The Effect of Supervisors on Police Behavior

Austin V. Smith

October 16, 2024

Abstract

I estimate the effects of first-line police supervisors — sergeants — on arrest outcomes. Using the exogenously-timed rotation of officers within the Dallas Police Department, I show that sergeants significantly influence the total quantity of officer arrests: moving an officer from a 10th percentile sergeant to one in the 90th percentile would increase monthly arrests by 42% relative to the mean. Sergeant-focused policies are more cost-effective than those targeted at officers: in the majority of cases, replacing one sergeant produces larger arrest changes than an equivalent replacement of a single officer. I find that sergeants affect arrests for serious crimes by incentivizing their officers to respond to more 911 calls and affect arrests for low-level crimes through discretionary drug enforcement, the latter of which results in disproportionate increases in use of force. However, sergeant effects along either dimension are independent. I show that reducing low-level enforcement induced by sergeants would have no negative effects on arrests for serious crimes. I conclude by demonstrating that performance on the exams used to determine promotions are predictive of a sergeant's enforcement preferences. My findings illustrate that supervision is an important source of discretionary law enforcement practices.

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 use of force against minority civilians. The most commonly proposed policy responses are directed toward front-line officers, such as increasing minority recruitment (Ba et al., 2021b) or modifying officer training (Dube et al., 2023). However, significantly less attention has been given to policies that target police management — in particular sergeants, the first-line supervisors of front-line officers.

The lack of policies directed toward sergeants is in contrast with the economic evidence from other industries suggesting that managers have large effects on their employees' productivity (see Roberts and Shaw (2022), for a review). Because they oversee many people at once, manager-specific interventions may be more cost-effective than worker-targeted policies. However the unique nature of policing makes it difficult to generalize these findings to law enforcement. For one, police officers have more discretion and less immediate oversight than workers in previously-studied contexts, such as production lines (Adhvaryu et al., 2023) or sales departments (Benson et al., 2019). This makes the scope of a sergeant's influence unclear. More pressing for the current policy debate is that, even if sergeants matter, we do not know if their supervision distinguishes between socially productive and harmful officer behavior.

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 career. Officers cannot control the timing of switches, which are determined by vacancies and predetermined 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.

I first 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 to one in the 90th percentile results in 1.6 additional arrests per month, representing a 42% increase relative to the mean. In total, sergeants account for 3.4% of the variation in officer arrests, which I estimate using a leave-one-out variance estimator that corrects for limited mobility bias in the second moment of the sergeant effects distribution (Kline et al.,

2020). I then compare the potential scope of sergeant-focused policies relative to those targeted at officers by considering sets of policies that replace sergeants (officers) at a given percentile of the distribution with an average sergeant (officer). I estimate that 77% of such policies would produce larger arrest changes (in absolute value) if targeted at sergeants rather than officers. This happens because sergeants manage an average of 6.33 officers at a time, which effectively multiplies their impacts. An event-study design around switching events reveals that the effects of moving to a new sergeant are immediate and persistent. The enduring nature of sergeant effects suggests that personnel changes within police management have the power to be longer-lasting than training interventions directed at officers, 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 ask whether sergeants can reduce socially unproductive arrests without reducing officer productivity along other dimensions. If sergeant effects operate primarily through increased officer effort, then it is not clear how sergeant policies can reduce *overly aggressive* enforcement, in particular. I answer this question by disaggregating sergeant effects into two separate arrest categories: serious — which include all violent and property crimes — and low-level — which consist of public order and quality-of-life crimes such as drug possession, disorderly conduct, and outstanding warrants. Prior evidence suggests that marginal changes to low-level arrests are highly discretionary and unlikely to affect public safety (Cho et al., 2023).

Strikingly, serious and low-level sergeant effects are only weakly correlated. This is especially true for the top half of the low-level effects distribution: for sergeants with above average low-level effects, the correlation coefficient is -0.02, and I can rule out a correlation larger than 0.13. This implies that sergeants with large low-level effects can reduce low-level arrests without causing similar reductions in serious crime enforcement. Indeed, I estimate that replacing the top 5% of sergeants in the low-level effects distribution with sergeants who are average along *both* dimensions of enforcement would lead to 974 fewer low-level arrests over 5 years, but 13 *more* serious arrests. Thus, even coarse sergeant policies that only target low-level enforcement can reduce costly police activities without significant collateral damages.

In contrast, serious and low-level *officer* effects are strongly positively correlated, suggesting that variation in arrests across officers is largely attributable to productivity differences. Sergeant-focused personnel interventions would thus be more effective at changing how officer effort is distributed between serious and low-level enforcement, rather than officer interventions which would simply change the total quantity of enforcement.

These 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. These effects are entirely

driven by arrests for simple possession and disproportionately impact Black civilians. In contrast, serious sergeant effects are associated with *reductions* in drug arrests. Sergeant low-level effects also increase other types of low-level arrests — such as disorderly conduct — however the magnitudes of these changes are considerably smaller. Consistent with these 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 are consistent with a model of behavior in which sergeants with large low-level effects incentivize their officers to engage in disorder policing strategies that target crimes which are indicative of broader social ills. Such strategies have long-been codified under "broken windows" theories of crime reduction, and are a relatively low-cost way for officers to demonstrate productivity since they do not depend on circumstances outside of an officer's control — such as victim statements and other physical evidence (Zhao et al., 2003). I find no evidence that civilian-reported crimes decrease when a sergeant with large low-level effects 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 to a 15% increase in use of force incidents relative to the mean.

In contrast, sergeants affect serious arrests through 911 response effort. A one standard deviation increase in serious sergeant effects leads officers to answer 2.2 more 911 calls per month, relative to a mean of 61. These increases are driven by officers volunteering for more calls outside of their assigned location when closer officers are unavailable. The marginal 911 call is more severe than the average, however there is no evidence that officers 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 aggressive. Under a sergeant with large low-level effects, the *probability* of making an arrest is elevated across call types. These changes are most prominent for the least severe categories, such as mischief and vandalism.

For serious sergeant effects, a greater focus on call activity translates into increased arrests for crimes that directly harm other people, 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 interactions.

Sergeant effects are strikingly large considering their management limitations. Most police interactions lack immediate supervision and sergeants have less opportunities to monitor officers compared to supervisors in private sector firms. I examine two potential mechanisms to explain sergeant influence on officer behavior. First, I ask whether sergeants "lead by example." Ethnographic evidence suggests that officers show more respect to "street sergeants" who conduct patrol activities in the field (Van Maanen, 1984).

4

From an officer's perspective, these sergeants understand the difficulties and complexities of patrol work. I proxy for leading by example using two measures of *sergeant* field activities: arrests and first-response to 911 calls. Second, I examine direct monitoring of officers using a sergeant's presence at their own officers' calls. Monitoring would allow sergeants to advise officers and impose their preferred policing strategies directly. Anecdotally, some officers frequently appeal to their sergeant's advice, meaning that sergeants who are more available to respond should be more effective at influencing officer behavior (Brooks, 2021).

I find evidence that low-level sergeant effects operate through leading by example and direct officer monitoring. 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. This also results in 0.54 more first-responder calls (mean = 4.26) and a 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. The most likely explanation is that serious sergeant effects 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. Indeed, survey studies show that officers are sensitive to the enforcement preferences of their sergeants (Engel and Worden, 2003), suggesting that these administrative controls are meaningful to officers. While data limitations prevent me from investigating 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 that these sergeants are more likely to approve overtime associated with call-responses.

I conclude by considering whether sergeant effects can be predicted using the information available to departments *before* someone is promoted to the position. These predictions can inform sergeant-selection mechanisms. In particular, I ask whether sergeants' pre-promotion characteristics are predictive of their effects on officer arrests. To do so, I leverage multiple sources of detailed DPD personnel records, including newly obtained data on exams that determine promotion to sergeant. Promotion exams for supervisory roles are ubiquitous across police agencies and are used as an objective measure of management capability. However, there is no consensus regarding their effectiveness as a selection mechanism (Bishopp, 2013).

I do not find evidence of differences in sergeant effects across race, gender, or age at the time of promotion. However, the distribution of sergeant effects differs significantly by exam performance. The low-level effect distribution for high scorers is shifted left relative to that of low scorers (Kolmogorov-Smirnov p-value = 0.034). This suggests that sergeants who are considered the best candidates for promotion are less likely to value low-level enforcement and those who do value low-level arrests tend to be marginal promotees. For serious effects, I find that high scorers have a *wider* distribution compared to low scorers (Kolmogorov-Smirnov p-value = 0.013). These results suggest that the top promotion candidates have more heterogeneous management styles — at least in terms of serious enforcement — than those at the

bottom of the promotion list. One plausible explanation for this heterogeneity is that two different types of sergeants score highly on the exam: those who are especially motivated to be effective managers and those who are good test-takers. These two types of sergeants may differ in their willingness to involve themselves in their officers' activities. While my data does not allow me to test this hypothesis, I discuss findings from previous ethnographic studies that support this interpretation.

Overall, my results show that first-line managers are a major determinant of police behavior and suggest that more research should be devoted to analyzing the effects of policies targeted at this crucial level of police organizations. In particular, the fact that sergeants who induce the most low-level arrests do not increase serious arrests on average indicates that policies affecting these sergeants' management styles could significantly reduce the costs of policing without worsening public safety. Because these sergeants tend to score worse on their promotional exams, one possibly effective policy could be to require additional management training for those who are on the margin of promotion.

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/or use of force are responsive 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 this literature that studies the importance of 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 primarily affect broad tactical strategies such as stops and searches (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.[1] In contrast to these papers, which primarily study the causal pathways of supervisor effects, my findings demonstrate the full magnitude 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

---

[1] 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.

large part by unobserved tastes. Despite the different takeaways from our studies, a shared conclusion is that first-line supervision crucially shapes police outcomes.[2]

Second, I contribute to work within labor economics on the impact of managers (Bertrand and Schoar, 2003; Bloom and 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 [3], 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, in particular (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 doing so, I contribute further to our understanding of how bureaucrat performance can be influenced by the incentives established through the organizational structure of bureaucracy in the absence of other traditional workplace incentives, such as performance pay (e.g. Bertrand et al., 2019).

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.

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 patrol, the largest section of most police departments, 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 1 of 35 sectors within

---

[2]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.

[3]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 these metrics cannot fully encompass the impact of teachers on students' life outcomes, specifically in the case of criminal involvement. Such findings highlight that even in settings where objectives are somewhat well-defined, a worker's impact may be experienced along other important, but not necessarily related, dimensions.

the city. There is at least one sergeant assigned to each sector on each of the three watches, or shifts.[4]

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 whom 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. This has also been well-documented in ethnographic research; one police officer who was interviewed by Van Maanen (1984) 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 whom 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 that 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 which enable them to incentivize officers to adhere to their preferences. Sergeants are responsible for writing recommendations for promotion 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 may provide strong incentives for officers to obey their superiors (King, 2005).

---

[4]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.

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 also respond to models of ideal patrol behavior that sergeants express through "leading by example." That is, if sergeants want their officers to make more arrests for drug crimes, they may choose to go out in the field and make these arrests themselves. Prior research suggests that officers are likely to be receptive to this style of management, since officers may believe that "street sergeants" understand the complexities of patrol work and therefore garner more respect (Van Maanen, 1984). Moreover, survey studies show that when sergeants engage in specific field activities, their officers are more likely to believe these activities will be used to assess their job performance (Engel and Worden, 2003).

