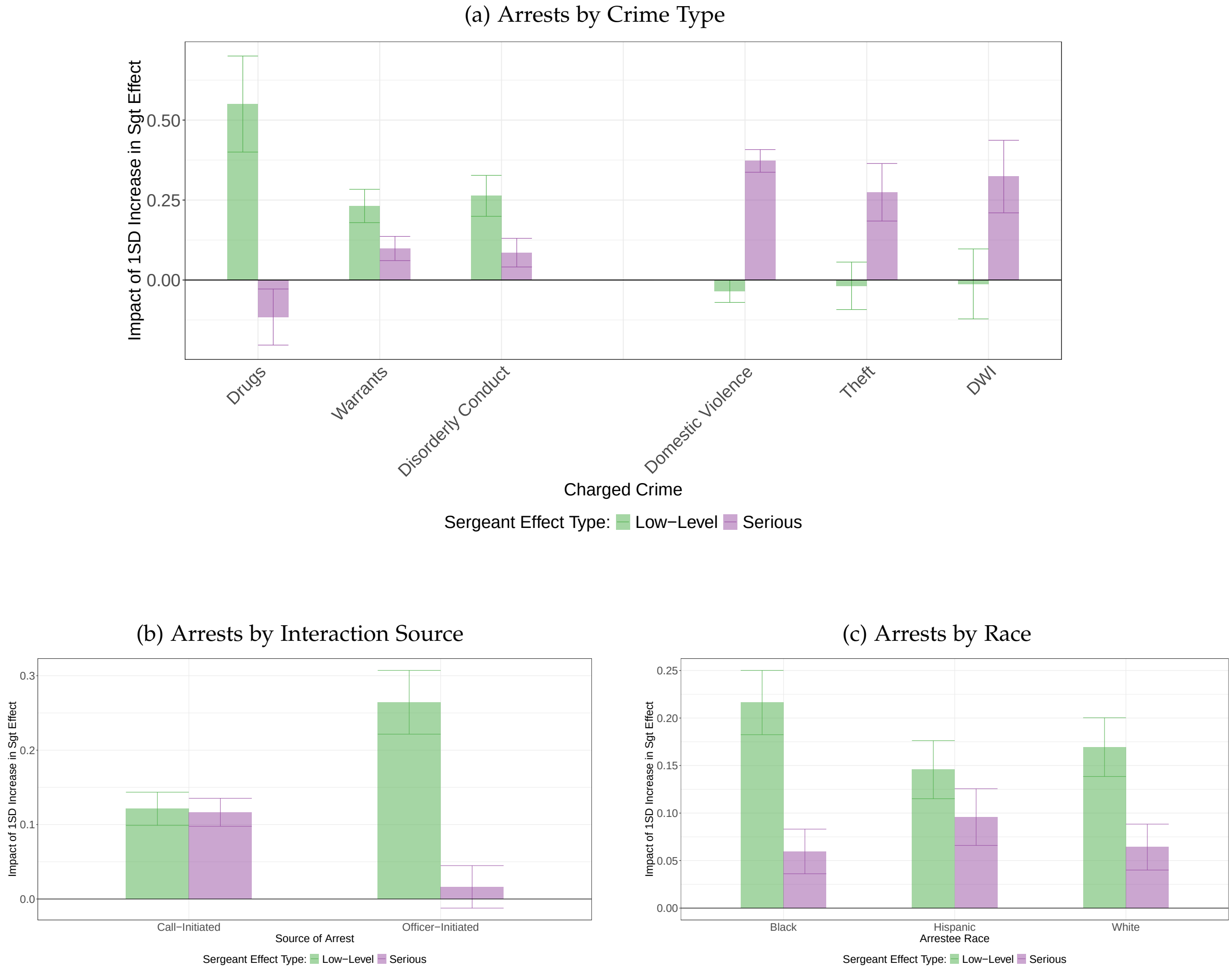


Figure A.13: Drivers of Sergeant Effects

Notes: This figure depicts estimates of coefficients α_L^c and α_S^c from equation 8, for arrests of various categories c. Estimates of α_L^c (α_S^c) are interpreted as the impact of increasing the low-level (serious) sergeant effect by 1 standard deviation. Green lines represent α_L^c estimates and purple lines represent α_S^c estimates. Arrest categories are given on the x-axis of each figure. For each category, the α_L^c and α_S^c are normalized to the average number of arrests in that category. 95% confidence intervals are depicted for each of the estimates, calculated using standard errors clustered at the officer level. Raw regression estimates for each of figures can be found in Table 3 (a), Table 4 (b), and Table B.8 (c).

