Tenure Effects in Investigatory Stops

Luke Brinkman and Andrew Jordan

October 1, 2024

PRELIMINARY. DO NOT DISTRIBUTE.

Abstract

We use data from Chicago to study how search and hit rates in pedestrian stops change with officer tenure. Search and hit rates are positively correlated, which is inconsistent with simple screening models. We propose a model that explains this pattern by allowing officers to make dynamic investments in search skill. The data are consistent with officers making minimal investments.

This research was supported by a grant from the Weidenbaum Center on the Economy, Government, and Public Policy. We have no other financial interests or conflicts of interest to disclose.

1 Introduction

Investigatory stops are a controversial aspect of policing. Proponents argue that police must vigorously exercise their right to stop and search drivers and pedestrians in order to control contraband and, under the "broken windows" policing philosophy, maintain general order in the community. Opponents argue that these stops have a disproportionate impact on certain demographics and unnecessarily create tense situations between police and citizens that can erupt into violence. It is therefore crucial to understand how officers make the decision to search and the consequences of that decision-making process.

Prior research on investigatory stops in economics has focused primarily on the threat of taste-based discrimination in the selection of pedestrians or drivers for stops. In these screening models, officers must trade off between the volume of stops they make and the success rate of those stops. Officers who make fewer stops leave behind the stops with the lowest probability of success, thus increasing their contraband hit rate. Joint information about stop and hit rates can therefore be used, along with a model of the decision, to discover whether officers set the same search threshold for different groups.

In the current paper, we instead focus on the efficiency of stop decisions made by police. Using data from Chicago, we identify a crucial characteristic of police officers, their tenure on the force, that affects stop and hit rates in a way that cannot be explained by a simple screening model. As officers spend more time on the force, they make fewer stops. However, this is not a case of older officers becoming more circumspect in their stop decisions. The stops made by high-tenure officers also have lower contraband hit rates than their low-tenure peers.

Our first goal is to establish the robustness of this pattern. We employ shift-level and stop-level data from the Chicago Police Department (CPD) to measure both searches per shift and hits per search. Our data contain detailed information about the location and timing of each shift worked by a CPD patrol officer. We use this information to construct an extensive series of fixed effects that control for the month, day-of-week, shift, and beat of each patrol (Ba et al. (2021)). We also use information about the joint assignment of officers to assemble the "cars" that they jointly patrolled in. We use these car assignments to investigate the possibility of joint decisions about whether to conduct a search.

Across all specifications, we persistently find that both search and hit rates fall with officer tenure. The search rate effects are precisely estimated owing to the large number of shifts, over 5 million, in our dataset and the large size of the effect. Relative to officers in their first year after probation, officers in their 20th year on the force search 9.3 percentage points less often. This is a decline of over 100% of the mean search rate across our sample of shifts.

Our measured hit rate effects are noisier because they can only be estimated within the sample of searches, which is substantially smaller. This is exacerbated by the lower search rates specifically among high-tenure officers. We therefore rely on coarser bins of tenure to maintain power. Regardless, we are still able to detect important tenure

effects. Relative to officers in their first 5 years on the force, officers in their 15th-20th years have hit rates 4.7 percentage points lower. That is a decline of 26% of the mean hit rate across our sample of searches.

Our second goal in this paper is to develop a theory of search decisions that can accommodate the patterns established in our analysis. Changes in immediate incentives alone cannot explain search and hit rates that decline together. Officers would never accept a change of that kind unless there was some other constraint or trade-off at play. We theorize that this added dimension relates to the information available to officers as they make their search decision and their skill in interpreting it. In particular, higher tenure officers have a reduced capacity to distinguish worthwhile stops from those that will not produce contraband.

This "search skill" is a crucial determinant not just of how efficient an officer is when selecting stops but also how many stops they make. We develop a simple decision-making model that incorporates this notion of skill. We use this model to show that hit rates are typically increasing in search skill, but search rates can either increase or decrease. When an officer primarily uses his greater skill to rule out unproductive searches, skill and search rates are negatively correlated. If instead the officer uses his greater skill to rule in productive searches, skill and search rates are positively correlated.

We also consider an extension to our model that allows officers to make endogenous investments in search skill. We find that officers may plausibly choose to make no investment in search skill. In turn, officers with minimal search skill become two-dimensional, searching frequently when the return to a random search is high but not at all when that return is low. This is a potential explanation for why police search activity tends to cluster only in specific neighborhoods. In some neighborhoods, investigatory stops are exceptionally rare, but in others, they are common and seemingly indiscriminate. Our model suggests that the distinguishing factor is the return to a random search, and that when police have low search skill, small changes in this return can produce large changes in search behavior.

We then further extend the model to consider dynamic investments in search skill. This requires officers to consider not just how useful search skill will be in the present moment but also to anticipate its usefulness later in their careers. In CPD, district assignments depend only on an officer's tenure, so officers anticipate getting choice assignments in low-crime districts later in their careers. In the dynamic model, this kind of career path disincentives investments in search skill. In a static setting, search skill investments can be beneficial to officers working in high-crime neighborhoods, but in a dynamic setting, these officers instead forego investment in anticipation of an easier future assignment.

Our model produces two useful insights into search skill accumulation. First, some situations where officers may choose to make no investment in search skill can be avoided by furnishing them with a strong baseline of skill. The zero investment equilibrium arises because officers find it too expensive to jump to an equilibrium where search skill has a high marginal return. Narrowing that gap encourages more officers to make the jump. Second, more stable assignment of officers between high- and low-crime neighborhoods avoids the dynamic trap described above. Officers who anticipate working in a high-crime neighborhood for their entire career will make more investments in the skills

needed to survive in that environment.

This paper contributes to a long literature in economics that studies screening decisions in police stop and search activity (Becker (1957), Becker (1993), Knowles et al. (2001), Anwar and Fang (2006), Marx (2021), Martin and Marx (2022)). These studies are primarily interested in the question of whether officers employ race-specific decision rules. The definitive study on this question, Pierson (2020), used traffic stop data from across the US to find small but significant racial biases in search decisions. However, they also found substantial cross-department heterogeneity in both stop and hit rates that was not readily explained by racial bias. The current paper focuses on pedestrian stops rather than traffic stops, which have some procedural differences. Gelman et al. (2007) apply the KPT model to pedestrian stops in New York City. They do not find evidence of racial bias in stop decisions but do note a pronounced disparate impact on Black pedestrians due to highly uneven stop rates across police precincts.

A newer literature considers the role of skill in screening decisions across a range of applications. Chan et al. (2022) studies variation across radiologists diagnosing pneumonia that can only be explained by a skill dimension. Several recent papers studying criminal justice allow for a skill dimension in their analysis of a screening decision (Marx (2021), Martin and Marx (2022), Arnold et al. (2022), Jordan (2024), Angelova et al. (2023)). Unlike those papers, this study documents an apparent difference in skill not across multiple decision makers but across the careers of single decision makers. This motivates our emphasis on investment decisions and their dynamics.

There has been limited criminological research into how to train police to make better stop and search decisions. We could locate only one study that focused specifically on a program to train officers to make better search decisions (Miller et al. (2020)). In that study, a randomly assigned, 1-day training on racial bias in stop and search produced no meaningful changes in search behavior. A police academy curriculum acquired from a FOIA request to the Illinois Law Enforcement Training and Standards Board did not contain a specific unit on investigatory stops of pedestrians.

Feigenberg and Miller (2021) is the previous study most related to ours. They find a lack of correlation between search and hit rates across officers making traffic stops. Officers in their data who searched frequently enjoyed approximately the same hit rates, conditional on searching, as officers who searched infrequently. This is consistent with the idea that officers have only the roughest sense of which searches might yield contraband. This paper builds upon that insight by documenting a different setting where police search outcomes are not adequately explained by a simple screening model and further considering the reasons for and consequences of low search skill.

This paper proceeds as follows. Section 2 describes the CPD stop and assignment data we use in our analysis. Section 3 presents our estimated tenure effects on search and hit rates. Section 4 presents a model that attempts to rationalize these findings. Section 5 concludes.

2 Data

The following section describes the sources of our data and the most relevant details of our sample construction. A more detailed account is included as Appendix C.

Our sample is constructed from Investigatory Stop Report (ISR) data and officer attendance records. ISRs are reports that officers are required to fill out whenever, acting on a reasonable suspicion, they briefly stop a person and make inquiries to confirm or dispel their suspicions. Due to a consent decree issued as a result of a Department of Justice investigation of CPD, ISR data are publicly available online. Each report describes an investigatory stop in detail, including whether any searches were performed and whether those searches uncovered any contraband. The ISRs also include information about protective pat downs (PPDs) performed by officers, which are less thorough than a search and do not require probable cause. In our analysis, we will consider both formal searches and PPDs to be searches. Each ISR also records up to two officers who participated in the stop.

We combine ISR data with attendance data obtained via FOIA request. To match officers across these datasets, we rely on a combination of full name and appointment date. These fields are recorded consistently across both datasets and plausibly assign a unique ID to each officer.

The attendance data describe the time, location, duties, and other details associated with police officers' shifts. We use the attendance data to create a dataset organized at the shift level. This has two advantages relative to using ISR data alone. First, it provides crucial information about shifts where officers made no stops. The frequency of no-stop shifts relative to shifts where at least one stop was made is by far the most important determinant of total stop rates. Second, our shift-level data provide an independent picture of which officers were involved in each stop. This is sometimes more than the two officers listed in the ISR because more than two officers were in the same police car or more than one car was involved in the stop.

Depending on our outcome of interest, we will use either shift-level data or stop-level data. Where our outcome of interest is the number of searches per shift, we use a dataset where the level of observation is a shift worked by a single officer. Where our outcome of interest is the probability that a search turns up contraband, the level of observation is an officer-search. A search that involved multiple officers will appear multiple times. In some robustness checks, we will instead use data organized at the "car" level rather than the officer level. A typical squad car contains two officers along with, potentially, one or more trainees. Shifts are determined to share a car if they have the same start time (including date), watch, car number, and beat number. Within a car, we label the officer with higher tenure the first officer (FO) and the less-experienced officer the second officer (SO). An officer in a car alone is considered both a first and second officer.