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 the officer is reassigned to a different sector within their division or if their current unit receives a new supervisor. Re-assignments occur because of officer or supervisor 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 .[5] Instead, officer vacancies may be filled within a division and watch at the discretion of command staff. When sergeant vacancies occur, other sergeants can interview for the opening, but the final transfer decision is unrelated to seniority.

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 divisions, shifts, and days of the week. If large scale staffing changes are needed, then the Chief can implement a Patrol Bid, which allows a designated set of sergeants and/or officers to choose their division, shift, and day-off groups in descending order according to time in rank. Since the bid may not occur every year,[6] eligibility for the bid is not known until 2 weeks prior, and it does not always occur 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 division commanders and, according my conversations with DPD officers, do not seem to follow a discernible pattern.

To become a sergeant, one must pass a tenure threshold and take a promotional exam that tests their knowledge of department procedures and leadership potential. Exam-takers are then ranked according

---

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

[6]In my sample, it happens 3 out of the 5 years

to their performance and, aside from a few year circumstances, promoted in order of their ranking on the list when openings arise. In the DPD, officers are required to spend at least 1 year as a senior corporal before qualifying for the sergeant's exam.[7] 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 other 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 via FOIA request from the Dallas Police Department and Dallas County District Attorney's Office, covering June 2014 to July 2019. I combine information on police incidents, personnel, officer activity, and court outcomes in order to construct a monthly panel of sergeant assignments for patrol officers that links officer enforcement activity to each of their sergeants throughout the sample. I focus on sergeant assignments for patrol officers whose primary job duties are answering civilian-initiated calls for service, patrolling their assigned beats, and responding to crimes observed "on-view." Patrol sergeants are linked to a sector of the city and a watch.

Dallas only maintains assignment data at the level of patrol divisions, a less granular geographic level, meaning that they do not keep records of sergeant assignments. However, the Computer Aided Dispatch (CAD) system used to allocate officers to police incidents stores the daily sector and watch assignments of responding officers (and sergeants) who are working a call (see Appendix C for details). I use these assignment data, combined with division assignments and promotion histories, to construct monthly sergeant assignments for patrol officers from June 2014 to July 2019. Specifically, I assign officers to the sector-watch in which they are assigned on the most days within the month and assign each sector-watch the sergeant who is observed with that assignment on 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.

In practice, officers may not be assigned the same sergeant each day they work. 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-

---

[7]Senior corporals are one rank above officer, but share all of the same duties as officers. The only difference is that senior corporals qualify for some roles that officers cannot, such as tactical team member and field training officer. For my analysis, I do not distinguish between officers and senior corporals.

assigned sergeant still carries administrative responsibilities for the officer, such as making recommendations or conducting performance evaluations. However, to the extent that officers receive advice and instruction from multiple sergeants within a month, I capture the effects of the sergeant to whom they are exposed most often. For the average sector-watch, I observe the assigned sergeant in CAD on 9.1 unique days within the month, suggesting it is unlikely that I consistently select fill-in sergeants who are more active than the one who is regularly-assigned.

In order to ensure my estimates are consistent with effects driven by an officer's permanent sergeant, I subject the sample to two filters. First, I require that officer-sergeant spells last at least 2 consecutive sample months. This minimizes any assignment errors that would be generated from officers working in a temporary assignment that has relatively 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 erroneous 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 determined. I show in Figure A.1 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 some time 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 individual officer activity from several sources. I use officer identifiers in the universe of arrest reports to count the number of arrests made by each officer in each 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), as well as several non-index crimes that have high social costs: simple assaults, any form of domestic violence, sexual assault, fraud, and DWI.[8] All other arrests are classified

---

[8]The only difference between my classification and Rivera (2022)'s is the inclusion of DWI in serious crimes, though classifying

as low-level. Low-level crimes primarily consist of outstanding warrants,[9] 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 use the most severe charge to classify each arrest. In other words, an arrest with any serious charge is classified as serious.

I link each arrest to court outcomes using records obtained from the Dallas County District Attorney's office and classify its conviction status.[10] 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 in order to separately evaluate civilian-initiated and proactive police encounters.[11] 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, and disciplinary action. Finally, I link each sergeant who was 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 given earlier than 2012, however 58% of sergeants in my sample 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 decisions. 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, because the event study sample requires officers to have successive stable patrol assignments, I am excluding some officers who have strong preferences for making arrests and may be more likely to transfer into a specialized team where they can make a large number of arrests, such as gang or narcotics enforcement. Since all of my analysis will include officer fixed effects, these sample differences should not significantly affect my findings.

---

these crimes as low-level would not meaningfully change my results.

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

[10]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 equalt 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 above. 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.

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

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 4 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 who are 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 average proportion of misdemeanor arrests, which account for 80% of all arrests according to estimates by Natapoff (2016).[12] However, officer arrests exhibit a substantial amount of variation. The standard deviation is 3.64 — nearly the same size as the mean. Figure A.4a shows that this variation is present across sergeants in my data by plotting the distribution of arrests across sergeants. Officers working for a sergeant in the right tail of the distribution make over 6 arrests per month. However, the average number of arrests made by officers working for sergeants cannot identify a sergeant's effect on arrests, since it cannot be disentangled from officer discretion. This discretion translates to even larger variation in arrests across officers (see Figure A.4b). Separating the effects of officers from the effects of sergeants requires observing changes in officer behavior under different sergeants. This is the crux of the empirical strategy, which I lay out in detail in the next section.

4 Empirical Strategy

4.1 Estimating Sergeant Effects

I first estimate sergeant effects on the quantity of officer arrests for 347 sergeants in my sample. To do so, I follow the two-way fixed effects approach pioneered by Abowd et al. (1999) and since 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 takes the following form:

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

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} includes sector-watch fixed effects in order to net out differences in arrests that are generated by spatial and temporal variation in crime. Since sector-watch and sergeant assignments overlap, separate identification of sergeant and sector-watch fixed effects requires all sector-watches in my sample to be managed by multiple sergeants. The data cleaning procedure described in Section 3 ensures this is the case in my analysis sample, and

---

[12]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, whereas possessing personal-use amounts of cocaine is a felony.

in practice each sector-watch has many sergeants — 6.95 on average (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. Additionally, I include a second degree polynomial of officer tenure to adjust for time-varying changes in arrest behavior that may be correlated with an officer's sergeant through their priority in schedule realignments.

Sergeant fixed effects are identified by officers who change sergeants. Specifically, sergeants in this model are only credited for *changes* in the behavior of officers who switch to or away 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, and tenure group. For ψ_j to identify the causal effect of supervisor j, I require that mobility of officers between sergeants is as-good-as random, conditional on officer fixed effects and the controls. In other words, sergeant assignments need to be uncorrelated with determinants of officer behavior that are not present 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 consider three forms of endogenous mobility that would lead to violations of the identifying assumptions.

First, sergeant assignments must be uncorrelated with 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 an officer whose preference for making arrests is increasing over time, then the model would erroneously attribute gains in arrests to the new sergeant. Moreover, if officers were moved systematically to neighborhoods whose demand for police enforcement was increasing, then I would also overstate the variation in sergeants effects. Such scenarios could occur 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-spots policing is a common strategy used by departments in which a large number of resources are focused on a few "hot-spot" areas with a high concentration of crime (Weisburd and Eck, 2004). If high-arrest sergeants are better at identifying crime hot-spots and ask for new officers, then I would observe increased arrests for officers who move 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, θ_{ij}, that are correlated with $\psi_{J(i,t)}$ and missing from the model.

14

Anecdotally, the nature of sergeant assignments in the DPD, as described in Section 2, makes endogenous movements unlikely. While officers and sergeants have some ability to select who they work with through the patrol bid, they cannot when the bid occurs or the timing of other moves that are generated by vacancies. Thus, they have limited ability to adjust their behavior in anticipation of a switch or sort on crime trends within 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 be a driving factor in officer or sergeant moves, as described. Moreover, new policing initiatives are often carried out by teams who are distinct from regular patrol officers. For example, hot-spot initiatives in Dallas introduced a specific team to carry out hot-spot patrol strategies during my 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 provide such event studies by splitting sergeants into terciles according to the average number of arrests made by officers whom 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 the sorting pattern of sergeants would lead estimates of $\hat{\theta}_i$ and $\hat{\beta}$ to be contaminated by sergeant effects. 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 arrests change suddenly and persistently. This is consistent with the fixed effects specification, in which the sergeant's effect "turns on" 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 movements do not appear to be systematically related to the direction of the switch. Third, Figure 1 suggests that officer-sergeant match quality is not an important determinant of moves. Sorting 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 on match quality is asymmetry in the effect 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:

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

whereas the same expectation for an equal move in the opposite direction is given by:

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

Under positive (negative) assortative matching, both of the expected match quality difference 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 higher arrest sergeant increase their arrests on average while those who move to a lower arrest sergeant decrease arrests. Moreover, moves in opposite directions appear to be symmetric: a move, for example, from the 3rd tercile to the 1st appears 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 appear to produce average changes in either direction. I additionally verify the symmetry across moves in Appendix Figure A.2, which plots the average change in residual arrests for upward moves against the average change in residual arrests for downward moves in the opposite direction. The points line up roughly along the -45 degree line.[13]

While the fixed effects from the model are, by assumption, unbiased, consistency requires that the number of observations tends to infinity within each officer-sergeant pair. 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 the amount of estimation error in the fixed effects, I adapt Empirical Bayes shrinkage procedures that are 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.[14] I then multiply each of the raw fixed effects by the Empirical Bayes shrinkage factor, which equals 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.

---

[13]Splitting sergeants into terciles provides relatively few cases of symmetric moves to evaluate. In Appendix Figure A.3, I perform the same event study analysis by splitting sergeants into quartiles. The conclusions of Figure 1 and Appendix Figure A.2 hold similarly in this case.

[14]For the bootstrap, I follow the procedure outlined by Best et al. (2023). I obtain residuals $\hat{v}_{it}$ 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 $\hat{\sigma}_\epsilon$.

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.

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

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

I focus on variation in pair-level average residualized arrests, since 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 outlined in the previous section reduces measurement error in the estimated fixed effects, the variance components may still be biased if there are too few officer movers in the data relative to the number of sergeants - the well-known *limited mobility bias* problem (Andrews et al., 2008). This would over-inflate the variance of sergeant fixed effects - causing us to conclude that sergeants have more of an impact than they actually do - and bias the covariance negatively due to the inverse correlation of measurement errors between the 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 these other contexts, the 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 adopt the bias-correction strategies developed by Andrews et al. (2008) and Kline et al. (2020) for the estimates of the variance components. The Andrews et al. (2008) method relies on a derivation of the bias term that requires homoskedastic errors, whereas Kline et al. (2020) - KSS, herafter - derive the bias term under unrestricted heteroskedasticity. The KSS bias term is a linear combination of each observation's variance weighted by each observation's 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 is the set of officers and sergeants who remain connected when any one officer is removed. The leave-one-out connected set only removes 3 sergeants and 3 officers from my data. In Section 5, I show that each of the bias-correction methods provide similar estimates of the variance and covariance components.