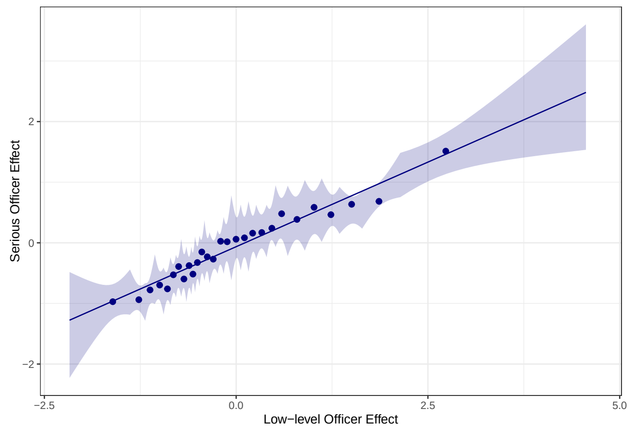
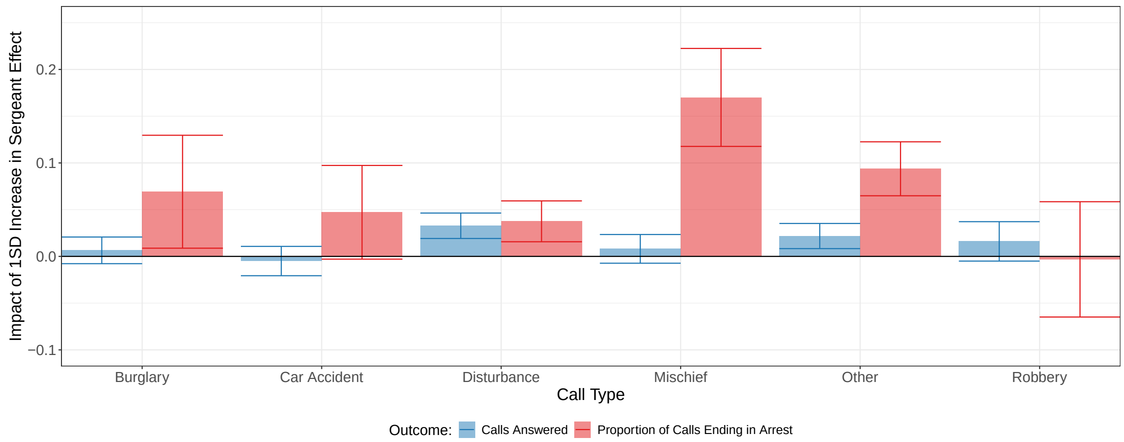


Figure A.14: Low-level and serious officer effects

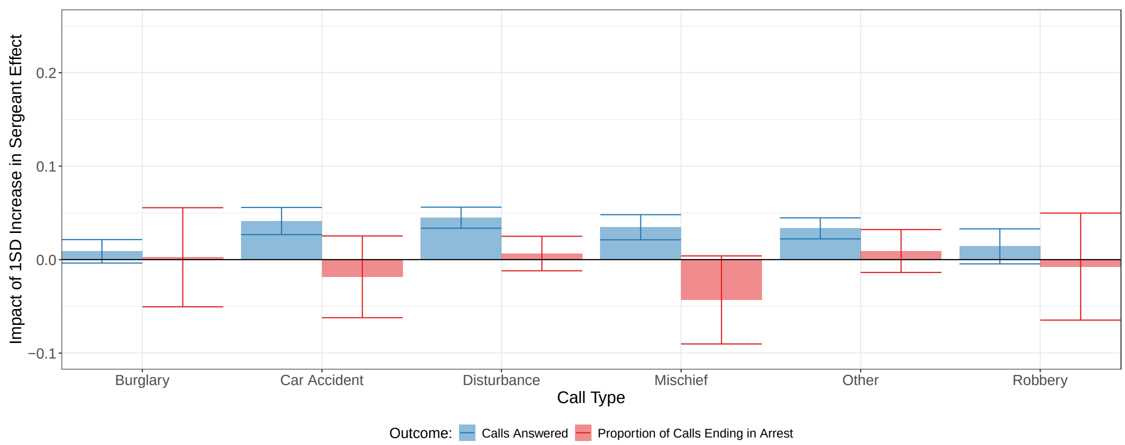
Notes: This figure displays a binned scatterplot of the relationship between the standardized low-level officer effects and standardized serious officer effects. The bins are chosen according to the procedure described by (Cattaneo et al., 2024). The blue line represents a linear fit and the purple field gives a 95% nonparametric confidence band.