2.1 Sample Selection

Our primary analysis uses observations from 2016, the first year covered by the ISR data, to 2020, the last year covered by the attendance data. After joining the shift data to the stop data, we observe 11,785,362 shifts. Keeping only shifts worked by CPD's district units, the units who provide typical beat cops for each of Chicago's 22 districts, drops this number to 6,800,309. Next, we drop shifts worked by officers with titles other than "Police Officer," such as "Sergeant of Police" and "Explosives Technician," bringing the number of observations to 5,960,507. CPD's scheduling operates on a watch system, with three main watches that always start during the same time of day combined with a fourth watch with more variable start times to create sufficient overlap to avoid low coverage during transitions between the primary three. Dropping shifts with abnormal watch numbers, associated with special duties, brings the sample to 5,828,043.

Finally, shifts associated with abnormal beat numbers are dropped. A beat is the physical location in which an officer patrols. Each beat covers a relatively small area of the city with 285 beats in total. A typical beat number is a four digit code with the first two representing the district and the latter two the specific beat within it. We drop both non-numerical beats and those with preceding digits that indicate atypical duties (00 and 81 are both associated with training and police academy). This restriction reduces the sample to 5,616,164 shifts.

We make an additional sample restrictions when constructing our car-level datasets, dropping shifts associated with car identifiers that indicate an assignment to something other than a typical squad car, includes both those that are indicative of officers not being in a position to make any stops at all (e.g. "ADMIN") as well as some that simply indicate slightly less typical duties (e.g. "BIKE"). This reduces the number of individual shifts under consideration by about 24%. This restriction occurs earlier in the sample construction process than the analogs to those described for the officer-level data above; in the officer-level case, the restriction on beat numbers generally removes shifts with car numbers such as "ADMIN," but not such as "BIKE," which are still regularly associated with ISRs despite their somewhat atypical nature. We also drop cars containing more than 2 senior officers or 3 trainees. This reduces the sample by an additional 3%.

The selection of the search-level sample is comprised of the same set of restrictions as described above, but the level of observation is records in the ISR that note that a search took place. As noted, this selection process as well as the construction of cars is covered in much more detail in Appendix C.

2.2 Key Variable Construction

As briefly noted above, our main dependent variables are searches per shift and the probability that a search discovers contraband, frequently referred to as a "hit rate." The number of searches in a shift is an immediate product of our joining of the ISR and attendance data. We define a hit within the ISR data as any search which recovers illegal drugs or weapons of any kind.

Additionally, our analysis centers on how officers' behaviors change over the course of their career. This requires the construction of a measure of officer tenure. An officer with CPD typically spends 6 months in police academy, then a year as a trainee, during which they are always assigned alongside a more experienced officer, before becoming a full officer. The goal is to start counting from their first day as a trainee (meaning a tenure value of less than 1 indicates the officer is a trainee).

When possible, we establish this start date using the attendance data, which extend back to the beginning of 2008. For officers with start dates in 2008 or later, tenure is calculated beginning on the first day we observe a shift worked that is not associated with training in the academy. For officers who joined the force before 2008, tenure is calculated beginning on the day 6 months after their listed appointment date. We prefer to use the attendance data because officers do not always begin their training period exactly 6 months after appointment (e.g. officers who started police academy in late 2019 waited longer than 6 months to join the force due to low staffing demands during lockdown).

2.3 Summary Statistics

Tables 1 and 2 present summary statistics for our 4 main samples: Table 1 contains the shift-level data while Table 2 contains the search-level data, each with separate columns for officer- and car-level observations. Table 1 makes clear that stops and searches are quite rare events. Additionally, with a mean number of officers per car of 1.41, solo officers are fairly common. Table 2 includes the demographic information of those stopped, who are overwhelmingly male and black. Additionally, the differences in the tenures reported across the two tables drives home the point that younger officers are much more likely to make stops and searches than older officers.

3 Tenure Effects

3.1 Observational Patterns

We begin with by documenting the simple tenure patterns in observational data. Figure 1 presents effects by officer tenure. Year 0 in this figure represents an officer's trainee year, during which they are assigned shifts alongside more senior officers in order to learn on the job. All regressions include all years of tenure, but we stop presenting estimates after 25 as cell sizes become small.

Panel A of Figure 1 focuses on search rates. Searches rise somewhat after an officer's trainee year, but they peak in year 3 around at a rate just under 15 searches per 100 shifts. After this, search rates fall gradually for each additional year of tenure. By an officer's 20th year on the force, he is performing barely 1 search per 100 shifts. Put simply, officers with advanced tenures conduct very few searches.

Next, we examine the question of how hit rates change over an officer's career. Panel B of Figure 1 now restricts

attention to ISRs where officers conducted a search and calculates the hit rate of those searches, again by officer tenure. We focus our attention on 5-year tenure bins. This is to make these results comparable to later specifications, which combine tenure years for power reasons.

Like searches, hit rates are highest when officers have the least experience, beginning at 20.8 percentage points for officers with less than 5 years of tenure. Over the next 5 years, they remain stable, but they decline thereafter. This decline is more gradual than in searches. By an officer's 20th year on the force, his hit rate has fallen by about 7 percentage points, settling at about $\frac{2}{3}$ of the hit rate enjoyed by officers in the youngest bin.

3.2 Controlling for Shift Characteristics

These observational patterns may, however, be driven by something other than tenure. Tenure determines priority for assignments, so officers with higher tenure tend to work less demanding assignments. To control for this potential confound, as well as any other aspects of shift place, time, or duties that may be correlated with both tenure and search rate, we estimate a regression specification that controls for month-day-shift-beat effects (Ba et al. (2021)). This entails a separate fixed effect for each interaction of month, day of week, shift (night, morning, or evening), and beat. The goal of these controls is to only compare officers of different tenures who are working very similar shifts

The full set of month-day-shift-beat (MDSB) cells is sometimes sparse. Following Ba et al. (2021), we restrict our attention to MDSBs where we have data from at least two officers of different tenures. This restriction causes us to drop some observations from each of our regression specifications. We present the ultimate sample size used below each figure.

Our estimating equation is:

$$Y_i = \beta_t + \delta_{mdsb} + \gamma_o + \epsilon_i$$

When we are focused on searches per shift, our estimating sample is all shifts, and Y_i is the number of searches an officer makes in shift i. When we are focused on hits per search, our estimating sample is all searches, and Y_i is an indicator for whether an officer finds contraband in search i. In either case, β_t is a vector of tenure fixed effects, δ_{mdsb} are month-day-shift-beat fixed effects, γ_o are officer fixed effects, and ϵ_i is idiosyncratic error.

The primary identifying assumption in our design is that, *within* an MDSB, the unobservable characteristics of an officer's patrol environment are unrelated to their tenure. This is a reasonable assumption because the MDSB controls hold constant everything that CPD might reasonably use to assign officers across shifts. A violation of this assumption would require not only that e.g. the first Tuesday of each month be a consistently different patrol environment from the second Tuesday, but also that supervisors consistently assign higher-tenure officers with that difference in mind. Even this unusual behavior is logistically impossible due to the 4-days-on, 2-days-off schedule followed by patrol officers. This prevents officers from always working the same days in each month or week. We make an additional identifying

assumption through the additive structure of our estimating equation, which implicitly assumes no interaction between tenure effects and MDSB effects.

It is important to note the officer fixed effects in our specification. These implicitly include fixed effects for the cohorts in which officers entered CPD. This means that our results are not simply a consequence of comparing young officers to old officers, nor can they be driven by different training regimes when entering CPD. Instead, we are estimating tenure effects *within* officers as they age over the 5 years of stop data we observe. Our full battery of tenure effects is then the result of stitching those short panels of tenure effects into a single, synthetic career.

Panel A of Figure 2 plots the estimated tenure effects, β_t, on search rates. We include a tenure effect for trainees but use the officer's first post-trainee year as the reference category. We make this normalization because some later specifications exclude trainees. The standard errors are clustered at the officer level for this and all following regressions. The results in this figure confirm that the decline in search rates with tenure is not driven by changes in beat assignment as officers gain more seniority. The MDSB controls leave intact the secular decline that we saw in Figure 1. The magnitude of the decline by tenure 20 is, if anything, slightly larger than that observed in Figure 1.

Panel B of Figure 2 plots the estimated tenure effects on hit rates within the sample of searches. We focus on 5-year bins when studying hit rates because searches are far less common than shifts, especially among officers with higher tenures, and estimates based on 1-year bins are underpowered to detect meaningful changes. As in searches, these hit rate results confirm that the observed tenure effects for hit rates are not driven by changes in shift characteristics as officers gain seniority. The slow decline in hit rates observed in Figure 1 is preserved even under officer and MDSB controls. By the 20-to-25-year tenure bin, standard errors are such that we cannot rule out a rebound in hit rates, but the potential magnitude of any positive estimate in our confidence intervals is only around 1 percentage point.

3.3 Robustness

One potential confound in this analysis is that stop and search decisions are not made by individual officers. Officers typically patrol with partners, and both partners are typically listed on the ISR paperwork when a stop occurs.[1] An officer therefore does not have sole discretion over which stops to make or not make, and their outcomes may be influenced by a partner who does not have the same tenure.

We examine this issue in several ways. Figures A1 and A2 focus on the senior and junior officers in a car, respectively. Figure A3 focuses just on cars where all officers had the same tenure, which includes cars with only one officer. In all cases, the estimates are somewhat noisier than our estimates using all officers, but they show the same pattern of declining search and hit rates.

[1]The ISR data only records two officers. When more than two officers are involved in a stop, for example if it involved two separate cars, some may be left out of the data. When constructing the analysis data, we associate a stop with all officers known to be partnering with a listed officer at that time, which is helpful in stops involving two cars when the two listed officers belong to different cars.

Figure A4 further probes the robustness of our main result using an analysis sample that excludes COVID lockdowns by dropping all data in 2020 and later. It confirms the general pattern documented in our main results: both searches per shift and hits per search are declining with officer tenure.