4.3 Disaggregating Sergeant Effects

Officers may allocate different levels of effort to enforcing low-level and serious crimes. Officers can make more low-level arrests by looking for evidence of low-level crimes. For example, officers can stop more civilians on the street or search for contraband at a higher rate during civilian 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 clear ex-ante whether a larger sergeant effect results from officers being induced to make more serious or low-level arrests, as the sergeant effect is a composite of serious and low-level effects:

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

The correlation between ψ_j^L and ψ_j^S reveals the complementarity between low-level and serious enforcement induced by sergeants. If sergeants who induce more arrests do so through uniform increases in officer effort, then we would expect ψ_j^L and ψ_j^S to be positively correlated. On the other hand, a negative correlation suggests that sergeant behaviors tend to favor one type of arrest of the other. A lack of correlation between suggests that sergeants induce serious and low-level arrests through independent behaviors.

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. As expected, shrinking the effects reduces variation and increases the mass around 0. However, even after shrinkage, a one standard deviation increase in sergeant effects increases arrests by 0.66 per month, or 17% relative to the mean. Though relatively symmetric around the average sergeant, the fixed effects distribution has a heavy left tail, implying the existence of a large amount of low-enforcement sergeants. 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).

One way to contextualize these magnitudes is to calculate the impact of replacing a sergeant with someone new.[15] Replacing all sergeants above the 90th percentile with average sergeants would result in 2,278 fewer arrests over the course of my sample, a 2.16% reduction. Notably, large changes can be achieved by replacing only a small number of sergeants — only 35 in the example given. While there is greater variation in officer effects compared to sergeant effects (see Figure A.5), the fact that sergeants manage many officers means that replacing a sergeant may produce larger changes to arrests. For example, a one-month replacement of a 90th percentile sergeant who manages the average number of officers with an average sergeant would lead to 2.63 fewer arrests made *ceteris paribus*.[16] On the other hand, an equivalent replacement of a 90th percentile officer would only lead to 2.25 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 an average 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 77% of cases, replacing one sergeant produces changes which are at least as large (in absolute value) as replacing an equivalent officer.[17] The largest difference between sergeant and officer replacement are in the left tail of the distribution, where there is a concentration of sergeants who induce far fewer arrests than the average. While replacing sergeants from the 60th to 90th percentiles also produces larger changes than an equivalent officer replacement, the gap becomes smaller past the 90th percentile, which is 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, over a majority of the distribution, replacing one sergeant produces changes that are at least as large as replacing an equivalent officer.

I present the results of the variance decomposition in Table 2, including the decompositions using raw fixed effects, the Bayes-shrunken fixed effects, and both variance bias-correction methods. As expected, the raw fixed effects overstate the contribution of sergeants to variation in officer arrests. All three of the bias-adjustment methods produce the variance in sergeant and officer effects. In the preferred KSS specification

---

[15]To avoid double-counting arrests that are credited to multiple officers, I calculate these impacts after re-estimating the sergeant and officer effects using a modified measure of arrests that only credits officers for half an arrest when another officer is listed on the arrest report. However, the results do not differ substantially when using the original arrest measure.

[16]The calculation is performed by taking the difference between the 90th percentile of the sergeant effects distribution and its mean — 0 — and multiplying that by the 6.33, the average number of officers managed by a sergeant in a month.

[17]Instead of assuming that each sergeant manages the average number of officers, one can also perform these calculations using the observed number of officers managed by each sergeant. Doing so leads to the conclusion that 67% of sergeant replacements produce larger changes than an equivalent change of officers. Calculations are available from the author upon request.

(Columns (7) and (8)), I find that variation in sergeants can explain 3.39% of the total variation in officer arrests. Officers, on the other hand, account for nearly three-fourths of the variation in arrests. Given that officers ultimately make the judgment call to arrest in any civilian interaction, the larger contribution of officers relative to sergeants is not surprising. However, as I 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 supervisor and officer effects. I find evidence across all specifications of high arrest officers sorting to low arrest sergeants. 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 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% to 3.4%). However, officers in my context explain a significantly larger portion of the variation than do workers in a factory line (5.4% to 72.3%). In Lazear et al. (2015)'s study of managers within a technology-based services firm, the authors estimate that a standard deviation change in manager increases productivity 2.6 times more than a standard deviation change in workers, assuming that the manager has an average-sized team. The same calculations in my context imply that changing a sergeant by 1 standard deviation produces an effect which is 1.44 times larger than a standard deviation change in officer.

A key caveat to comparing my estimates to the manager effects literature is that the managers studied by Adhvaryu et al. (2020) and Lazear et al. (2015) oversee workers who perform relatively routinized tasks with little discretion and clear objectives. Since front-line workers within policing have significantly more discretion, it is not surprising that police management has a smaller effect on worker behavior. The fact that managerial changes still outperform worker changes in policing is noteworthy. One of the central questions asked by this paper is whether front-line managers are capable of changing police behavior given the limitations of their position. The results presented in this section suggests that they indeed can.

5.2 Diagnostic Checks

This section conducts diagnostic checks in order to alleviate concerns regarding the validity of the fixed effects estimates from equation 1. To begin, I provide evidence in support of the identifying assumption made in Section 4. Section 4 discusses three forms of endogenous officer mobility that would 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. One might be concerned that officers are assigned to sergeants on the basis of recent changes in arrests. For example, an officer may start making fewer arrests 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 that I attribute to the new sergeant would come from the declining trend prior to 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:

$$y_{et} = \alpha_e + \sum_{k \neq 0} [\pi_0^k D_{et}^k + \pi_1^k D_{et}^k (\Delta \hat{\psi}_e)] + x_{et}' \beta + \epsilon_{et}. \tag{6}$$

Here, e indexes a switching event - uniquely determined by the officer i and the switch month T - and k indexes months relative to the switching month. I include the model controls (tenure, sector-watch fixed effects, and day-off group fixed effects) in order 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 that depends on the size of the change in sergeant effect. I test for endogenous reassignment by evaluating the pre-move event study coefficients. The event study model also nests a test for general misspecification of the sergeant effects, since 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 an officer 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 duration of the panel, in line with the insights from the nonparametric event study in Figure 1.[18]

The previous test cannot capture whether officers sort to sergeants based on trends in crime rates. 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.

---

[18]The size of the effect after moving is also close to 1, which is reassuring since 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.

This would result in officers making more arrests after a move, but 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). I report these estimates in Table B.4, separately for officers moving into and out of each sergeant's unit.[19] 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.[20]

A second source of endogenous mobility is unobserved shocks that are 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 interrogate the presence of correlated contemporaneous shocks using a placebo test on "incumbent" officers. Incumbents are officers who already work with a switching officer's new sergeant when the switcher joins the team. 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. 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 $l \neq i$ who are managed by sergeant j' 5 months before the switch and 4 months after:

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

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 between the effects of sergeants j' and j increases by 1. This model also provides a secondary test of endogenous crime trends, since we would expect arrests to be increasing for incumbent officers prior to a positive switch in sergeant effects if larger sergeant effects are driven by growing demand for making 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. These results indicate that sergeant effects are unlikely to be contaminated by contemporaneous changes in enforcement policy.

The third identification concern relates to match-specific error components. If officers sort to sergeants with whom they have a comparative advantage for making arrests, then the model will be misspecified

---

[19]I use 911 calls as a measure of crime rather than crime reports since crime reporting will be endogenous to police activity (Weisburd, 2021). To the extent that aggressive policing can also 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.

[20]The previous 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.

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 that sergeant and officer effects are additively separable. If sergeant effects were generally 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 of 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 model to systematically under- or over-estimate arrests for some 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 (negative) average residuals for top (bottom) quintile officers matched with bottom quintile supervisors. Appendix Figure A.9 demonstrates that the mean residuals do not exhibit any clear pattern 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 my 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 a sergeant.

5.3 Disaggregating Sergeant Effects by Crime Type

In this section, I disaggregate the sergeant effects in order to evaluate the potential for sergeants to be able to reduce costly forms of police enforcement without changing socially beneficial police efforts. Figure 4 depicts a binned scatterplot with low-level sergeant effects on the horizontal axis and serious sergeant effects on the vertical axis, along with a linear fit and nonparametric 95% confidence band (Cattaneo et al., 2024). The linear fit implies a positive relationship between low-level and serious effect and the estimated correlation is a small but statistically significant 0.11. However the confidence band does not allow me to rule out independence of the two effects. Moreover, visual inspection of the plot suggests that sergeants who induce far fewer low-level arrests than the average tend to induce fewer serious arrests; however, throughout the rest of the distribution the relationship is flat. Indeed, if I remove just the bottom 5% of supervisors in the low-level distribution, the estimated correlation falls to 0.07 and is statistically insignificant. More importantly, 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. The evidence suggests that the strategies that sergeants use to increase low-level arrests are independent of strategies that increase serious arrests. For sergeants at the top of the low-level effects distribution in particular, it is unlikely that reducing low-level arrests would reduce the overall enforcement of serious crimes.

To make this point clear, I calculate the change in low-level and serious arrests that would be generated by replacing the top 5% of low-level effect sergeants with sergeants who are average in both serious and low-level dimensions. The hypothetical policy here is necessarily coarse: I target only low-level sergeant effects but, as a consequence, also impact the distribution of serious effects. I estimate that this change in the sergeant effects distributions would lead to 974 fewer low-level arrests over the 5-year period, but they would *increase* serious arrests by 13. Because sergeants a roughly even proportion of sergeants with large low-level effects have positive and negative serious effects (see Figure A.11), replacing all of these sergeants would significantly reduce low-level enforcement without affecting arrests for serious crimes.

The independence of low-level and serious sergeant effects is particularly, since these effects for officers are strongly and positively correlated (see Figure A.12). The complementarity of low-level and serious arrests for officers suggests that variation in arrests across officers can be attributed to productivity, since officers who make more arrests tend to do so for both serious and low-level crimes. In contrast, sergeants affect different dimensions of arrests through distinct actions. These findings suggest that sergeants are a more effective target for policies to reduce the over-enforcement of low-level crimes, since a policy intervention that targets officers who make many low-level arrests would likely have the negative side effect of reducing their overall effort.