Figure A.15: Impact of sergeant effects on calls and arrests by call type

(a) Low-level Sergeant Effects

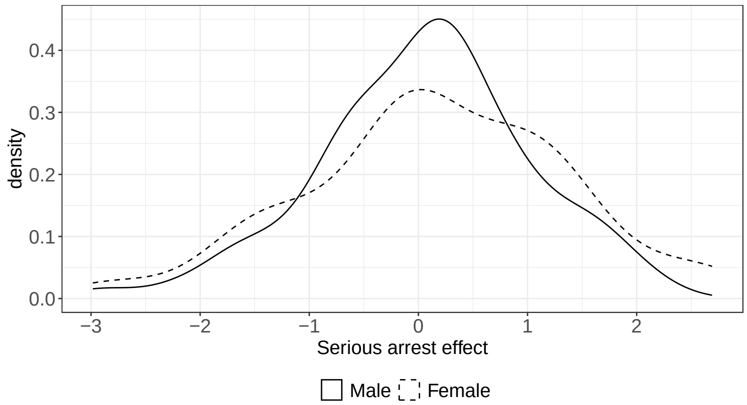


(b) Serious Sergeant Effects



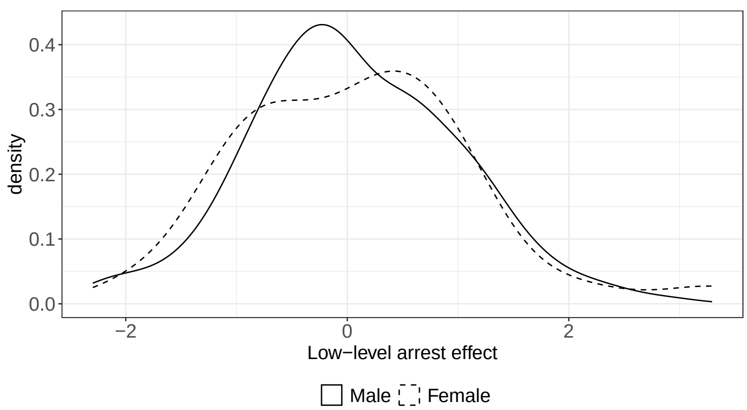
Notes: This figure displays estimates and 95% confidence intervals from regressions of calls answered (blue) and proportion of calls ending in arrests (red) on low-level and serious sergeant effects, along with the controls specified in equation 8. Panel (a) shows the estimates for the low-level sergeant effects and panel (b) shows the estimates for the serious sergeant effects. All estimates are normalized to the mean of each respective outcome.

Figure A.16: Serious Sergeant Effects Distribution by Gender



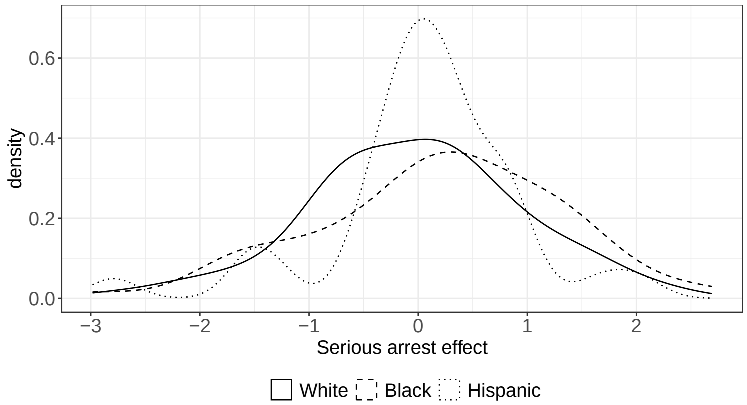
Notes: This figure plots the distribution of serious sergeant effects by sergeant gender. Kolmogorov-Smirnov p-value = 0.415.

Figure A.17: Low-level Sergeant Effects by Gender



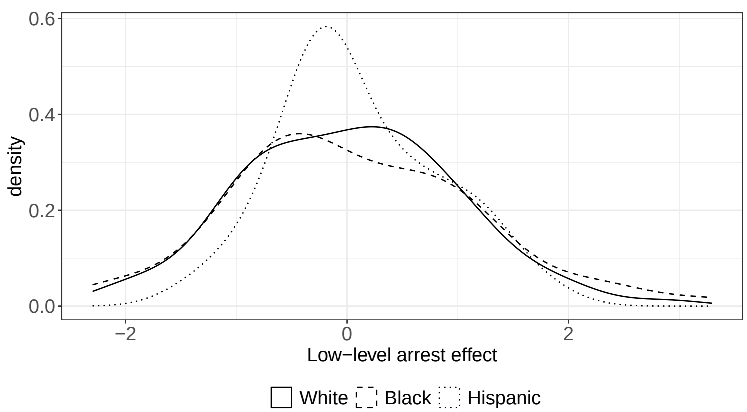
Notes: This figure plots the distribution of low-level sergeant effects by sergeant gender. Kolmogorov-Smirnov p-value = 0.817.

Figure A.18: Serious Sergeant Effects by Race



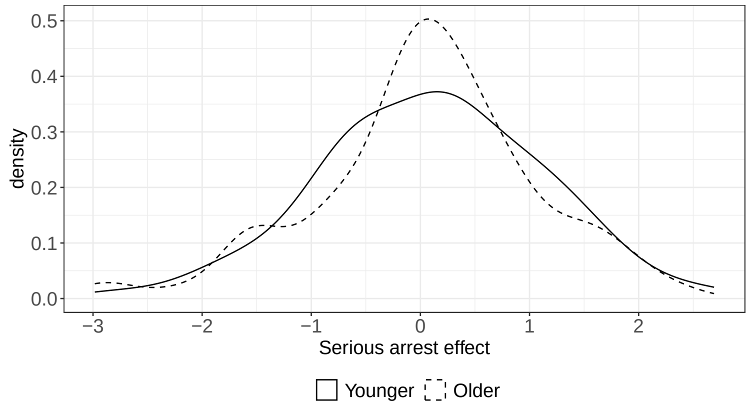
Notes: This figure plots the distribution of serious sergeant effects by sergeant race. Black/Hispanic Kolmogorov-Smirnov p-value = 0.239. Black/White Kolmogorov-Smirnov p-value = 0.219. White/Hispanic Kolmogorov-Smirnov p-value = 0.130.