4 A Model of Search Skill

The findings in the previous section are notable because they run counter to our typical understanding about the relationship between search frequency and hit rates. In a simple screening model, an officer who conducts fewer stops or searches than his peers should have a *higher* success rate. This arises from the assumption that officers who search less are electing only to make searches when the chance of finding contraband is particularly high. This could be because they perceive a particularly high cost or low benefit of searching.

Indeed, it is possible to recover this type of pattern across dimensions of the search data other than tenure. Figure 3 presents scatterplots that capture the observational relationship between search and hit rates with data aggregated across different dimensions. Panel A presents search and hit rates by month, Panel B by beat, Panel C by officer, and Panel D by officer tenure. In panel A, higher search rates in a month are associated with lower hit rates, a pattern consistent with simple screening models. Panels B and C show an overall positive correlation between search and hit rates across beats and officers. However, these plots are dominated by officers and beats with very few stops, and the estimated linear relationships are only somewhat positive. Panel D shows a strong positive relationship between search and hit rates across years of tenure.

The effects we document in the tenure dimension are therefore unusual and need a considered explanation. We require some officer characteristic other than search incentives/threshold to be changing over the career. In this section, we explore the theory that the tenure effects we document are driven by the evolution of officer skill. We first develop a static model that describes officer skill and how it affects search decisions and outcomes. We then consider how officers will choose skill when it is endogenous. We pay special attention to settings where officers may choose not to accumulate any skill, as these are empirically relevant. Finally, we sketch a dynamic version of this skill choice model in order to illustrate the importance of anticipated assignment changes late in the career.

4.1 Static Model

Consider a police officer who must make decisions about stopping pedestrians with the goal of discovering contraband. Each shift, he observes N candidates for stops, indexed by i. He must choose whether to stop each candidate, $D_i \in \{0, 1\}$. Upon being stopped, the true contraband state of a candidate, $Y_i \in \{0, 1\}$, is revealed. The average contraband carrying rate among candidate stops is $\mu = E[Y_i]$. When $Y_i = 1$, the officer receives a payoff which we normalize to 1. When $Y_i = 0$, the officer instead loses $\gamma > 0$.

The officer makes their decision on the basis of a signal, λ_i. When $Y_i = 1$, λ_i is drawn from a distribution $F(\lambda|\theta)$. When $Y_i = 0$, λ_i is instead drawn from $F(\lambda|\theta = 0)$.[2] Here, θ is a measure of the officer's skill. That is, his ability to distinguish pedestrians with and without contraband. We place the following conditions on $F(\lambda|\theta)$.

1. $F(\lambda|\theta)$ is continuously differentiable with respect to θ
2. $F(\lambda|\theta)$ satisfies the monotone likelihood property: for any $\theta_1 > \theta_2$, $\frac{f(\lambda|\theta_1)}{f(\lambda|\theta_2)}$ is strictly increasing in λ.
3. Increasing θ induces some rightward shift in the entire distribution: $\frac{\partial}{\partial\theta}F(\lambda|\theta) \leq 0 \; \forall \; \lambda, \theta$
4. There is no effect of skill in the extreme tails of the signal distribution: $\lim_{\lambda\to\infty} \frac{\partial}{\partial\theta}F(\lambda|\theta) = \lim_{\lambda\to-\infty} \frac{\partial}{\partial\theta}F(\lambda|\theta) = 0 \; \forall \; \theta$

Together, these conditions on $F(\lambda|\theta)$ communicate a notion of skill that is smooth, consistent, and requires discretion. The first condition means that small changes in the officer's skill create small changes in his ability to find contraband - no sudden jumps. The second and third condition ensure that this improvement applies across a wide range of scenarios, and that there are no situations where the officer becomes worse at finding contraband as his skill increases. Finally, the fourth condition states that accumulating skill only has an effect if the officer is going to use it. When the critical signal, λ, approaches positive (negative) infinity, this means that the officer is searching no (all) pedestrians. In these scenarios skill ought to have no effect on outcomes. Note that Anwar and Fang (2006) develop a model like this one but allow for only one θ for all officers. With this restriction, only assumption (2) is relevant.

One canonical example of a distribution with the properties described here is a standard normal with a mean shift in θ (Macmillan and Creelman (2004)). That is, $F(\lambda|\theta) = \Phi(\lambda - \theta)$. This is sometimes called a "binormal" distribution because λ_i is drawn either from a standard normal (when $Y_i = 0$) or a standard normal shifted to the right by θ (when $Y_i = 1$).

Given the signal λ_i, the officer forms a Bayesian posterior estimate of the probability of contraband:

$$q(\lambda_i) = \frac{\mu f(\lambda_i|\theta)}{\mu f(\lambda_i|\theta) + (1-\mu) f(\lambda_i|0)}$$

The officer then makes a stop in any case where, given the probability of contraband, his expected utility is positive, $q(\lambda_i) - \gamma(1 - q(\lambda_i)) > 0$. This can be rearranged to $q(\lambda_i) > \frac{\gamma}{1+\gamma}$ and plugged into the Bayesian prediction to get a decision rule just in terms of λ and primitives. Search if:

$$\frac{f(\lambda|\theta)}{f(\lambda|0)} \geq \gamma \frac{1-\mu}{\mu}$$

[2]Without loss of generality, we will consider values of $\theta \geq 0$. Values of $\theta < 0$ can also create separation between the $Y_i = 1$ and $Y_i = 0$ signalling distributions, but this confuses the intuition that more θ means more informative signals.

Appendix B shows that the general decision rule is solved by a unique λ^* where the officer searches all pedestrians with signals $\lambda > \lambda^*$. In the binormal specification, there is a closed-form expression for λ^*:

$$\lambda^*(\theta) = \frac{\theta}{2} - \frac{\ln(\frac{\mu}{\gamma(1-\mu)})}{\theta}$$

Appendix B also confirms that in this model, as in the basic screening model, increasing the cost of a failed search leads to fewer searches with a higher average success rate. However, the new skill dimension need not follow this pattern. Increasing an officer's skill typically increases his hit rate, but it has an inconsistent effect on search rate. Intuitively, an increase in skill produces two changes. First, the officer knows to forego certain bad stops that he would have made previously. Second, the officer knows to initiate certain good stops he would have ignored previously. How these two effects change overall search rates depends on which one is larger.

Which effect dominates the change in search rates depends on the types of mistakes officers tend to make when they have no skill endowment. Appendix B shows that this, in turn, depends on whether the return to a random search, $\mu - \gamma(1-\mu)$, is positive or negative. When it is positive, an officer with no means of differentiating good searches from bad will opt to search all pedestrians who show probable cause. Conversely, when $\mu - \gamma(1-\mu)$ is negative, an uninformed officer will choose to search no pedestrians. We theorize that a decrease in skill over the career leads to fewer stops, which is consistent with a negative return to a random stop. This makes sense in our setting, which is characterized by close monitoring of investigatory stops and skepticism about the value of widespread stop and frisk as a policing tactic.

4.2 Endogenizing Skill Investment

The next step in developing the model is to give the officer a choice in the level of θ they have. Suppose the officer starts with an initial endowment of skill, θ_0. The officer can select an investment in skill $a \geq 0$ but must pay a convex cost $c(a)$ with $c(0) = 0$ and $c'(0) = 0$. The officer then uses the post-investment value of $\theta = \theta_0 + a$ across N potential searches. This leads to the following optimization problem:

$$\max_{a \geq 0} N[\mu(1 - F(\lambda^*(\theta_0 + a)|\theta_0 + a) - \gamma(1-\mu)(1 - F(\lambda^*(\theta_0 + a)|0)] - c(a)$$

Across each of N potential searches, the officer will search and find contraband in a proportion $\mu(1 - F(\lambda^*(\theta_0 + a)|\theta_0 + a)$, earning a payout of 1. In another proportion $(1-\mu)(1-F(\lambda^*(\theta_0+a)|0)$, the officer will search and not find contraband, paying a penalty of $-\gamma$. The officer can improve his ability to discern guilty from innocent pedestrians, and this can potentially affect both the number of searches the officer makes, through $\lambda^*(\theta_0 + a)$, and the success rate of those searches, through $F(\lambda|\theta_0 + a)$. In exchange for any investment, the officer must pay $c(a)$.

As we have written the officer's problem, large investments in skill do nothing to reduce opportunities to apply that skill to stopping or searching pedestrians. This is in contrast to many models of skill accumulation, where e.g. education and labor must share a time budget. We do not force skill investment and pedestrian stops to compete because police officers have many responsibilities, of which pedestrians stops are only a small part. In our data, officers only initiate investigatory stops in 15% of shifts, less than once a week. Accurately modeling the time trade-off between the accumulation and application of skill in this environment would require giving serious attention to these other responsibilities and their respective shadow prices of time. Instead, we indirectly model the officer's time allocation problem with a generic cost function. This says that taking more time for skill investment makes it "harder" for the officer to fulfill his other responsibilities, but it does not specify precisely how.

One thing to notice about the officer's problem is that N is no longer neutral. In simpler problems of this kind, the number of potential searches serves only to scale the officer's payout, so it is immaterial to the trade-off between successful and unsuccessful searches. However, once the officer can make a costly investment that improves his performance across all searches, it matters how many such searches he will have the opportunity to make. Officers expecting to encounter many pedestrians who give probable cause for a search will benefit much more from an investment in θ than will officers expecting to encounter very few such pedestrians.

In Appendix B, we discuss the officer's optimal choice of skill investment. We find there is always at least one local maximum and possibly multiple maxima. This means that the officer's optimal investment as a function of other parameters may not be smooth.