5.4 Channels of Officer Behavior

What behaviors do sergeants encourage in their officers in order to induce more low-level or serious arrests? I first investigate this question by estimating how serious and low-level sergeant effects change the specific crimes for which officers make arrests. To do so, I leverage regressions of the following form:

$$y_{it}^c = \alpha_L^c \hat{\psi}_{J(i,t)}^{L*} + \alpha_S^c \hat{\psi}_{J(i,t)}^{S*} + \theta_i^c + + x_{it}' \beta^c + \nu_{it}^c, \tag{8}$$

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

In Table 3, I report estimates for the 3 most frequent serious and low-level charges. Increasing a sergeant's serious effect results in statistically significant increases across the three largest 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 average total) but increase by 37% when officers are assigned to a sergeant who has a one standard deviation larger serious effect. Additionally, sergeants with larger low-level effects actually *reduce* domestic violence arrests by 0.016 (3.5% relative to the mean) per month. Sergeants indeed make crime-specific tradeoffs, even though the two dimensions of sergeant effects are uncorrelated in the 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 is for drug arrests. A one standard deviation increase in the low-level effect of a supervisor increases an officer's drug arrests by 0.17 per month, over 50% on the mean. However, the same size increase for a sergeant's serious effect reduces drug arrests by .036 per month (10% on the mean). On the other hand, both dimensions of sergeant effects are positively associated with warrant and disorderly conduct arrests, though the magnitudes are larger for the low-level effect.[21]

Patrol officers can make arrests through two channels: self-initiated interactions — such as traffic stops, investigating abandon buildings, or stopping citizens on the street — or 911 calls. I assess how sergeant effects impact arrests through each channel by estimating equation 8 using officer-initiated arrests and call-initiated arrests as the outcome variables. The results (Table 4) show that sergeant-induced serious

---

[21]These findings have important implications for racial disparities. Black civilians make up a disproportionate share of drug arrests (55% relative to 50%), meaning that aggressive low-level enforcement may widen existing disparities. Indeed, I find Black arrests increase at a faster rate than Hispanic or white arrests for positive changes in low-level sergeant effects, even when compared to the higher baseline arrest rates for Black civilians (Table B.7). However, changes in serious sergeant effects increase arrests across all races in proportion with their baseline arrest rates, which is sensible given the lack of clear compositional changes for serious sergeant effects.

arrests are entirely initiated from calls (column 2) while low-level arrests originate from a mixture of calls and officer-initiated interactions. However, for low-level sergeant effects, the officer-initiated interactions account for over 60% of the total increase in arrests, a stark result since officer-initiated arrests are less prevalent than arrests from 911 calls. These results suggest that low-level sergeant effects are associated with disorder styles of policing, where sergeants ask their officers to heavily enforce low-level crimes in order to reduce crime more broadly. However, such arrests may be particularly costly to society on-net, since these behaviors have not imposed a large enough cost 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 change would change as a result of more 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 not available, which means that some portion of sergeant effects may operate purely through greater levels of call activity. I clarify the mechanisms for changes in call-initiated arrests by regressing monthly call-specific outcomes on the low-level and serious sergeant effects, for which I present results in Table 5. I find that the number of calls answered per month increases with positive changes to both the serious and low-level sergeant effects (column 1). A one standard deviation increase in the serious sergeant effect is causes subordinates to answer 2.2 more calls per month relative 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 to calls answered, low-level effects lead to more substantial changes in the probability that officers make an arrest at the calls they respond to (column 2). A 1SD 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 only induce low-level arrests at calls and higher serious effects are associated primarily with increases in serious call arrests (columns 3 and 4). I also find evidence of a small increase in low-level call arrests for sergeants with a larger serious effect. These findings reinforce the earlier observation that management behaviors that increase low-level arrests in isolation do not meaningfully impact enforcement for serious crimes.

Figure A.13 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 effects on arrest probability are largest

for the least severe call type, Mischief. However, the positive overall effect on call arrest percentage for sergeants with high serious effects are driven entirely by officers responding to more serious calls. There is no evidence that increasing the serious sergeant effects increases arrest probability after conditioning on call type. In other words, serious sergeant effects operate entirely through call-response effort, while low-level sergeant effects operate through aggressive discretionary enforcement, even at civilian-initiated 911 calls.

The results thus far suggest that sergeant incentivize their officers to change the quantity of serious and low-level arrests. However, do sergeants value arrest *quality*? 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 0 arrests in a month, which means I am unable to study conviction rates directly. Instead, I follow the approach used by Gudgeon et al. (2023) and estimate the impact of sergeant serious (low-level) effects on the number of convicted arrests and total arrests separately. I use these estimates to calculate how changing the serious (low-level) sergeant effect changes the ratio of convicted arrests to total arrests and compare it to the ratio of the 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 conviction rates overall. This suggests that sergeants who induce arrests of either kind do not do so through encouraging officers to make low-quality arrests that will eventually be thrown out in court. However, the results for serious and low-level conviction rates indicate that the sergeant-induced arrests are not necessarily higher quality, either. 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 changes 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 that is 4.75 times higher than the average low-level arrest, so 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 these arrests decreases low-level conviction rates for high serious effect sergeants.

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, 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 .02 more uses of force per month, a 14% increase relative to the mean, compared to a change of 0.006 from an equivalent

increase in serious effects. In column 2, I find that low-level sergeant effects increase complaints and serious sergeant effects decrease complaints, however both coefficients are imprecisely estimated. Both serious and low-level sergeant effects result in more officer activity and more formal interactions with civilians, which likely contributes to increased use of force. However, the stark difference in the effect sizes suggests that targeted low-level enforcement leads to violent escalation that is likely disproportionate with the costs of the crimes that it sought to address.

Overall, the results in this section reveal that, even though sergeants have a substantial ability to shape 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 sergeants can change officer behavior along one dimension without significantly altering behavior along the other. Indeed, inducing low-level arrests and serious arrests implies very different downstream officer behaviors. Officers make more serious arrests through call activity. On the other hand, to induce more low-level arrests, sergeants incentivize their officers to proactively detect low-level crimes on patrol, particularly those involving drug possession.

5.5   Mechanisms: Leading By Example and Monitoring Officers

I now consider *how* sergeants are able to change the actions of their officers. Since police not only have discretion but operate largely outside the direct view of their sergeants, there has been a long-standing debate over whether supervision *can* change officer behavior, much less if 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 — via modeling the activities you want to incentivize — my 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 that sergeants attend as the first-responder. Arrests made by sergeants may be observed by officers, either directly or second-hand through stories told by another officer or the sergeant themselves. 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 responding first to a call may model activity as desirable to their officers, and their officers can observe this call-response through their in-car computer terminals, which show who is assigned to calls within their division.

Sergeants can also impose their preferences through increased monitoring of their officers. They can do so by assigning themselves to their officers' calls more often. In addition to overcoming the inherent monitoring limitations of supervision in policing, these situations may also increase the likelihood that officers reach out to their sergeants in for advice in future uncertain situations, since they know 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 using 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 (x_u) fixed effects in order to control for variation in sergeant behaviors that are driven by the time and location of their assignment. 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 substantially 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 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 SD increase in low-level effects (column 5).

On the other hand, I do find no evidence that serious sergeant effects are associated with leading by example or enhanced officer monitoring in the field. The point estimates in each column are small and imprecisely estimated. It is likely that sergeants induce serious arrests through other forms of behavior that cannot be captured in 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 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). While I do not have access to overtime data, I can use officer shift data to measure the number of calls and arrests that take place outside of their regular hours. I show in Table B.6 that close to half of the total increase in calls answered as a result of a 1SD change in serious sergeant effects are driven by calls outside of an officer's regular hours. Moreover, overtime *arrests* increase, but predominantly for serious crimes. I also find that low-level effects are associated with significant increases in calls and arrests outside an officer's shift. While these measures are imperfect proxies for overtime, they suggest that sergeants may be willing to use their administrative control of overtime approval to shape the behaviors of their officers.

6 Predicting Sergeant Effects

My results demonstrate that sergeants can be a powerful target for police reform policy. To what extent are sergeant effects mediated by observable characteristics that are determined *before* someone has been promoted to sergeant? Answering this question is of interest for two reasons. First, it has important implications for structuring sergeant-focused policies. If sergeant effects can be predicted by performance at the rank of officer, then departments could use performance to inform promotion decisions or target officers for special management training before they take over sergeant duties full-time. 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 for 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., 2021b,a). The extent to which this variation carries over into managerial preferences provides insight into this question.

I evaluate the differences in sergeant effect distributions across four observable pre-promotion characteristics: race, gender, age at the time of the promotional exam, and score on the promotional 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 split sergeants into two categories of age at the time of exam and promotional score. I call sergeants "older" if they were above the average across all exams and I call them "high-scorers" if they were above the average score for their particular exam.

In Figure 6a, I present densities separately by exam score. I find a striking difference between high and low scorers in the distributions of both serious and low-level effects. For low-level effects, the distribution of high scorers is shifted to the left to he low score distribution, and a Kolmogorov-Smirnov suggests that the this difference is statistically significant (p-value = 0.034). Thus, on average, those who score below average on the promotional exams induce *more* low-level arrests than those who score above the 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. To the extent that marginal promotees — who are barely promoted by virtue of their low exam score — are more likely to value low-level arrests, my results indicate that the knowledge required to perform well on promotional exams may be inversely correlated with effects on aggressive policing tactics.

In Figure 6b, I also find evidence of a statistically significant difference in the distribution of serious effects between high and low scorers (Kolmogorv-Smirnov p-value = 0.013). In contrast to the low-level effects, it does not appear that the differences are driven by a monotonic shift in one direction. Instead, high-scorers have a wider distribution over serious effects than low-scorers. The distribution for low-scorers

is 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 scores on the exam may indicate someone is particularly motivated to be a successful sergeant, or it may simply mean someone is a good test-taker. It's possible that the motivated candidates are more willing to involve themselves in their officers' patrol activities, whereas the good test-takers choose to avoid intervening on their officers. While I lack the data necessary to test this story, there is suggestive support of this hypothesis from the ethnographic literature. Van Maanen (1984) suggests that officers who are administratively-inclined and relatively bookish are likely to perform well in a standardized exam setting, but are often not be willing to be actively involved in patrol decisions as a sergeant. On the other hand, it is not likely that those with the most natural test-taking talents are the *only* successful exam-takers, and motivated officers could achieve high scores through extra studying.

In the appendix, I present the empirical density of low-level and serious effects, separately by race, gender, and age groups. I do not find evidence of significant differences in the distribution of either supervisor 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. I do, however, find evidence that low-level *officer* effects differ by race in a way that is 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. It is possible that the sergeant selection mechanism is designed in a way that filters out officers within different racial categories that contribute to this variation. This question warrants deeper analysis that exceeds the scope of this paper.

7 Conclusion

This paper shows that supervision matters for police enforcement decisions. Critically, supervisors' preferences are heterogeneous and operate through distinct forms of officer behavior. My findings have several important implications for police reform policy.

First, training or personnel realignments that target first-line supervisors may be cost-effective interventions to change the way that police use discretion. Other police reforms, such as training programs, have shown to be effective but can be degrade in impact over time (Owens et al., 2018). My findings that supervisors provide persistent incentives for their officers to police in a particular way suggests that policies that can change the preferences of supervisors could be effective in a long-lasting way. Moreover, because supervisors represent a smaller portion of police agencies, even policies that would require supervisors to be regularly re-trained would be less costly than effective programs that would require the entire stock of frontline officers to be re-trained at regular intervals.

Second, policing activities that aggressively target low-level crimes are often disconnected from activi-

ties targeted at abating more serious crimes. That supervisory effects along these dimensions appear to be largely independent supports a burgeoning body of literature that reducing violent and property crimes does not necessarily require harsh enforcement for low-level crimes, which may be more connected to public health and civilians' overall quality of life (Cho et al., 2023). Moreover, I provide evidence that actions that specifically target low-level crimes produce more collateral damage via use of force than actions that increase arrests for serious crimes.

Third, this paper highlights the importance of future research regarding promotion mechanisms in the public sector broadly and policing in particular. By showing that the preferences of managers vary significantly by their performance in the standardized promotion process, my findings indicate that even "objective" promotion tools can produce unexpected tradeoffs for public organizations when selecting staff to occupy positions with supervisory powers.

Ultimately, my paper indicates that policy interventions that target first-line police management would be 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., Burgess, R., Chawla, A., and Xu, G. (2019). The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies*.

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.

Bishopp, S. A. (2013). An evaluation of the promotional processes in a large texas metropolitan police department. *Policing*, 36:51–69.