Figure A.19: Low-level Sergeant Effects by Race



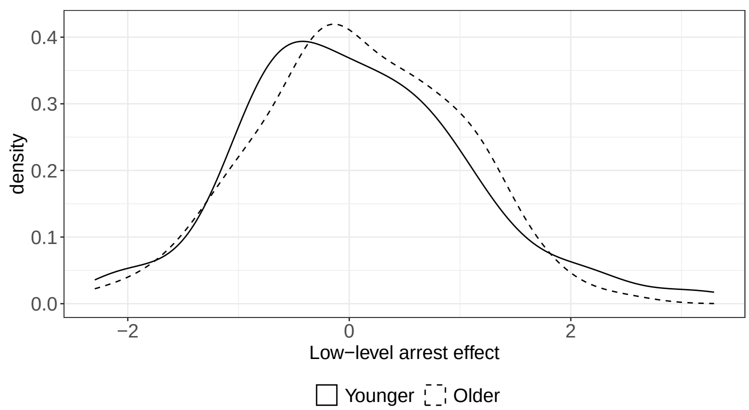
Notes: This figure plots the distribution of low-level sergeant effects by sergeant race. Black/Hispanic Kolmogorov-Smirnov p-value = 0.218. Black/White Kolmogorov-Smirnov p-value = 0.241. White/Hispanic Kolmogorov-Smirnov p-value = 0.855.

Figure A.20: Serious Sergeant Effects by Age at the Time of Exam



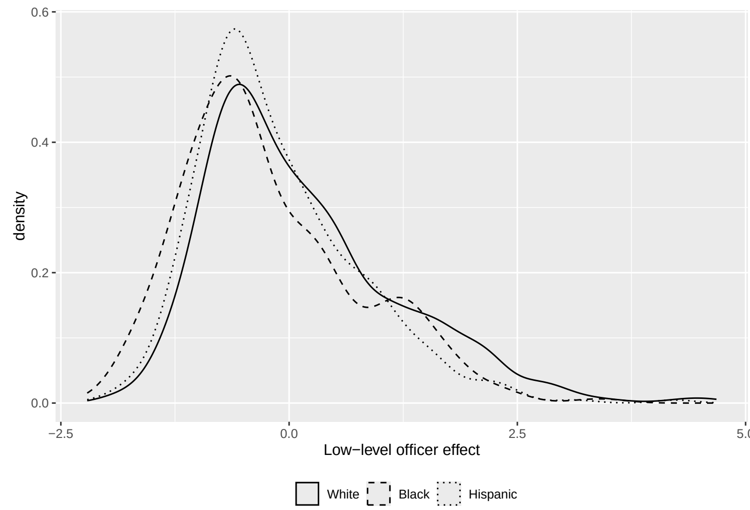
Notes: This figure plots the distribution of serious sergeant effects by age at the time of the promotional exam. Older sergeants are those who are above the average age across all exams in my sample, while younger sergeants are below the average. Kolmogorov-Smirnov p-value = 0.587.

Figure A.21: Low-level Sergeant Effects by Age at the Time of Exam



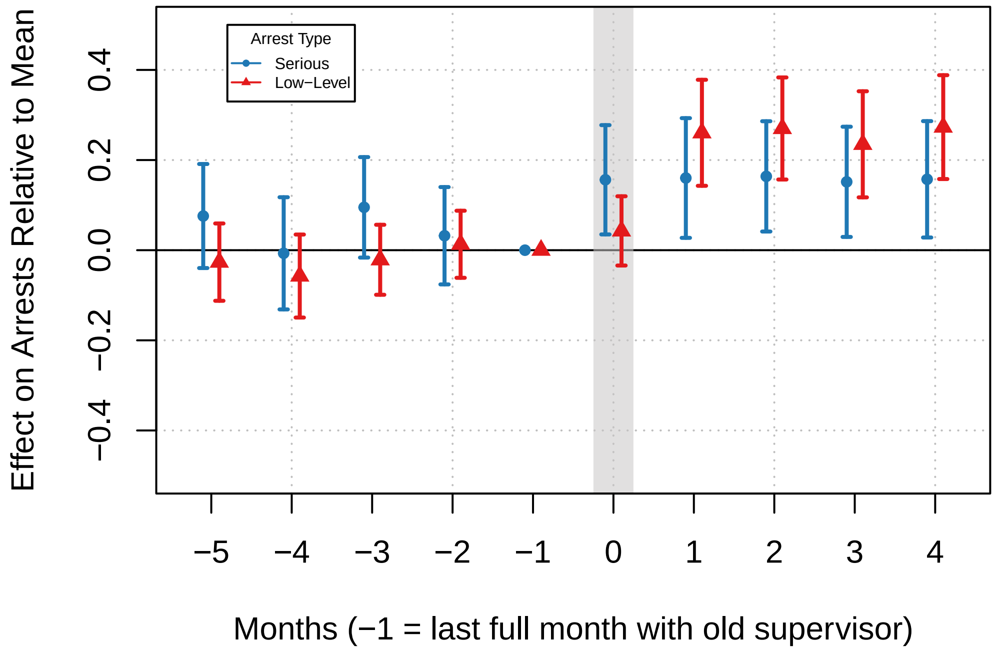
Notes: This figure plots the distribution of low-level sergeant effects by age at the time of the promotional exam. Older sergeants are those who are above the average age across all exams in my sample, while younger sergeants are below the average. Kolmogorov-Smirnov p-value = 0.587.