We present an example using the binormal model. Figure 4 shows the marginal benefit curve for the binormal model with $\mu = 0.5$, $\gamma = 3$, and $\theta_0 = 0$. We also plot two marginal cost curves associated with cost functions $c(a) = \frac{0.1a^2}{2}$ and $c(a) = \frac{0.2a^2}{2}$. For the lower cost curve, marginal cost and marginal benefit intersect three times. The first time is very close to $a = 0$. This is not precisely the degenerate solution discussed in Appendix B, but it is similar in spirit. The officer is already searching *nearly* every pedestrian, so investment in skill is *nearly* worthless. The second intersection is a local minimum. The third is a significant investment, representing the equilibrium where the officer makes an effort to distinguish innocent from guilty pedestrians. When a local maximum of this kind exists, the officer typically prefers it to the nearly degenerate equilibrium. However, the higher cost curve in Figure 4 shows that the high-effort equilibrium may not exist. In this example, it stops existing because the cost of investing in skill becomes too high. Below, we will discuss other reasons why the high-effort equilibrium may not exist.

One point emphasized by this analysis is that near-degenerate equilibria where officers make trivial investments in skill are a credible threat. When initial skill, θ_0, is low, officers have little basis for distinguishing pedestrians, so they choose either to stop most or let most go. That leads to extreme values of λ^*. Marginal skill investments then affect only a small number of potential stops, so they are not worth the cost. Officers will therefore make meaningful investments only if it is feasible to jump all the way to a high-effort equilibrium that is not guaranteed to exist.

Figure 5 shows graphically how increasing initial skill can eliminate near-degenerate equilibria. It depicts the same system as Figure 4 except with $\theta_0 = 1.5$. Now that even small investments can produce appreciable results, the optimal investment becomes continuous. High costs still discourage accumulating skill, but there is no sudden drop into minimal investment. All else equal, officers who are choosing not to invest in skill may be induced to by mandatory training, even if it only partially closes the gap between zero investment and optimal investment.

4.3 Dynamic Considerations

In this section, we expand the search skill investment model discussed in the previous section into a multi-period dynamic model. The purpose of this section is not to develop this model in full detail but rather to gain more insight into why the data indicate that search skill begins to decay after just a few years on the force.

Consider an officer who intends to work for T years. After a mandatory training period, he begins with a skill endowment of θ_0. Each period, $t \in \{1, \ldots, T\}$, he must choose an investment, a_t, which is added to his stock of skill before any stop decisions are made that period. Simultaneously, his stock of skill depreciates according to an additive term δ. Hence, the officer's law of motion for skill is:

$$\theta_t = \max\{0, \theta_{t-1} - \delta + a_t\}$$

The officer makes investment choices while considering the entire path of the crime environment over his anticipated tenure, N_t and μ_t for $t \in \{1, \ldots, T\}$. We acknowledge it is somewhat implausible that officers will have perfect knowledge of how rates of suspicious pedestrians and contraband carrying will evolve over time. However, uncertainty about overall crime trends is not our focus. Instead, we wish to call attention to a more predictable way that N_t and μ_t tend to change over an officer's tenure. Junior officers are typically placed in the Chicago neighborhoods with the highest crime rates. As officers gain seniority, they get priority for transfers to districts with less criminal activity. An individual officer therefore expects both N_t and μ_t to decline as his career progresses.

Putting these elements together with a standard discount rate, β, we can write the officer's value function in a period $t < T$:

$$V_t(\theta_t) = N_t[\mu_t(1 - F(\lambda^*|\theta_{t-1} - \delta + a_t) - \gamma(1 - \mu_t)(1 - F(\lambda^*|0)] - c(a_t) + \beta V_{t+1}(\theta_{t-1} - \delta + a_t)$$

and in $t = T$

$$V_T(\theta_T) = N_T[\mu_T(1 - F(\lambda^*|\theta_{T-1} - \delta + a_T) - \gamma(1 - \mu_T)(1 - F(\lambda^*|0)] - c(a_T)$$

This sketch of a more involved dynamic model highlights two potential explanations for low investment in skill beyond high learning costs and low initial endowments. The first is depreciation. Naturally, if the skill accumulation technology has a high rate of depreciation, it will be more difficult to acquire and maintain a large stock of skill. This is even more important if depreciation is not additive, as we have modeled it above, but multiplicative, which imposes a greater penalty the more skill an officer carries. For our purposes, however, we will roll depreciation into a broader notion of the officer's learning technology. Insofar as that includes explicit training programs, it is no great insight that the effectiveness of these training programs depends not just on how much skill they impart in the moment but on how well that knowledge is retained over the long run.

Perhaps more interesting is the path of N_t and μ_t over an officer's career. As described above, the typical pattern of neighborhood assignment starts officers with high values of N_t and μ_t, but officers anticipate transitioning to a quieter neighborhood with lower values of N_t and μ_t after they have accumulated enough seniority. Officers are forward looking, so they anticipate that a large stock of contraband-finding skill will be largely useless after this reassignment. This can depress skill investment in early periods relative to officers who do not anticipate this reassignment.

The simple simulation in Figure 6 illustrates this point. In the "High then Low" scenario, all officers begin their careers in a "high-crime neighborhood" with $\mu = 0.3$ and $N = 2$. Officers work in this neighborhood for the first 5 years of a 20 year career. For the last 15 years, they work in a "low-crime neighborhood" with $\mu = 0.06$ and $N = 0.5$. Other parameters are set at $\beta = 0.98$, $\gamma = 2$, $\delta = 0.1$, $\theta_0 = 0.5$, and $c(a) = \frac{0.1a^2}{2}$.

The first row of graphs in Figure 6 show the results. Officers with this career path never make significant investments in search skill. Though skill could be useful in the high-crime neighborhood, officers anticipate moving to the low-crime neighborhood, where it will be less useful, so they don't invest even when working in the high-crime neighborhood. This has two important effects. First, search rates in the upper-left panel tend towards extremes. Officers search many pedestrians in the high-crime neighborhood because they have little basis for selecting searches, but their expected payoff from searching at random in that neighborhood is positive. Conversely, officers search few pedestrians once they transition to the low-crime area in their 6th year because they continue to have little basis for selecting searches, but now their expected payoff from a random search is negative. The second effect is that searches, when they happen, are no more likely to succeed than if officers were searching at random.

The "High or Low" scenario in the second row of graphs in Figure 6 suggests an alternative way of assigning officers over neighborhoods and across time. In this scenario, one quarter of all officers are assigned to the high-crime neighborhood for the entirety of their careers. The other three quarters of officers work in the low-crime neighborhood, likewise for their entire careers. Now, officers in the high-crime neighborhood have sufficient incentive to invest in skill. This simultaneously reduces search rates and increases contraband hit rates relative to the "High then Low" scenario. This is because officers can now have the skill to ignore some pedestrians who potentially appear suspicious but are not actually carrying contraband. The situation in the low-crime neighborhood is similar to before. Both search

and hit rates remain low. Even so, the combined average effect across the two neighborhoods is a reduction in search rates and an increase in contraband hit rates.

We acknowledge that there are significant institutional barriers to implementing the second scenario in actuality. Police unions have a strong commitment to ensuring equal treatment of their members, both in job duties and in pay. They would therefore stand against a system that "tracked" officers into permanently different assignments. Even so, we hope that the hypothetical scenarios presented here illustrate the importance of career trajectories to the investments that officers make in their own skill.

5 Conclusion

In this paper, we showed robust evidence that pedestrian search rates and hit rates *both* decline as CPD officers accumulate tenure on the force. These effects persist even in the presence of officer and month-day-shift-beat fixed effects. This joint pattern of search and hit rates cannot be explained by a simple screening model. Our proposed explanation is that officers rely on a stock of "search skill" that decays over time, and that the typical career path in CPD disincentives investment in this skill.

While still speculative, this theory ties together much of what is currently known about pedestrian stops. Pierson (2020) found large differences in traffic search and hit rates across jurisdictions, far in excess of the part that could be explained by the races of stopped motorists. Gelman et al. (2007) find large differences in pedestrian search rates across NYC precincts that were difficult to rationalize with a simple screening models. Feigenberg and Miller (2021) find that the volume of searches conducted by a given officer is largely uncorrelated with the hit rate of those searches. All these findings, along with our finding of a positive correlation between search and hit rates over the career, are consistent with a world where officers have a limited ability to distinguish high-quality from low-quality stops. There is some set of plausible stops (i.e. those that can show probable cause), but officers draw stops from this set essentially at random. In this world, the return from a random stop is a crucial determinant of officer behavior and may make a contribution to disparate stop rates across populations. When going from a "low-crime" to a "high-crime" neighborhood, a low-skill officer's decision rule can easily flip from "no stops" to "all stops."

Going forward, it is important to consider how the problem of low officer search skill may be addressed. Our model suggests that baseline training in how to select pedestrians to search can pay dividends insofar as it pushes some officers maintain or even increase an investment in search skill that would have been too large to build up from zero. To our knowledge, however, explicit training in search selection is sparse, as is research on its effectiveness. Our model also recommends a policy that pushes officers to specialize in situations where maintaining an investment in high search skill is worthwhile. This is an attractive idea because it only involves rearranging how officers are assigned to police districts, though it is not a given that the additional wages needed to compensate officers persistently assigned

to high-crime neighborhoods will equal the wage cut that would be accepted by the officers persistently assigned to low-crime neighborhoods.

References

Angelova, Victoria, Dobbie, Will and Yang, Crystal. (2023), Algorithmic Recommendations and Human Discretion. Working Paper.

Anwar, Shamena and Fang, Hanming. (2006). 'An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence', *American Economic Review* 96(1), 127–151.

Arnold, David, Dobbie, Will and Hull, Peter. (2022). 'Measuring Racial Discrimination in Bail Decisions', *American Economic Review* 112(9), 2992–3038.

Ba, Bocar A., Knox, Dean, Mummolo, Jonathan and Rivera, Roman. (2021). 'The role of officer race and gender in police-civilian interactions in Chicago', *Science* 371(6530), 696–702.

Becker, Gary S. (1957), *The Economics of Discrimination*, University of Chicago Press.

Becker, Gary S. (1993). 'Nobel Lecture: The Economic Way of Looking at Behavior', *Journal of Political Economy* 101(3), 385–409.