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 Reenen, J. V. (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.

Brooks, R. (2021). *Tangled Up in Blue: Policing the American City*. Penguin Press.

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$_p$*process*$_a$*nd*$_t$*imeline* $- iap - 07 - 03$.

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.
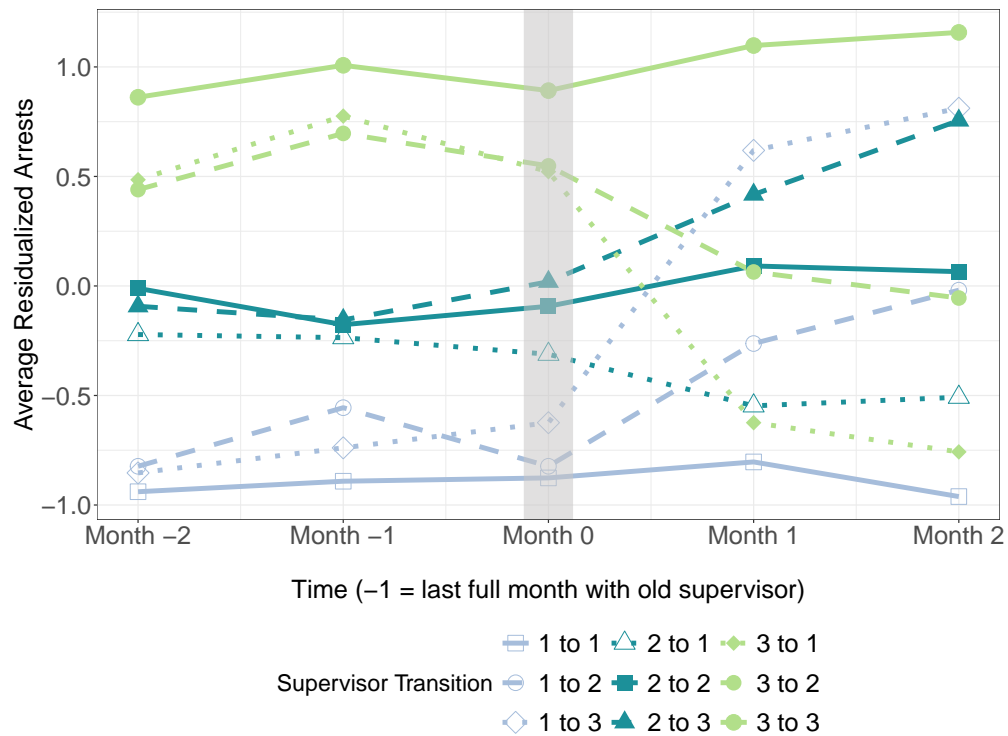
8 Tables and Figures

Figure 1: Event Study Around Moves

Notes: This figure plots the average number of arrests in the months around a supervisory change by the magnitude of the change. In particular, I group supervisors 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 supervisor are described by "Supervisor Transition." Arrests are residualized by a second-degree polynomial of officer tenure and officer, sector-watch, and day-off group fixed effects.
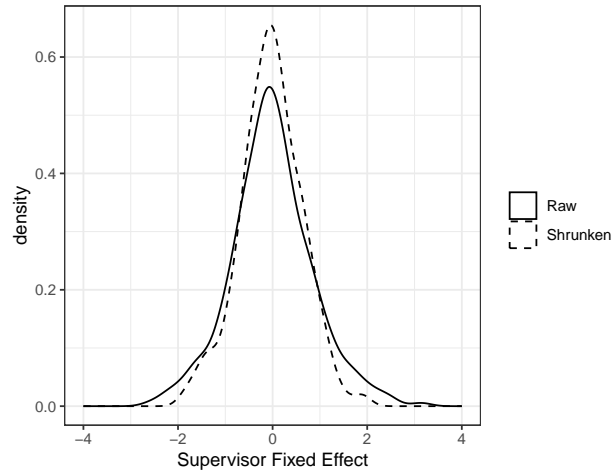
Figure 2: PDF of Supervisor Fixed Effects

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

Figure 3: Event Study Coefficients
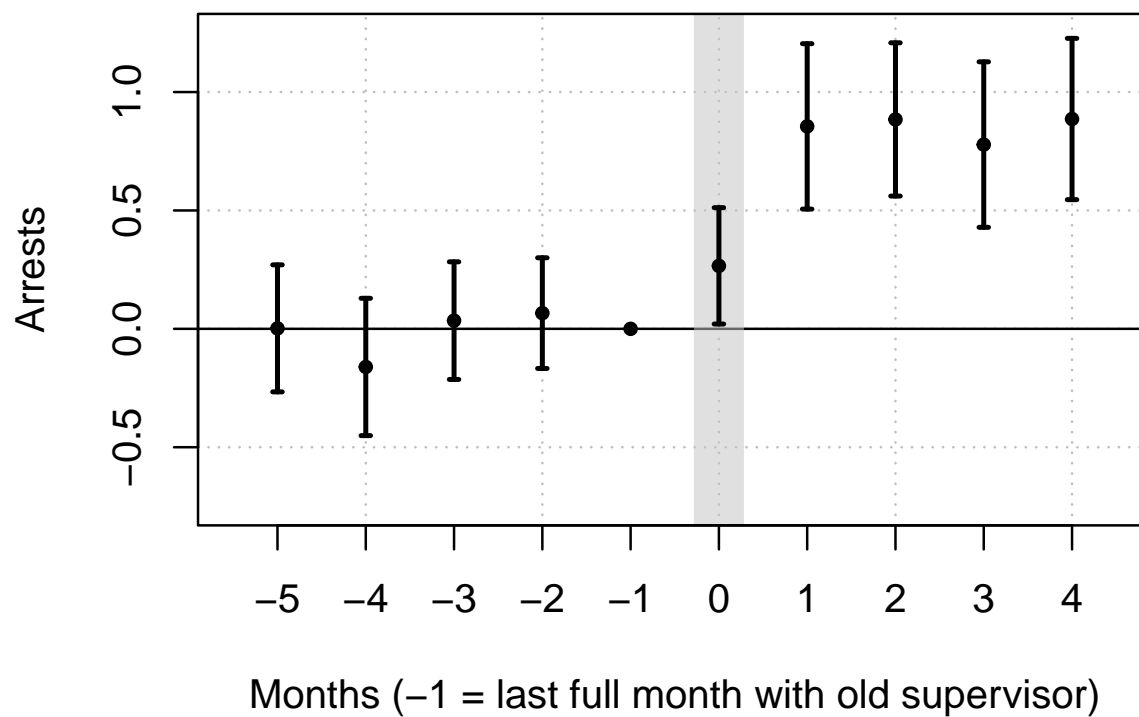
Notes: This presents the event study coefficients for equation 6. Standard errors are clustered at the officer level. Month -1, the last full month an officer spends with their old supervisor, is used as the reference month. The model is estimated using the event study data that are balanced on [-5, 4].
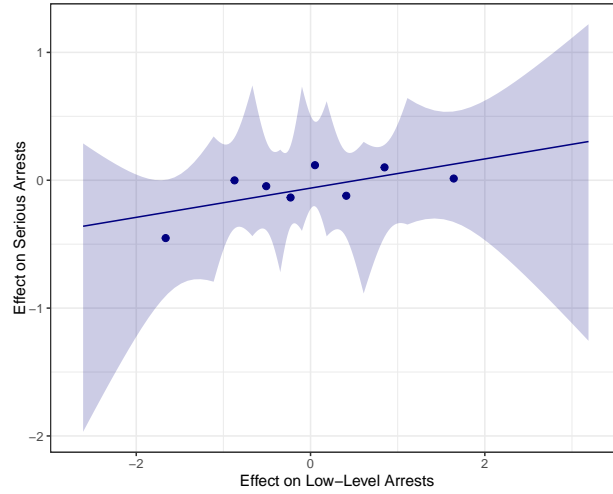
Figure 4: Comparing Low-level and serious arrest effects

Notes: This figure displays a binned scatterplot of the relationship between the standardized low-level supervisor effects and standardized serious supervisor 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.
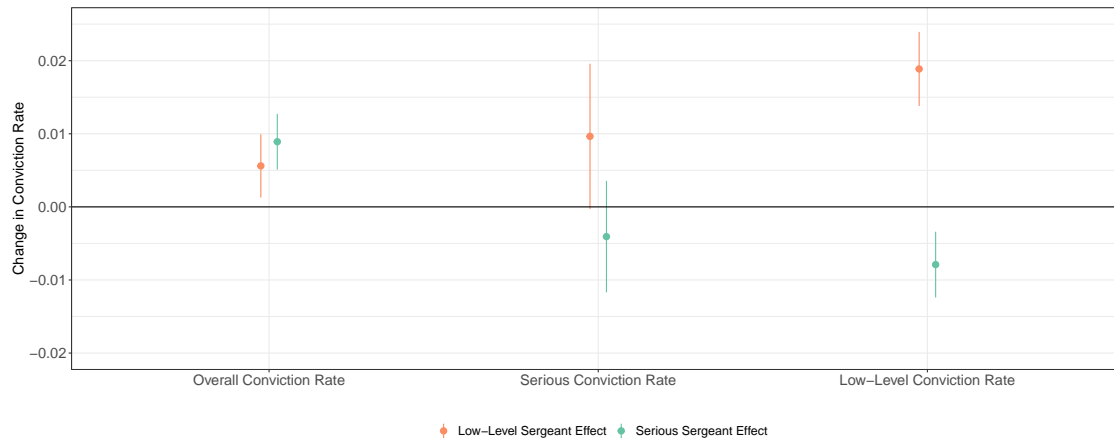
Figure 5: Effect on Conviction Rates

Figure 6: Sergeant Effects Distributions by Promotional Exam Performance

(a) Low-level effect distribution

(b) Serious effect distribution
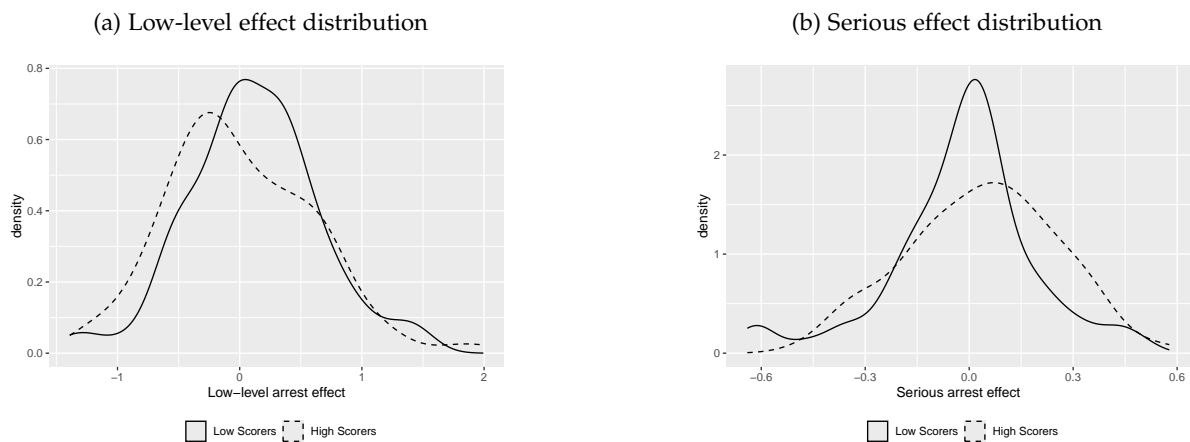
Notes: These figures plot distributions of shrunken sergeant fixed effects separately by the sergeant's performance on the promotional exam they took to attain their current position. I define high scorers as sergeants who score above the average for their exam and low scorers as those who score below the average.