Figure A.22: Low-level Officer Effects by Race



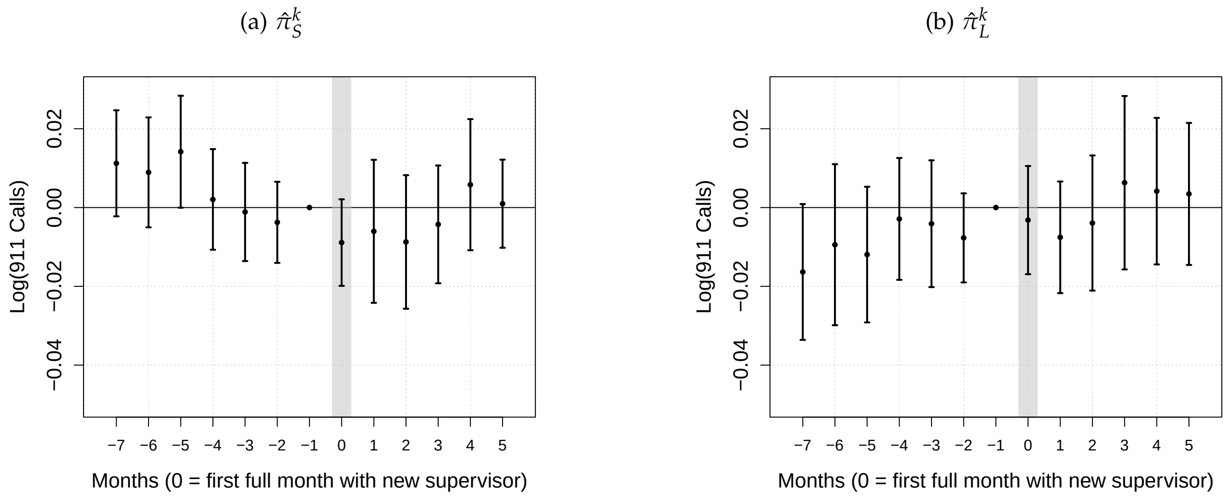
Notes: This figure plots the distribution of low-level officer effects by officer race. Black/Hispanic Kolmogorov-Smirnov p-value = 0.047. Black/White Kolmogorov-Smirnov p-value = 0.000. White/Hispanic Kolmogorov-Smirnov p-value = 0.003.

Figure A.23: Event Study by Arrest Type



Notes: This figure plots event study coefficients for equation 6, separately for models that use serious arrests (defined as index arrests as well as domestic violence, fraud, simple assault, and DUI) and low-level arrests as the dependent variable. Serious arrest results are given in blue and low-level arrest results are given in red. Month -1, the last full month that the officer spends with the old supervisor, is the reference month in all specifications. For each severity level, the effects are normalized to the average number of arrests in month -1. The model is estimated using the event study data that are balanced on [-5, 4].

Figure A.24: Sergeant effects on crime



Notes: The figures report for the event study coefficients estimated in equation 15. Panel (a) plots the coefficients for the serious sergeant effects and Panel (b) plots the coefficients for the low-level sergeant effects. Standard errors are clustered at the sector-watch level.

B Tables

Table B.1: Data Sources

Data Source (1)	Variables (2)
Computer Aided Dispatch Entries (2014-2019)	Assignments/911 Calls
Arrest Reports (2014-2019)	Number of arrests
Charge Reports (2014-2019)	Type of arrest
Use of Force Reports (2014-2019)	Use of force incidents
Civilian Complaints (2014-2019)	Number of complaints
Disposed Cases, Dallas County DA (2014-2019)	Conviction
Various Personnel Records (2014-2019)	Watch/day-off group/promotion dates
Sergeants Exam Results (2012; 2014; 2018)	Composite Promotional Score

Notes: This table describes each of the data sources used to construct the analysis sample.