Chan, David C, Gentzkow, Matthew and Yu, Chuan. (2022). 'Selection with Variation in Diagnostic Skill: Evidence from Radiologists*', *The Quarterly Journal of Economics* 137(2), 729–783.

Feigenberg, Benjamin and Miller, Conrad. (2021). 'Would Eliminating Racial Disparities in Motor Vehicle Searches have Efficiency Costs?*', *The Quarterly Journal of Economics* 137(1), 49–113.

Gelman, Andrew, Fagan, Jeffrey and Kiss, Alex. (2007). 'An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias', *Journal of the American Statistical Association* 102(479), 813–823.

Jordan, Andrew. (2024), Racial Patterns in Approval of Felony Charges. Working Paper.

Knowles, John, Persico, Nicola and Todd, Petra. (2001). 'Racial Bias in Motor Vehicle Searches: Theory and Evidence', *Journal of Political Economy* 109(1), 203–229.

Macmillan, Neil A. and Creelman, C. Douglas. (2004), *Decetion Theory: A User's Guide*, Psychology Press.

Martin, Daniel and Marx, Philip. (2022). 'A Robust Test of Prejudice for Discrimination Experiments', *Management Science* 68(6), 4527–4536.

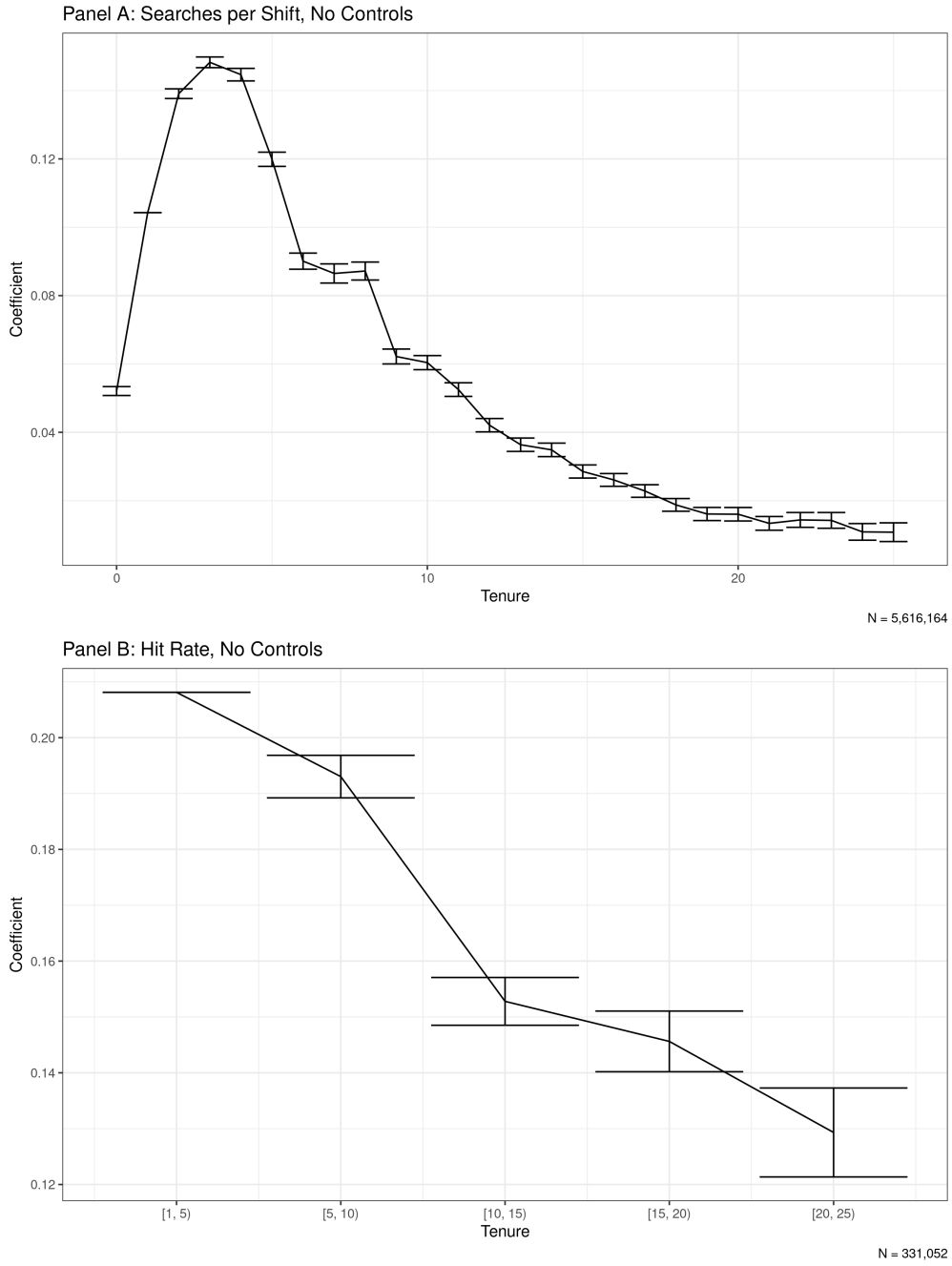
Marx, Philip. (2021). 'An Absolute Test of Racial Prejudice', *The Journal of Law, Economics, and Organization* 38(1), 42–91.

Miller, Joel, Quinton, Paul, Alexandrou, Banos and Packham, Daniel. (2020). 'Can police training reduce ethnic/racial disparities in stop and search? Evidence from a multisite UK trial', *Criminology & Public Policy* 19(4), 1259–1287.

Pierson, E., Simoiu C. Overgoor J. et al. (2020). 'A large-scale analysis of racial disparities in police stops across the United States', *Nature Human Behavior* 4, 736–745.