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 switching events in which the focal officer is observed with the pre-switch sergeant at least 6 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 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 effects of changes in the two dimensions of supervisor preferences on arrests for 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) arrest effect is given by the standardized Bayes-shrunken supervisor fixed 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 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 supervisors	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 supervisor 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 supervisors and officers decrease. This implementation adapts Julia code that is provided publicly by the authors and developed by Paul Courcera.

Table 4: Source of Civilian Interaction

	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 effects of changes in the two dimensions of supervisor preferences on arrests from different sources of civilian interaction. The low-level (serious) arrest effect is given by the standardized Bayes-shrunken supervisor fixed 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 effects of changes in the two dimensions of supervisor preferences on officer activity relating to 911 call responses. Calls answered refer to calls in which the officer is among the first units dispatched to the scene. Arrest probability is measured according to arrests in which the focal officer is present on the arrest report. The low-level (serious) arrest effect is given by the standardized Bayes-shrunken supervisor fixed 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 effects of changes in the two dimensions of supervisor preferences on the number of use of force incidents and complaints involving the supervisor's officer. The low-level (serious) arrest effect is given by the standardized Bayes-shrunken supervisor fixed 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 supervisor behavior on the estimated low-level and serious supervisor 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 supervisor responds to in which their subordinates are also present. Standard errors are clustered at the supervisor level. ***$p < 0.01$,** $p < 0.05$,* $p < 0.1$.

Appendices

A Figures

Figure A.1: Robustness to Alternative Sampling Decisions

(a) Unrestricted

Correlation = 0.8720
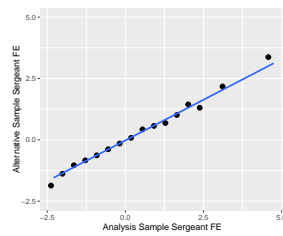
(b) Impute Missing Within Spell
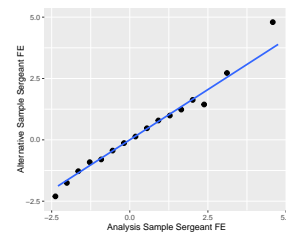
Correlation = 0.8666
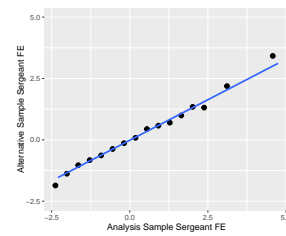
(c) Impute Missing Within Spell, remove everything else
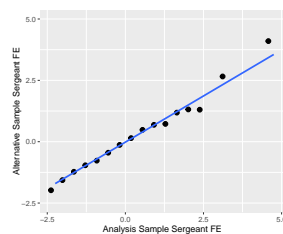
Correlation = 0.9873
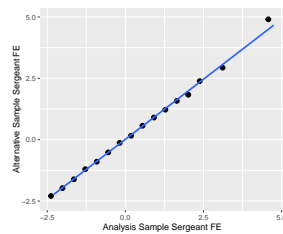
(d) Impute all temporary assignments
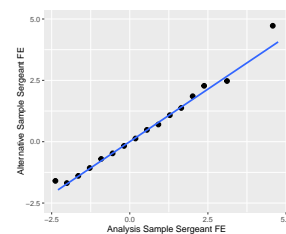
Correlation = 0.9362

(e) Keep temporary, remove missing

Correlation = 0.9153


(f) Keep missing, remove temporary

Correlation = 0.9311

Notes: This figure presents the correlation between supervisor 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.2: Symmetry in moves

Notes: Each crosshair represents a pair of symmetric moves between supervisors in different terciles. 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.

Figure A.3: Nonparametric event study using sergeant quartiles

(a) Event Study Figure

(b) Symmetry Test
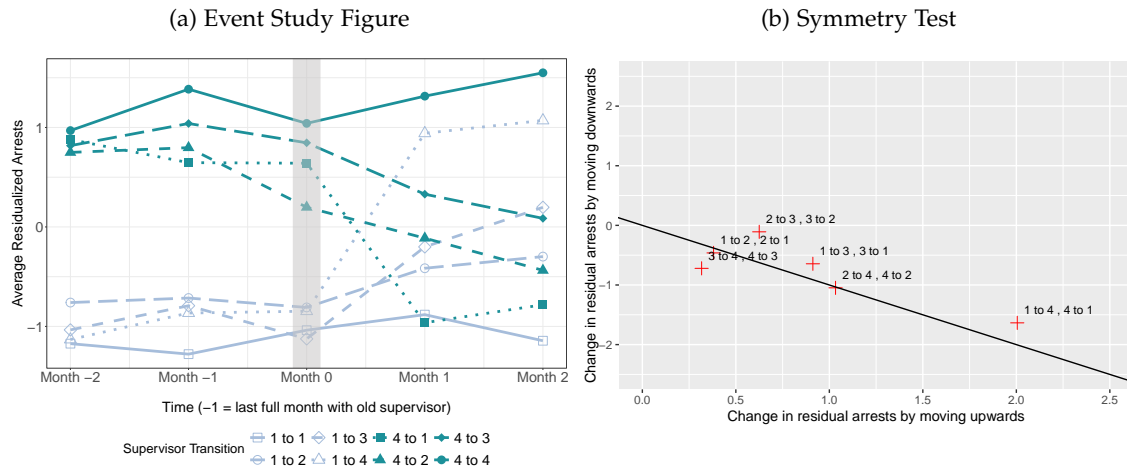
Notes: These figures present the same information as Figure 1 and A.2, instead splitting supervisors into quartiles rather than terciles. To limit the amount of lines on the figure, I only plot transitions from supervisors in the highest and lowest quartiles in (a).

Figure A.4: Distribution of Arrests

(a) Across Sergeants

(b) Across Officers
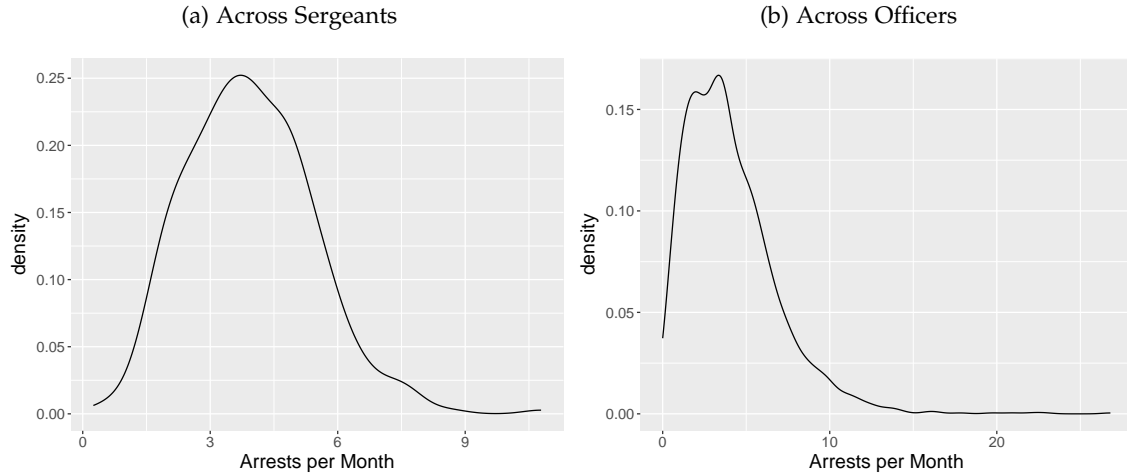
Notes: These figures present empirical distributions for monthly arrests. For a given sergeant, I calculate the average number of monthly arrests made by officers who work for them. A.4a then plots the distribution of this average. Then, for each officer, I calculate the average number of arrests they make in a month across all months they are in the sample. I plot the distribution of this average in A.4b.
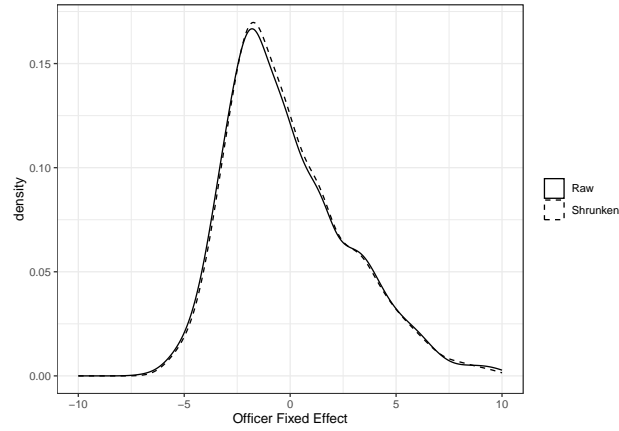
Figure A.5: Distribution of Officer Fixed Effects

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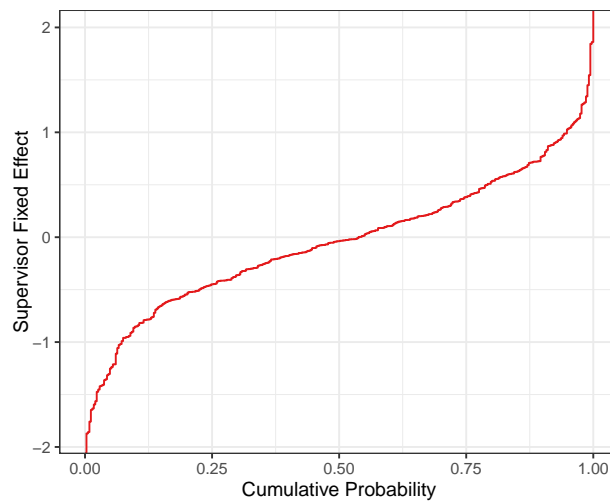