Table B.2: Event-study around sergeant switches

k	Total Arrests	Serious Arrests	Low-Level Arrests
	(1)	(2)	(3)
-5	0.0020	0.0732	-0.0712
	(0.1366)	(0.0568)	(0.1175)
-4	-0.1609	-0.0067	-0.1542
	(0.1477)	(0.0613)	(0.1261)
-3	0.0347	0.0919*	-0.0572
	(0.1265)	(0.0549)	(0.1064)
-2	0.0664	0.0310	0.0354
	(0.1191)	(0.0532)	(0.1020)
0	0.2661**	0.1510**	0.1151
	(0.1252)	(0.0597)	(0.1052)
1	0.8550***	0.1548**	0.7002***
	(0.1779)	(0.0654)	(0.1611)
2	0.8843***	0.1583***	0.7260***
	(0.1649)	(0.0602)	(0.1551)
3	0.7782***	0.1466**	0.6316***
	(0.1780)	(0.0602)	(0.1612)
4	0.8862***	0.1520**	0.7342***
	(0.1735)	(0.0636)	(0.1578)
Observations	12,770	12,770	12,770
Y mean	3.6488	0.89742	2.7514
Pre-trends F stat	0.8473	1.5706	0.6234
p-value	0.8473	0.1791	0.6458

Notes: This table presents the event-study coefficients used to make Figure 3. The regressions use switching events that are balanced around 5 sample months prior to the move and 4 sample months after the move. Month -1 is the reference point and the switch occurs at in month 0. The pre-trends F statistic is calculated from an F-test of joint significance of the coefficients for which $k < -1$. Standard errors are clustered at the officer level.

Table B.3: Analysis of Variance

	Arrests				
	(1)	(2)	(3)	(4)	(5)
R^2	0.166379	0.516861	0.527611	0.202556	0.623947
Adjusted R^2	0.164421	0.497526	0.505139	0.195091	0.559736
Controls	✓	✓	✓	✓	✓
Officer FE		✓	✓		
Sergeant FE			✓	✓	
Sergeant-by-Officer FE					✓
Observations	49,923	49,923	49,923	49,923	49,923

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: This table reports R^2 and adjusted R^2 for models that vary the included fixed effects. Controls include a second degree polynomial of officer tenure, and sector-watch and day-off group fixed effects.

Table B.4: Trends in crime do not predict changes in sergeant effects

	All Months		Months With Movers	
	$E[\Delta\hat{\psi}_{out}]$	$E[\Delta\hat{\psi}_{in}]$	$E[\Delta\hat{\psi}_{out}]$	$E[\Delta\hat{\psi}_{in}]$
	(1)	(2)	(3)	(4)
$Log(911Calls)_{-1}$	-0.0056	0.0285	-0.0052	0.1420
	(0.0401)	(0.0424)	(0.1630)	(0.1302)
$Log(911Calls)_{-2}$	-0.0022	0.0029	-0.0052	0.0053
	(0.0490)	(0.0525)	(0.1818)	(0.1700)
$Log(911Calls)_{-3}$	0.0018	-0.0799	-0.0609	-0.3495*
	(0.0535)	(0.0526)	(0.1635)	(0.1824)
$Log(911Calls)_{-4}$	-0.0647	0.0335	-0.2065	0.1106
	(0.0488)	(0.0546)	(0.1924)	(0.1488)
$Log(911Calls)_{-5}$	0.0704*	-0.0152	0.1350	0.0411
	(0.0423)	(0.0475)	(0.1884)	(0.1653)
Observations	5,525	5,525	1,387	1,424
Y mean	0.00738	0.00401	0.02938	0.01555
Joint F p-value	0.62913	0.58831	0.83440	0.39364
Sector-Watch fixed effects	✓	✓	✓	✓

Notes: This table examines the correlation of crime trends with sergeant switches. Regressions are performed at the sector-watch by month level. For each sector-watch and month, I calculate the average change in sergeant effects separately for officers who move in and out of the sector-watch in that month. The table reports results for a regression of the sergeant effect changes on the natural logarithm of 911 calls originating in that sector-watch in each of the 5 months prior to the focal month. Regressions include fixed effects for the sector-watch. Dependent variables are the average change in the sergeant effects for out-movers (columns 1 and 3) and in-movers (columns 2 and 4). Out-movers are officers who leave the sector-watch and in-movers are officers who join the sector-watch in a given month. Columns 1 and 2 use all monthly observations for each sector-watch. Columns 3 and 4 only use the monthly observations in which at least one out-move (column 3) or one in-move (column 4) occurs. Standard errors are clustered at the sector-watch level.