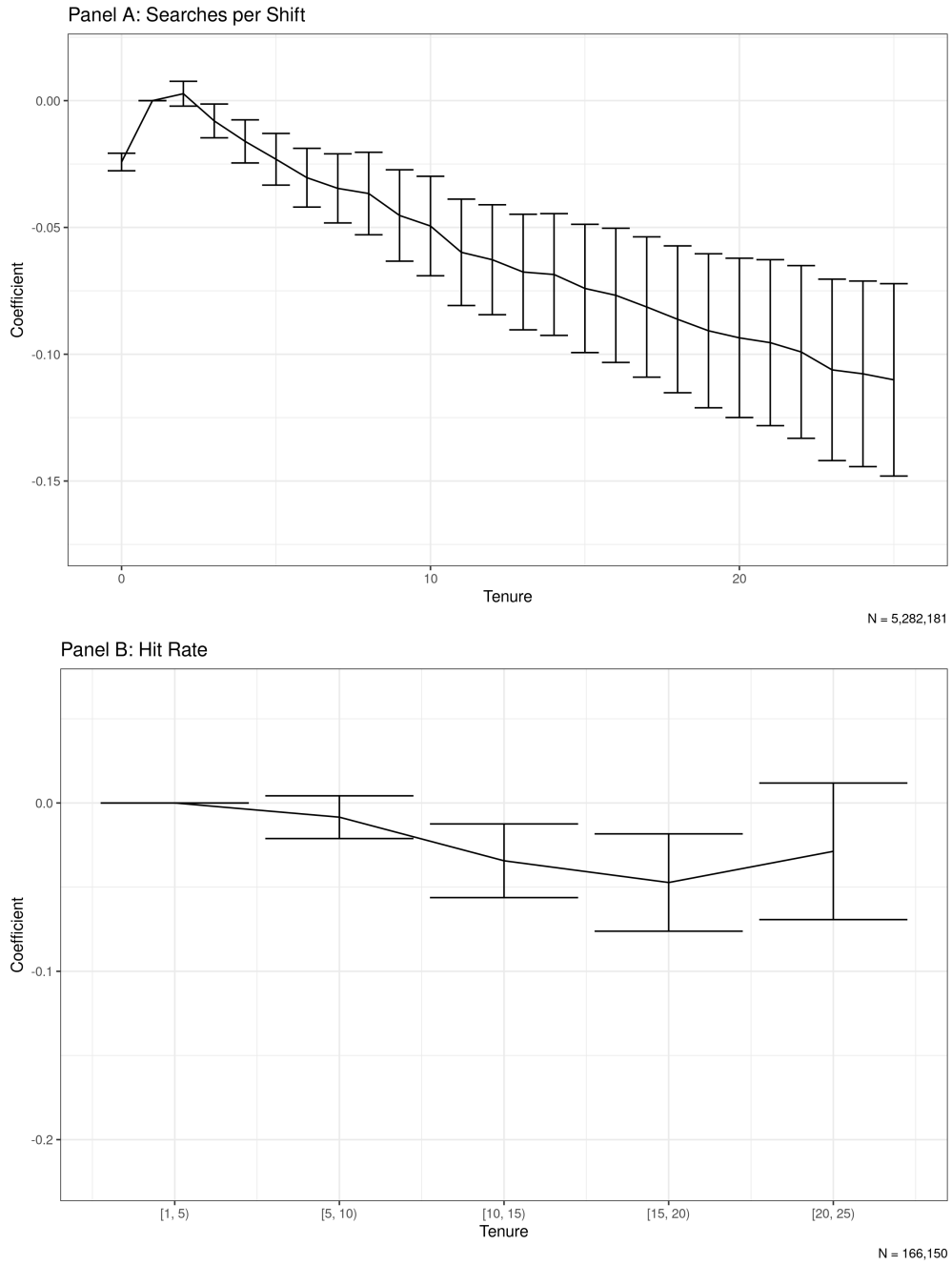
Figures and Tables

Figure 1: Tenure Effects: No Controls



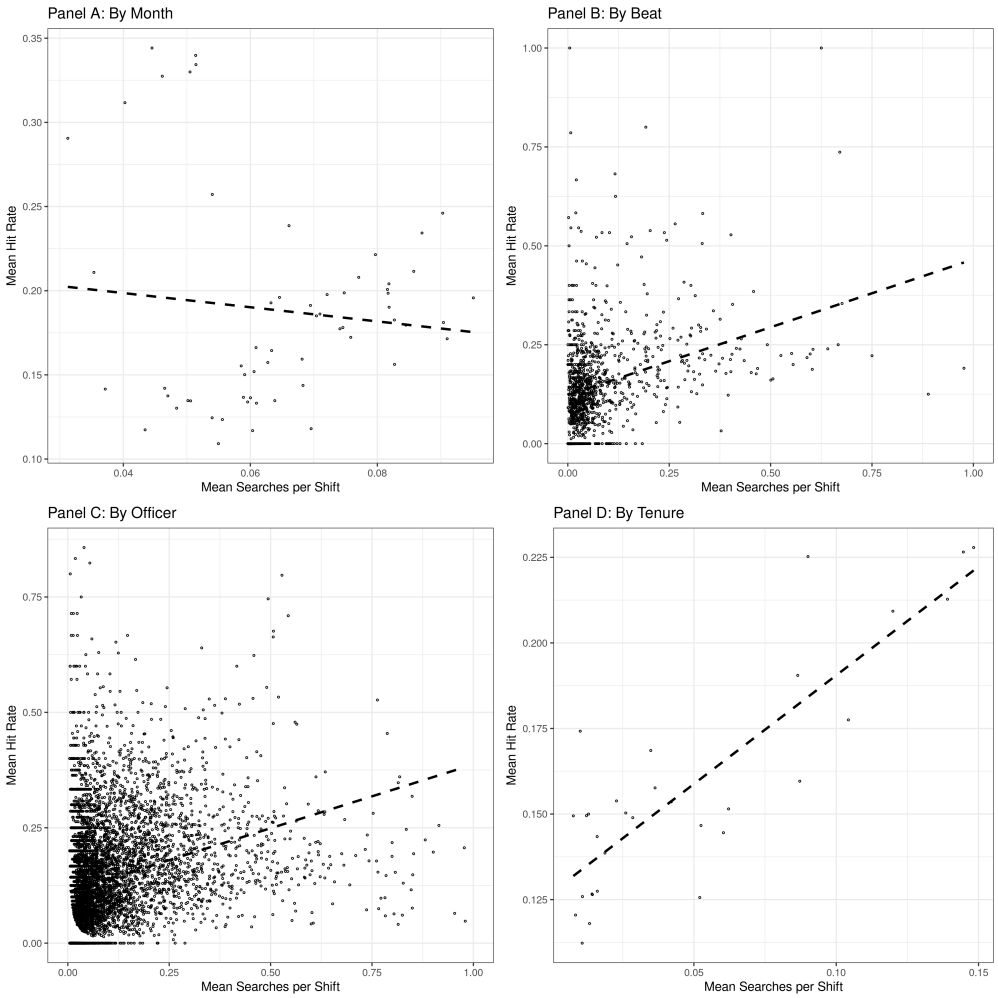
Panel A shows rate of pedestrian searches per shift for CPD patrol officers by years of tenure on the force. Panel B shows, across bins of tenure, the rate at which searches revealed contraband. Bars indicate 95% confidence intervals.

Figure 2: Tenure Effects: Main Specification



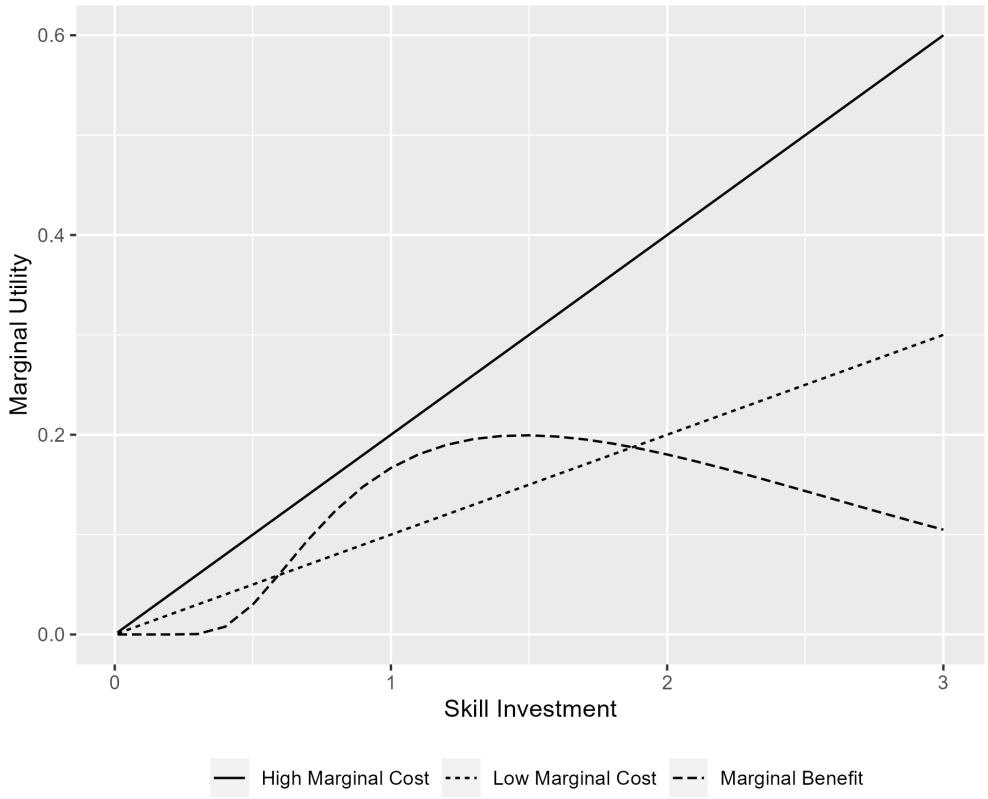
Panel A shows the estimated effect on searches of indicators for tenure by year. A tenure of 1 year is the excluded category. Regressions include month-day-shift-beat fixed effects and officer fixed effects. Panel B shows the estimated effect on hits (within the sample of searches) for bins of tenure. Tenures of 1-4 years are the excluded category. Relative to our main analysis sample, some observations are dropped because they do not appear in the same month-day-shift-beat cell as any observations with a different tenure. Bars indicate 95% confidence intervals.

Figure 3: Search Rate-Hit Rate Correlations



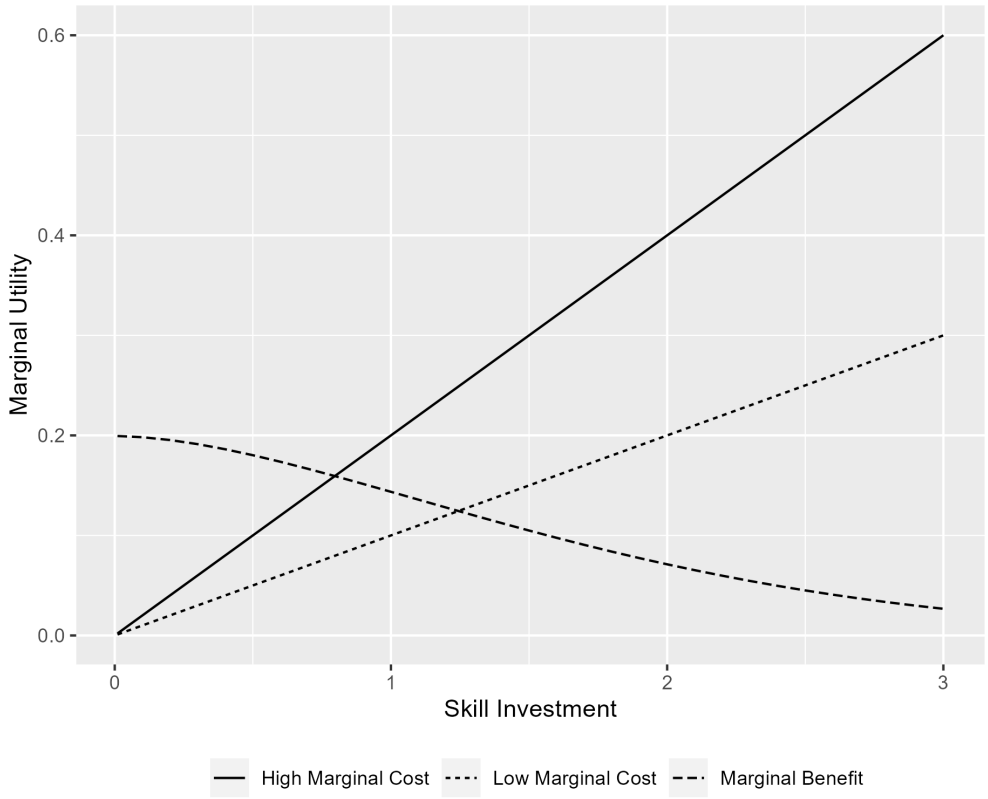
This figure shows scatterplots relating search rates and hit rates across various dimensions. Panel A calculates each rate by month in our data. Panel B aggregates by physical beat in the city. Panel C aggregates by officer. Panel C aggregates by year of tenure.