Figure A.6: CDF of Supervisor Fixed Effects

Notes: This figure displays the CDF of the supervisor fixed effects estimated using equation 1 and multipled by the Bayesian shrinkage factor described in Section 4.
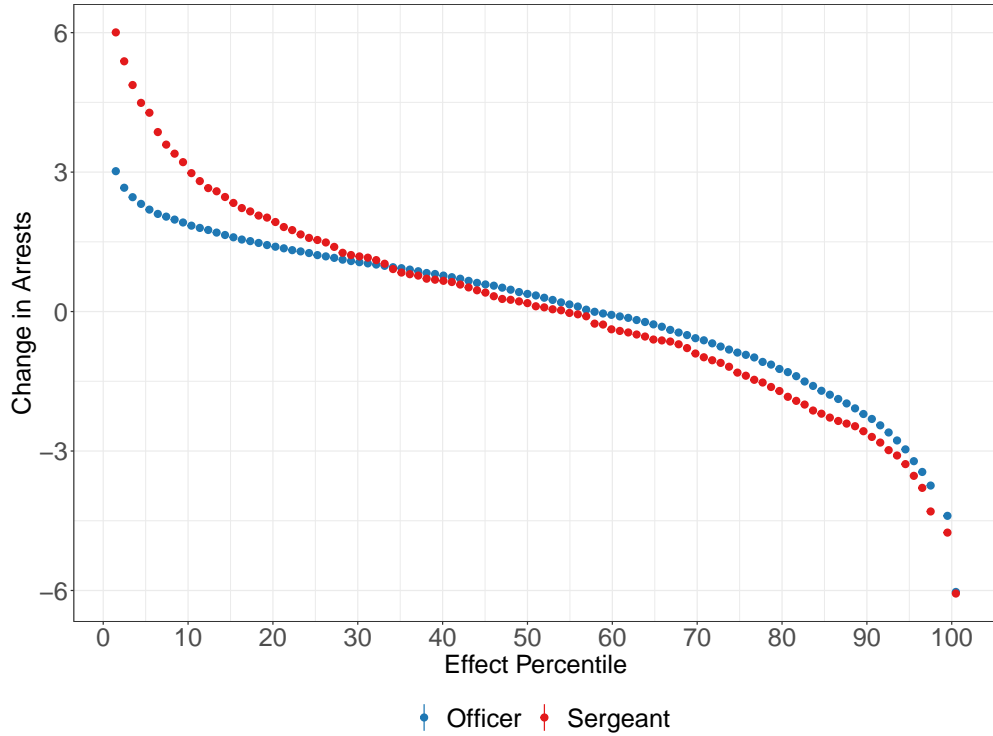
Figure A.7: Effect of a one-month replacement with the average

Notes: This figure plots the calculated effect of one-month replacements of sergeants (in red) and officers (in blue) with an average 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 an average sergeant is calculated by subtracting each sergeant's effect from 0, the average of the sergeant's distribution by construction, 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.
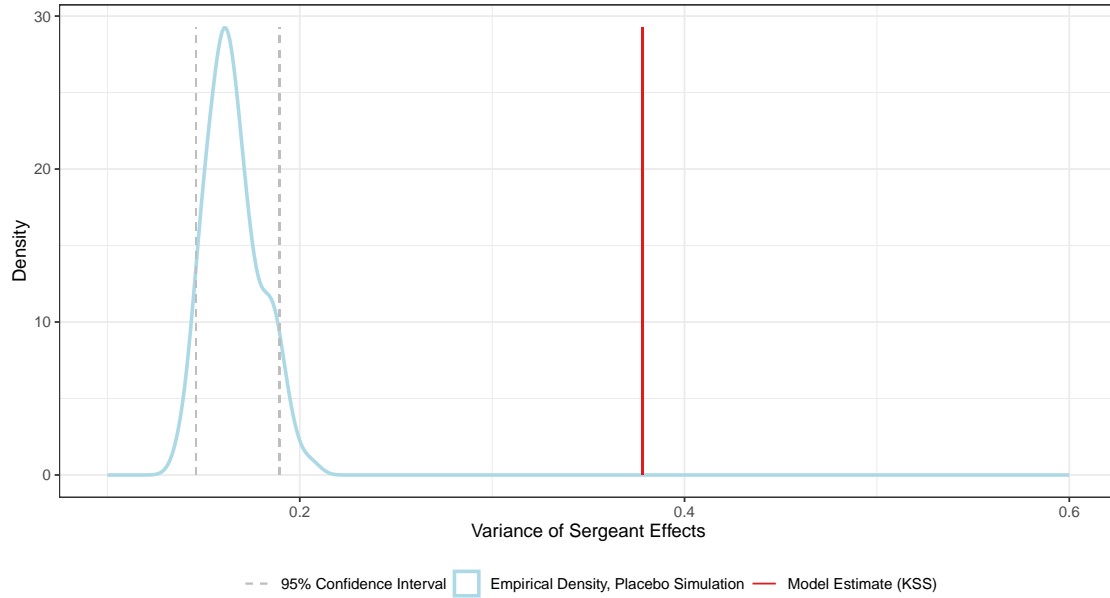
Figure A.8: Placebo Test, Sergeants randomly reassigned to officers

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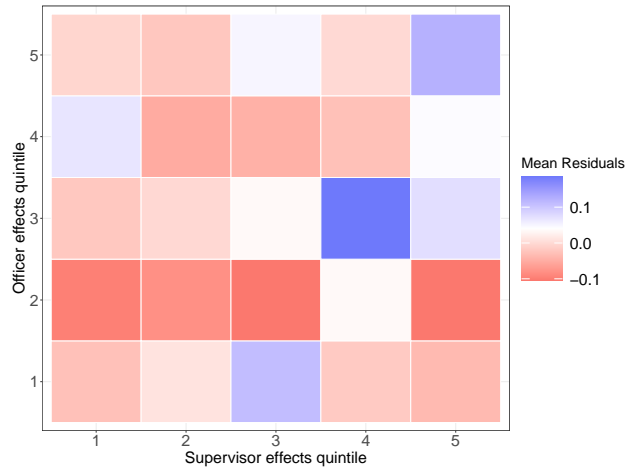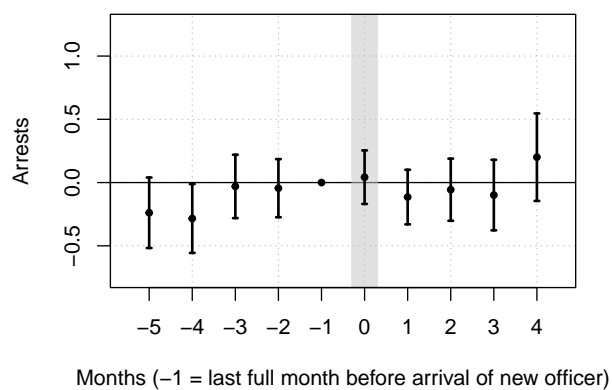
Figure A.9: Residuals by quintile of officer and supervisor arrest effects

Notes: This figure reports the average residuals by quintiles of officer and 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: Distribution of Sergeants between Serious and Low-level effects

	1	2	3
3	11.2%	11.2%	11.0%
2	7.8%	12.7%	12.7%
1	14.4%	9.2%	9.8%

Tercile of Serious Effect

Tercile of Low-level Effect

Notes: This figure displays the percentage of sergeants within each tercile of low-level and serious arrest effects.
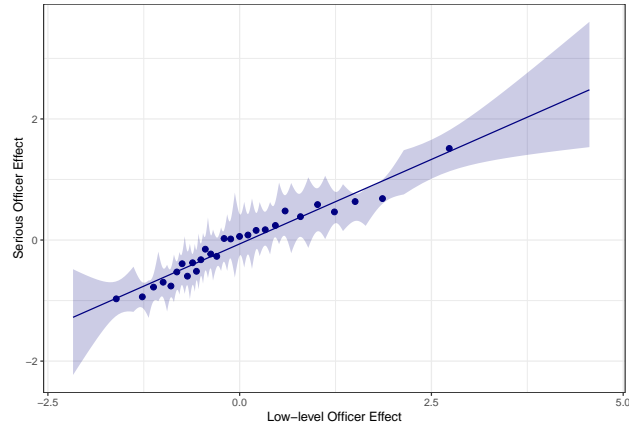
Figure A.12: Low-level and serious officer effects

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.
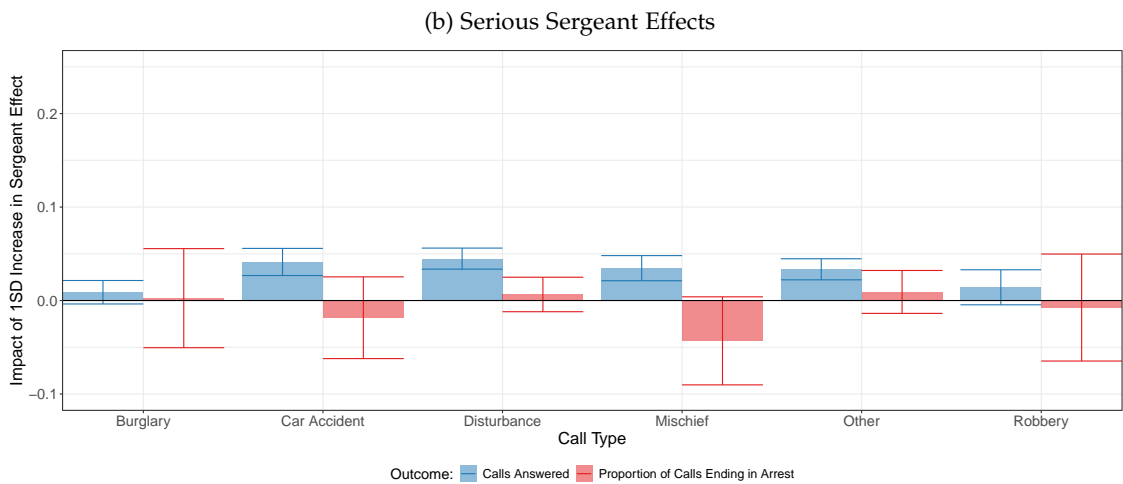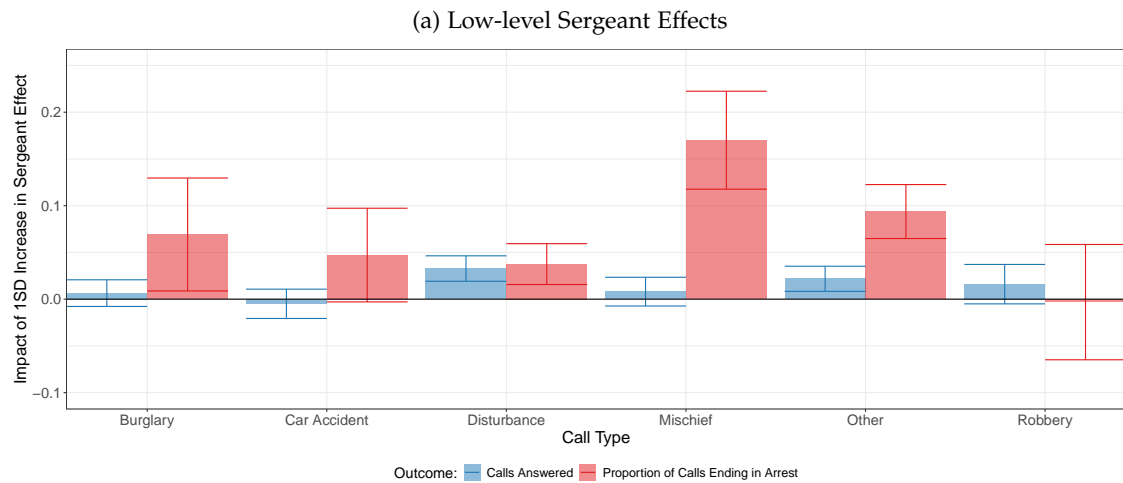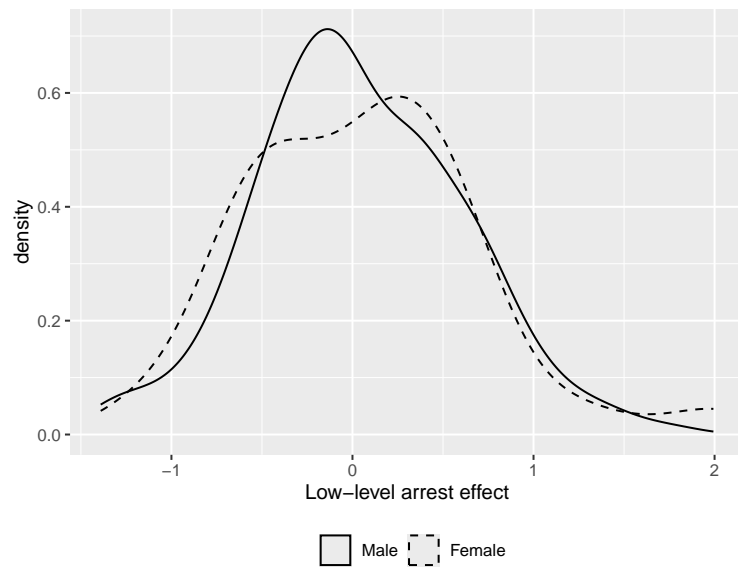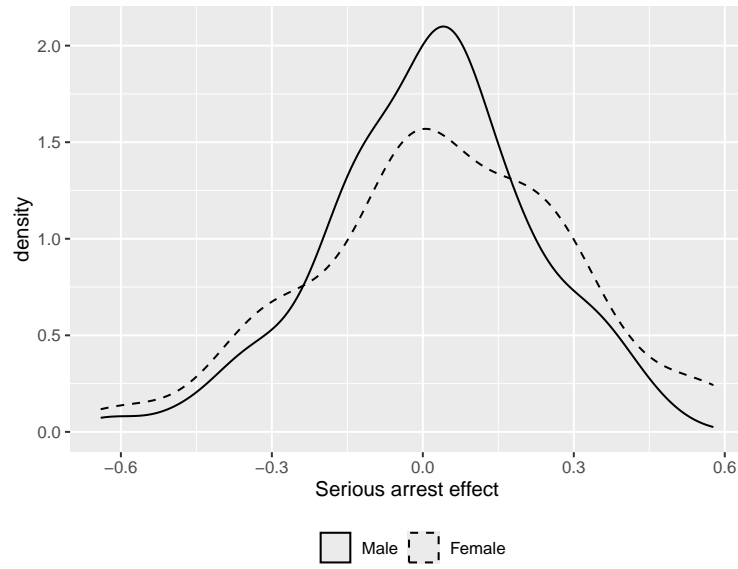
(a) Low-level Sergeant Effects