Table B.5: Crime Trends Don't Predict Sergeant Effects

	Sergeant Effect
	(1)
$Log(911Calls)_{-1}$	0.0712
	(0.0455)
$Log(911Calls)_{-2}$	0.0254
	(0.0319)
$Log(911Calls)_{-3}$	-0.0352
	(0.0297)
$Log(911Calls)_{-4}$	-0.0544
	(0.0347)
$Log(911Calls)_{-5}$	-0.0275
	(0.0485)
Observations	5,525
Y mean	-0.03592
Joint F p-value	0.21720
Sector-Watch fixed effects	✓

Notes: This table presents results for a regression of sergeant effects on the natural logarithm of 911 calls in the sergeant's assigned sector-watch between 1 and 5 months before the focal month. The data are at the sergeant-by-month level. Standard errors are clustered at the sergeant level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.6: Drug Arrest Types

	Possession Arrests	Distribution Arrests
	(1)	(2)
Low-level Sergeant Effect	0.1543^{***}	0.0163^{***}
	(0.0210)	(0.0041)
Serious Sergeant Effect	-0.0314^{**}	-0.0051^{***}
	(0.0132)	(0.0018)
Baseline Controls	✓	✓
Observations	49,923	49,923
Y mean	0.28955	0.01839

Notes: The table reports results for a regression of possession (column 1) and distribution (column 2) arrests on low-level and serious sergeant effects, along with standard model controls from equation 8. I classify a drug arrest as "possession" if the charge description only mentions possession and not sale or manufacturing. If the charge description mentions sale or manufacturing, then the drug arrest is classified as "distribution," so that the two categories are mutually exclusive. The effect sizes for all drug arrests are 0.1713 for low-level sergeant effects and -0.0361 for serious sergeant effects, taken from Table 3. 99.6% of drug arrests are either for possession or distribution. Standard errors are clustered at the officer level. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

Table B.7: Sergeant effects and officer overtime activities

	Overtime Calls	Overtime Low-level Arrests	Overtime Serious Arrests
	(1)	(2)	(3)
Low-level Sergeant Effect	0.6303***	0.0312***	-0.0011
	(0.1646)	(0.0057)	(0.0031)
Serious Sergeant Effect	0.9050***	0.0103*	0.0219***
	(0.1464)	(0.0056)	(0.0037)
Baseline Controls	✓	✓	✓
Observations	49,923	49,923	49,923
Y mean	6.0084	0.11327	0.04926

Notes: This table presents results for a regression of overtime calls and arrests on low-level and serious sergeant effects, along with standard model controls from equation 8. I call a call or arrest overtime if it occurs outside of the officer's shift hours listed in the assignments data. Standard errors are clustered at the officer level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.8: Sergeant impacts by race

	Black Arrests	Hispanic Arrests	White Arrests
	(1)	(2)	(3)
Low-level Sergeant Effect	0.4172***	0.1469***	0.1370***
	(0.0333)	(0.0157)	(0.0127)
Serious Sergeant Effect	0.1150***	0.0966***	0.0519***
	(0.0231)	(0.0154)	(0.0100)
Observations	49,923	49,923	49,923
Y mean	1.9291	1.0082	0.80869

This table reports the estimated coefficients for the sergeant effects in equation 8, using arrestee race as the dependent variable. The baseline controls include officer fixed effects and the full set of controls used in equation 1. Standard errors are clustered at the officer level. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

Table B.9: Sergeant Effects and Conviction Rates

	Difference in Conviction Ratio	Difference in Serious Conviction Ratio	Difference in Low-Level Conviction Ratio
	(1)	(2)	(3)
Low-level Sergeant Effect	0.0056**	0.0097*	0.0189***
	(0.0022)	(0.0051)	(0.0138)
Serious Sergeant Effect	0.0089***	-0.0040	-0.0079***
	(0.0019)	(0.0039)	(0.0230)
Baseline Controls	✓	✓	✓
Observations	49,923	49,923	49,923
Mean ratio	0.2046	0.4282	0.1322

Notes: This table reports the changes in officer conviction rates that results from increasing the low-level and serious sergeant effects by one standard deviation, which are plotted in Figure 5. I calculate the change in conviction rate in two steps. I first regress the number of total and convicted arrests separately on the standardized low-level and serious sergeant effects, along with the model controls as in equation 8. Then, for each of the two sergeant effects, I add the regression coefficients for convicted and total arrests to their respective sample means and take the difference between this ratio and the ratio of the means. For each estimate, I calculate a 95% confidence interval using a bootstrap with 100 resamples. I do this procedure separately for all arrests ('Overall Conviction Rate), serious arrests ('Serious Conviction Rate), and low-level arrests ('Low-Level Conviction Rate'). $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

C Constructing Sergeant Assignments

The Computer Aided Dispatch (CAD) system used by the Dallas Police Department stores assignment indicators for every sworn employee who is assigned to a call. These assignment indicators are known internally as "element numbers." Element numbers are assigned every day to each separate patrol car and describe the watch and beat assignment of the car. Beats are smaller geographic sectors within patrol sectors that individual officers are assigned to patrol.

Watches are described by letters A-F, where A/B/C denote overnight/day/evening watches and D/E/F are variants for day/overnight/evening that allow for multiple units to be assigned to one beat at a time depending on department needs. Beats are given by a 3-digit numeric. Thus, an example of an element number within CAD is A135, which means that the officer is working the overnight shift patrolling beat 135. The first and second digits of the beat code identify the sector in which the beat is located. Returning to the previous example, beat 135 is part of sector 130.

Sergeants are given element numbers that denote the sector and watch to which they are assigned. The sergeant for an officer with the element number A135 has the element number A130. In the case of variant units within a sector, there will be one sergeant in charge of each unit. That is, an officer in the variant overnight unit E135 would have a sergeant with the element number E130. I use this pattern in the element numbers to identify the most common sector-watch assignments for officers and sergeants within each month of the data, as described in Section 3.