Figure 4: Solutions to Officer's Investment Decision



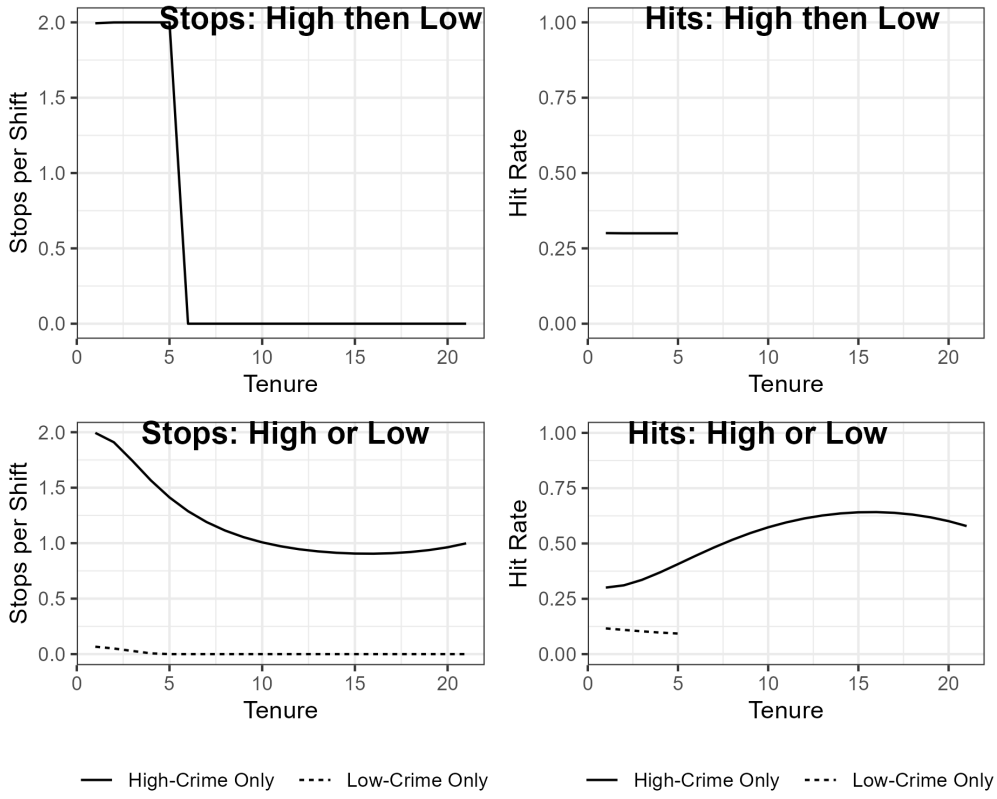
This figure visualizes the officer's decision to invest in search skill in terms of its marginal cost and marginal benefit. It uses parameters $\mu = 0.5$, $gamma = 3$, and $\theta_0 = 0$. The low marginal cost curve is generated by cost function $c(a) = \frac{0.1a^2}{2}$, and the high marginal cost curve uses $c(a) = \frac{0.2a^2}{2}$.

Figure 5: Solutions to Officer's Investment Decision: Greater θ_0



This figure visualizes the officer's decision to invest in search skill in terms of its marginal cost and marginal benefit. It uses parameters $\mu = 0.5$, $gamma = 3$, and $\theta_0 = 1.5$. The low marginal cost curve is generated by cost function $c(a) = \frac{0.1a^2}{2}$, and the high marginal cost curve uses $c(a) = \frac{0.2a^2}{2}$.

Figure 6: Stop and Hit Rate Paths



This figure shows the path of stop and hit rates given dynamic decisions about investment in search skill made by officers. The top row shows the "High then Low" scenario, in which officers spend the first five years in a "high-crime" neighborhood with $\mu = 0.3$ and $N = 2$ and the last 15 years in a "low-crime" neighborhood with $\mu = 0.06$ and $N = 0.5$. The bottom row shows the "High or Low" scenario in which officers spend their entire career either in the high-crime or low-crime neighborhood. Other parameters are set at $\beta = 0.98$, $\gamma = 2$, $\delta = 0.1$, $\theta_0 = 0.5$, and $c(a) = \frac{0.1a^2}{2}$.

Table 1: Shift-Level Summary Statistics

	Officer-Level	Car-Level
N	5,616,164	2,792,804
ISRs	0.15	0.18
Searches	0.03	0.04
PPDs	0.05	0.06
N Officers		1.41
N Trainees		0.16
Tenure	10.46	
FO Tenure		12.00
SO Tenure		10.40

This table gives summary statistics for our shift-level analysis data. The car-level data combines up to two officers into a car per shift, designating the officer with greater tenure the First Officer (FO) and the officer with lesser tenure the second officer (SO)

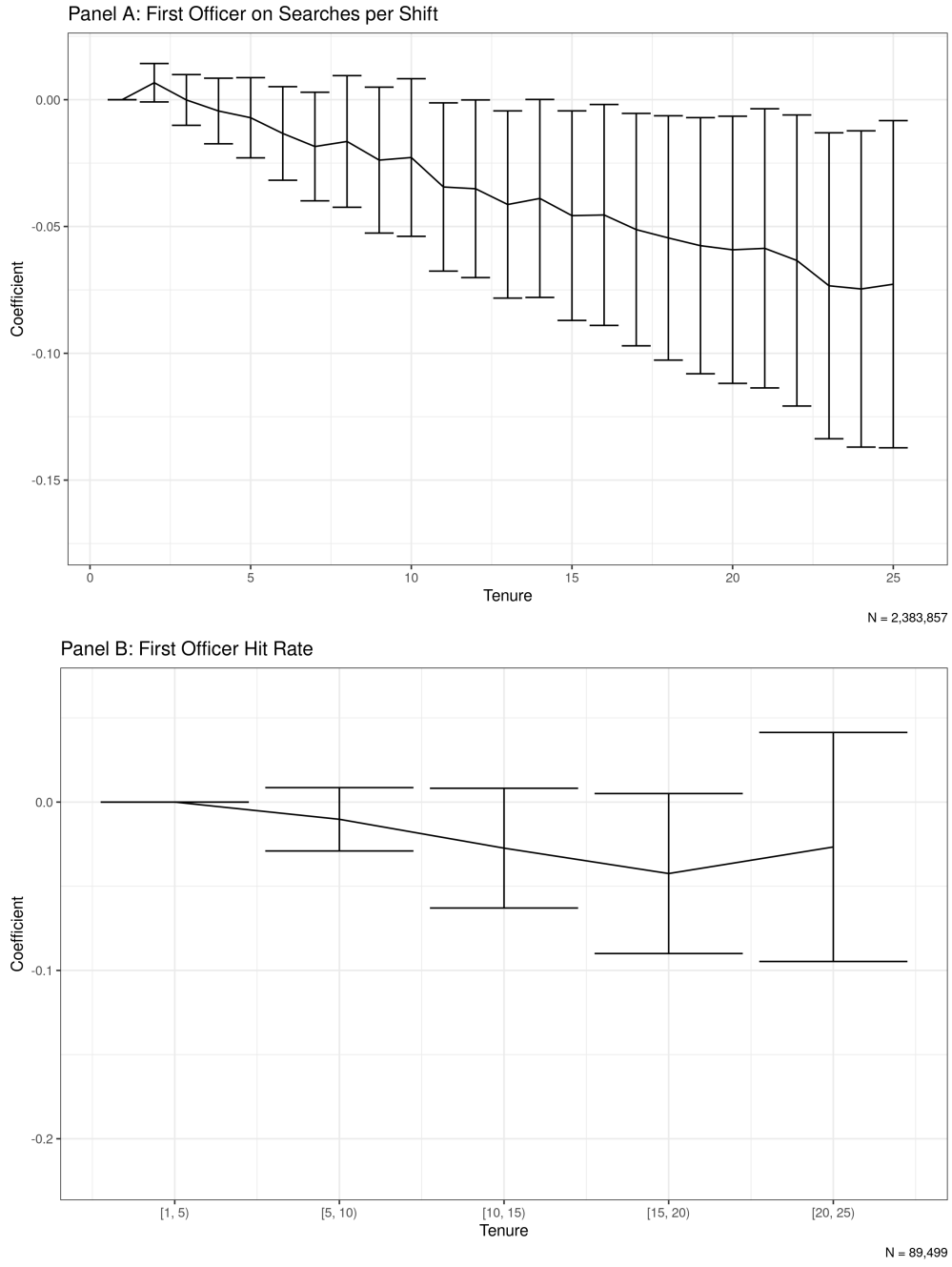
Table 2: Search-Level Summary Statistics

	Officer-Level	Car-Level
N	365,307	222,488
Hit Rate	0.19	0.19
Male	0.92	0.92
Black	0.73	0.73
N Officers		1.53
N Trainees		0.14
Tenure	5.68	
FO Tenure		7.19
SO Tenure		5.77

This table gives summary statistics for our search-level analysis data. The car-level data combines up to two officers into a car per search, designating the officer with greater tenure the First Officer (FO) and the officer with lesser tenure the second officer (SO)