(b) Serious Sergeant Effects

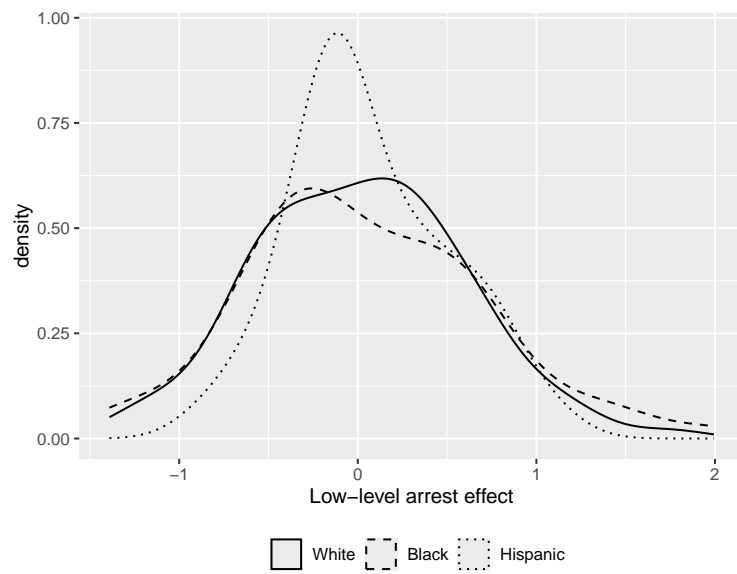Figure A.13: Impact of sergeant effects on calls and arrests by call type
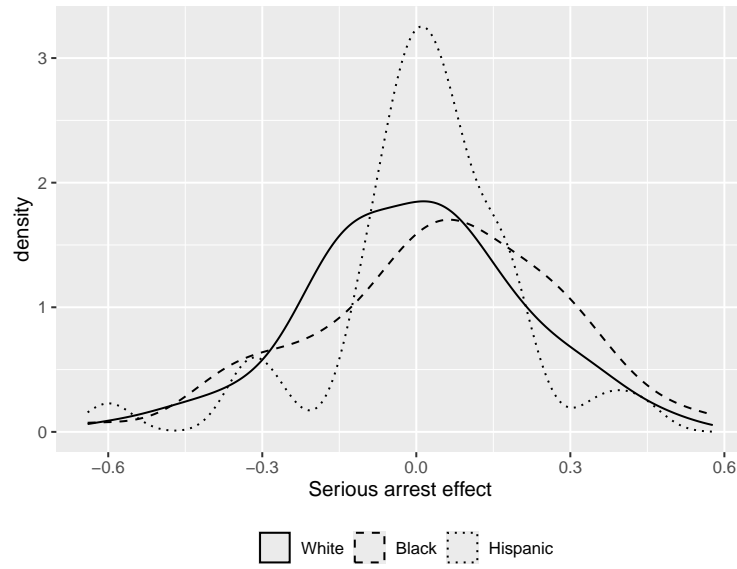
Figure A.14: Event Study by Arrest Type
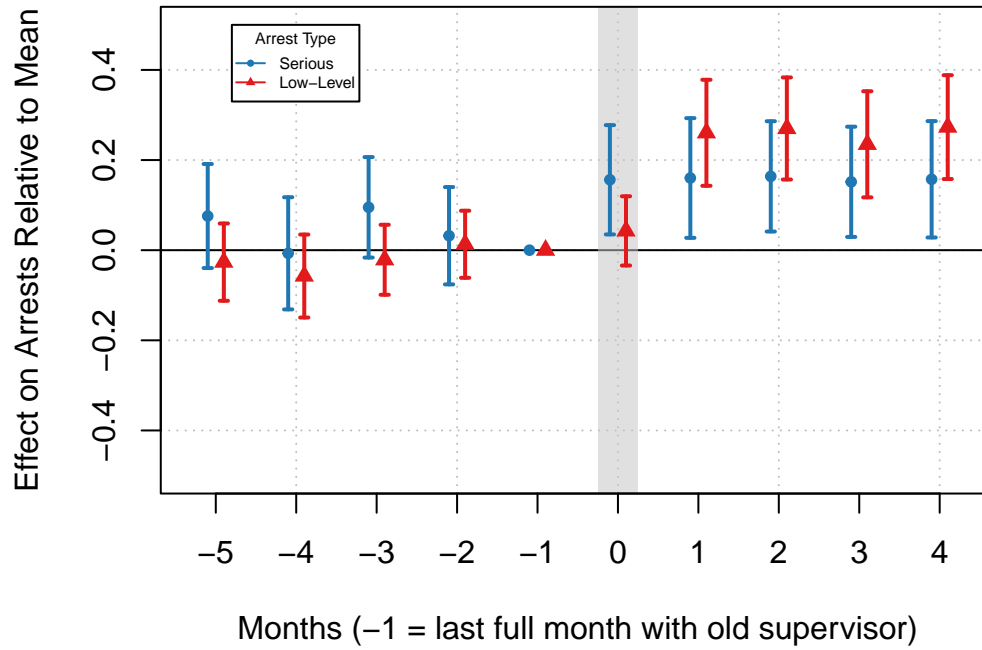
Notes: This figure plots event study coefficients for equation 6, separately for models that use serious arrests (defined as index arrests as well as domestic violence, fraud, simple assault, and DUI) and low-level arrests as the dependent variable. Serious arrest results are given in blue and low-level 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.15: Supervisor effects on crime

(a) $\hat{\pi}_S^k$

(b) $\hat{\pi}_L^k$
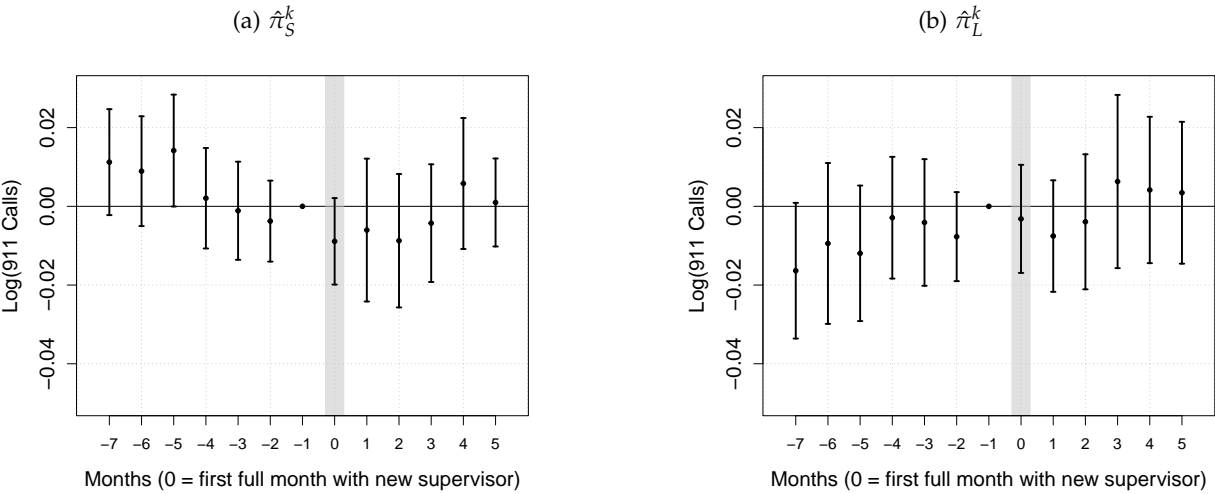
Notes: The figures report for the event study coefficients estimated in equation 15.

B Tables

Table B.1: Data Sources

Data Source	Variables
(1)	(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

Table B.2: Event-study around sergeant switches

k	Total Arrests (1)	Serious Arrests (2)	Low-Level Arrests (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. 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.

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	✓

Table B.6: 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)
Observations	49,923	49,923	49,923
Y mean	6.0084	0.11327	0.04926

Table B.7: 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

Table B.8: 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	
Mean ratio	0.2046	0.4282	0.1322

Notes: ***$p < 0.01$,** $p < 0.05$,* $p < 0.1$.

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 $\psi_j \sim \mathcal{N}(0, \sigma_\psi^2)$, $\epsilon_j \sim \mathcal{N}(0, \sigma_{\epsilon_j}^2)$, 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

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

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 $\lambda_j = \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_{\epsilon_j}^2}$. 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 $\hat{\psi}_j^k$, where k= 1,...,1000. I estimate the error variance of each $\hat{\psi}_j$ using the sample variance of the bootstrap distribution: $\hat{\sigma}_{\epsilon_j}^2 = \frac{1}{k-1} \sum\limits_{k=1}^{1000} (\hat{\psi}_j^k - \bar{\hat{\psi}}_j^k)^2$. I then estimate $\hat{\sigma}_\psi^2$ using the variance estimator proposed by Morris (1983):

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

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

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

One can also use the weights proposed by Morris (1983): $\frac{1}{\hat{\psi}_j^2 + \hat{\sigma}_j^2}$, which requires one to estimate $\hat{\sigma}_\psi^2$ 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 supervisor-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.15b and the estimates for π_S^k in Figure A.15a. 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.