The assignments that I construct exclude officer spells in specialty patrol units whose element number does not match a geographic sector. Based on conversations with DPD, these units perform distinct duties from regular patrol officers, as evidenced by the fact that they are not assigned to specific geographic beats. DPD did not provide me with the specific details of the job duties related to these assignments for reasons related to officer safety. But, consistent with a separate and distinct role, officers in these specialty units exhibit significant more variation in arrests than regular patrol officers and sergeants cannot be reliably identified for these units.

D Empirical Bayes Shrinkage

The raw supervisor fixed effects are estimated with error. Suppose that the estimates are given by:

$$\hat{\psi}_j = \psi_j + \epsilon_j, \tag{10}$$

where $\psi_j \sim \mathcal{N}(0, \sigma^2_{\psi})$, $\epsilon_j \sim \mathcal{N}(0, \sigma^2_{\epsilon_j})$, and ψ_j and ϵ_j are independently distributed across the population of 347 supervisors. The mean of the supervisor fixed effects is 0 by construction, since the true mean is unidentified in the model. Under these distributional assumptions, we have that

$$\hat{\psi}_j | \psi_j \sim \mathcal{N}(\psi_j, \sigma^2_{\epsilon}). \tag{11}$$

Hence, it is implied that each of the fixed effects are unbiased estimates of supervisor j's effect, as is the case under the identifying assumptions laid out in Section 4. As shown by Morris (1983), one can construct a more efficient estimator of ψ_j using the posterior mean of ψ_j conditional on the estimate $\hat{\psi}_j$:

$$E[\psi_j | \hat{\psi}_j] = \lambda_j \hat{\psi}_j, \tag{12}$$

where $\lambda_j = \frac{\sigma^2_{\psi}}{\sigma^2_{\psi} + \sigma^2_{\epsilon_j}}$. As described in the text, I estimate the shrinkage factor $\hat{\lambda}_j$ by bootstrapping the estimation of equation 1. For each supervisor j, I obtain bootstrap estimates of the fixed effect $\hat{\psi}^k_j$, where k= 1,...,1000. I estimate the error variance of each $\hat{\psi}_j$ using the sample variance of the bootstrap distribution: $\hat{\sigma}^2_{\epsilon_j} = \frac{1}{k-1} \sum\limits_{k=1}^{1000} (\hat{\psi}^k_j - \bar{\hat{\psi}}^k_j)^2$. I then estimate $\hat{\sigma}^2_{\psi}$ using the variance estimator proposed by Morris (1983):

$$\hat{\sigma}^2_{\psi} = \frac{\sum W_j (\hat{\psi}^2_j - \hat{\sigma}^2_j)}{\sum W_j}. \tag{13}$$

For my main estimates, I use weights $W_j = 1$, so that the estimate takes the form:

$$\hat{\sigma}^2_{\psi} = Var(\hat{\psi}_j) - E_j(\hat{\sigma}^2_j). \tag{14}$$

One can also use the weights proposed by Morris (1983): $\frac{1}{\hat{\psi}^2_j + \hat{\sigma}^2_j}$, which requires one to estimate $\hat{\sigma}^2_{\psi}$ iteratively by first plugging in a guess of the across-supervisor variance and calculating as in equation 14 until the values are sufficiently close. Using this weighted estimate provides similar shrinkage factor.

It is also possible to estimate the variance components directly from the regression residuals using a method proposed by Guarino et al. (2015) and implemented in the policing context by Weisburst (2024). I note that using this method produces nearly identical shrunken estimates to the one used in the main text.

E Crime effects

Do sergeant-induced arrests either low-level or serious dimensions improve public safety? In order to answer this question, I leverage the rotation of supervisors between sector-watches in an event study design similar to the one used in Section 5.1, by estimating how the effect of a sector-watch changing supervisors on 911 calls varies by the magnitude of the change in a supervisor's low-level and serious arrest effect:

$$Log(911Calls)_{et} = \alpha_e + \sum_{k \neq 0} [\pi_0^k D_{et}^k + \pi_L^k D_{et}^k (\Delta \hat{\psi}_e^L)] + \pi_S^k D_{et}^k (\Delta \hat{\psi}_e^S)] + x'_{et}\beta + \epsilon_{et}. \tag{15}$$

I control for event fixed effects and, as in Section 5.5, the average low-level serious arrest propensities of officers and effect for the officers working within a unit each month.

I plot the estimates for π_L^k in Figure A.24b and the estimates for π_S^k in Figure A.24a. Once again, there is no evidence that supervisor switches along either dimension are driven by trends in crime within an area. Moreover, I find no evidence of that supervisor variation along either dimension of arrests leads to reductions in crime crime. The point estimates are below 1% in magnitude for each arrest type in all months following the switch and are statistically insignificant. I can rule out crime reductions larger than 2% related to changes along each dimension.