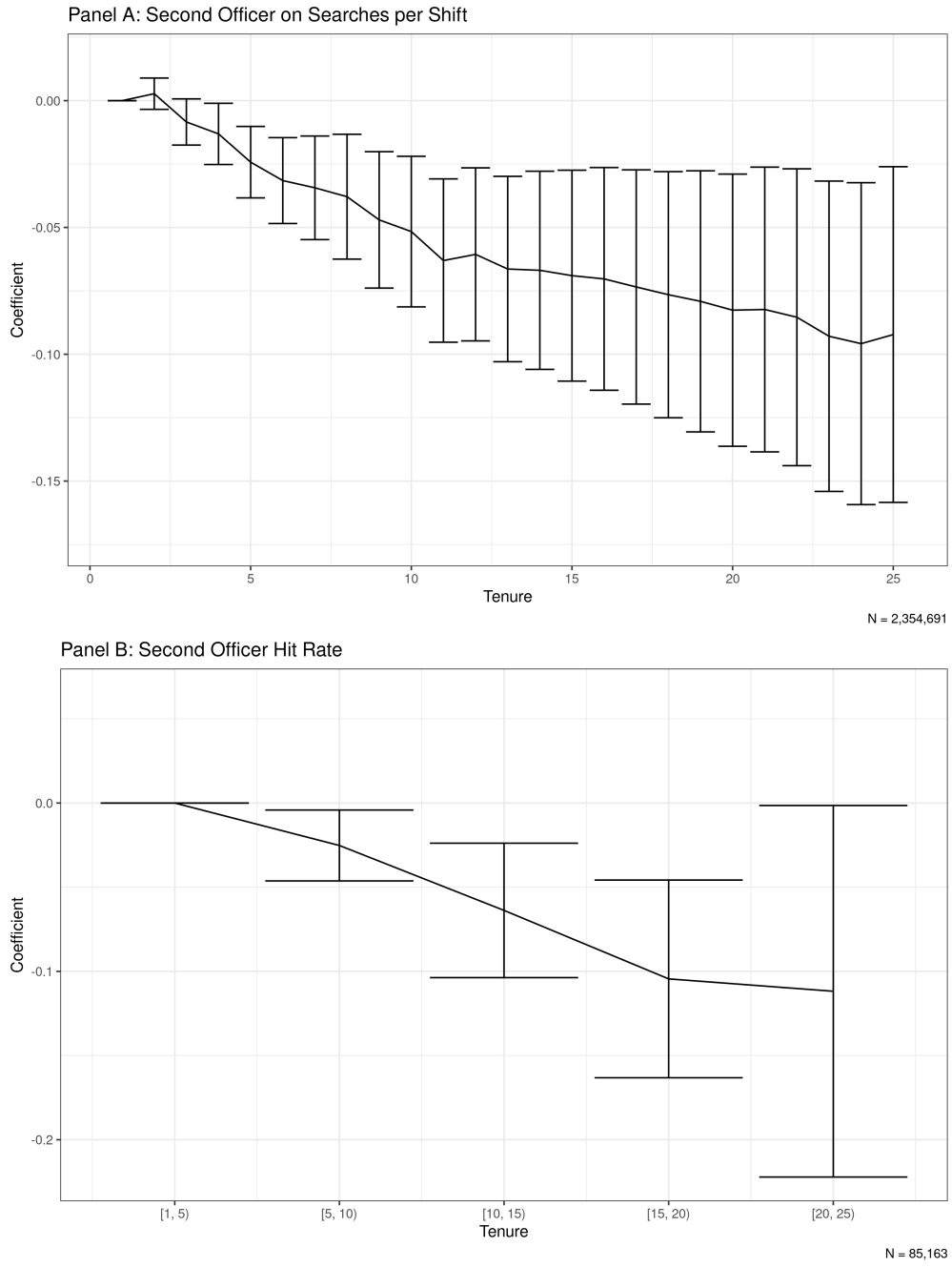
Appendix A: Additional Figures

Figure A1: Tenure Effects: Just First Officers



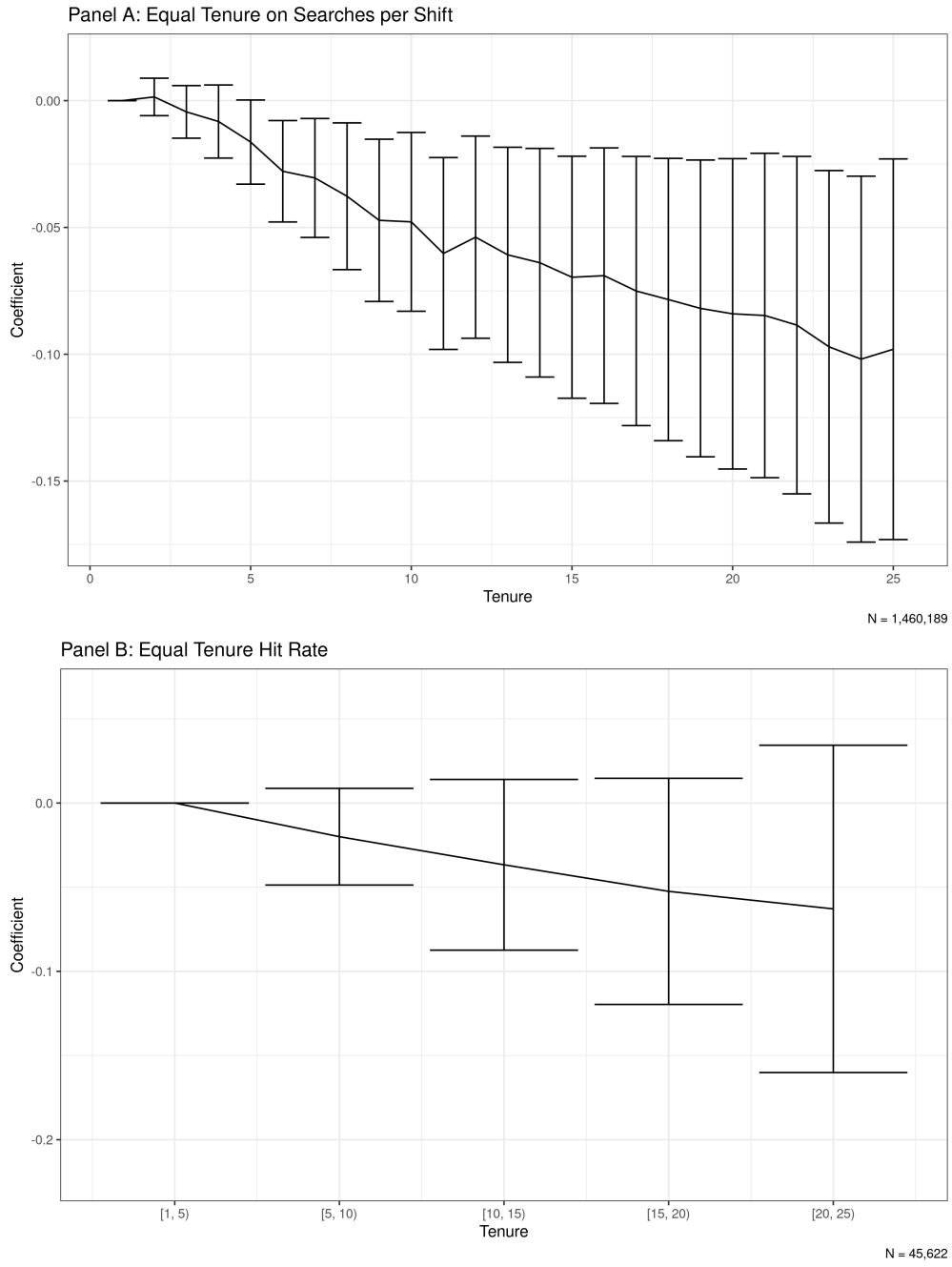
This figure replicates Figure 2 restricting attention to only the more senior officer in the car. When only one officer was present, he is counted as the senior officer.

Figure A2: Tenure Effects: Just Second Officers



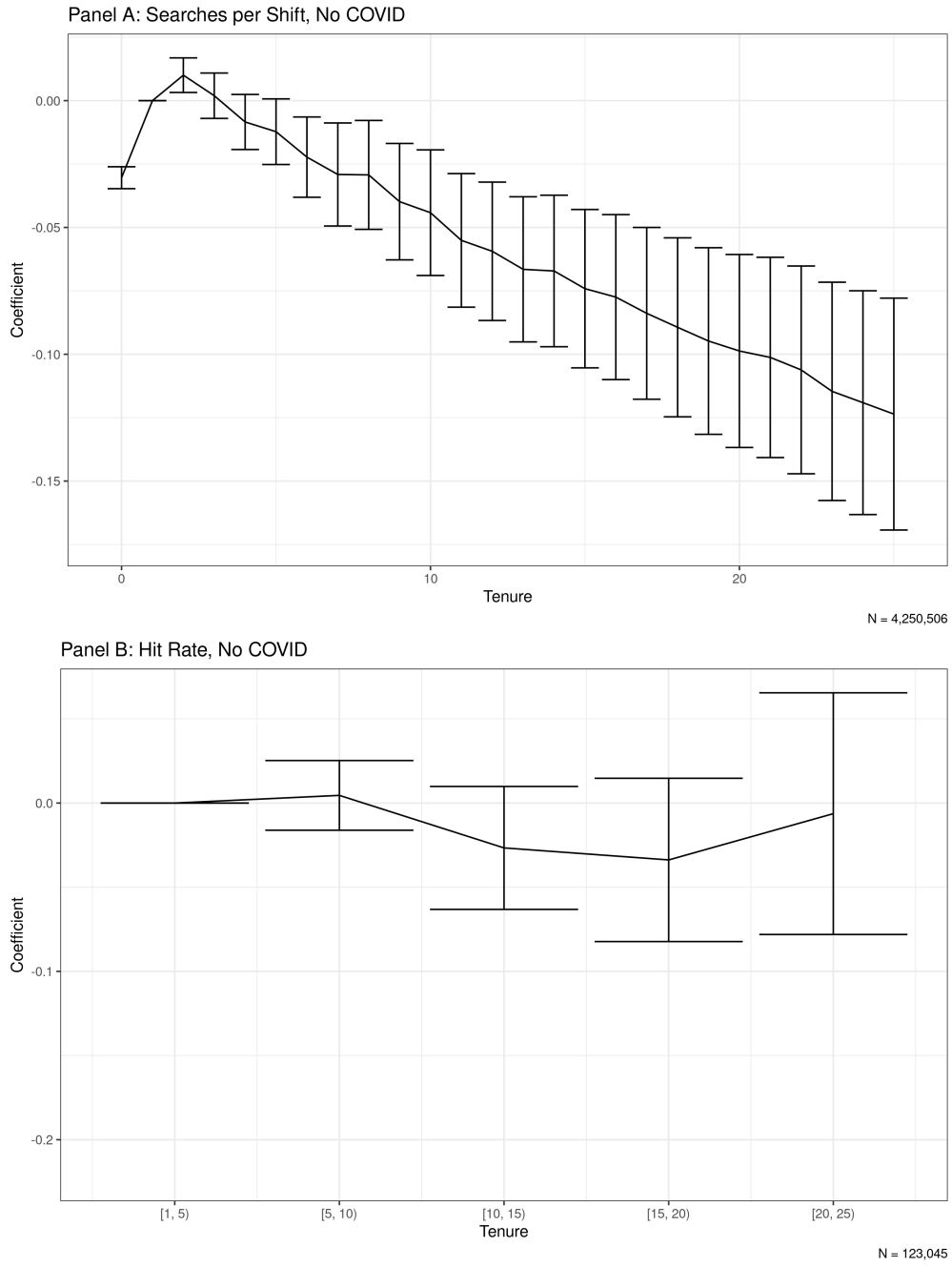
This figure replicates Figure 2 restricting attention to only the more junior officer in the car. When only one officer was present, he is counted as the junior officer.

Figure A3: Tenure Effects: Just Cars with Equal Tenure



This figure replicates Figure 2 restricting attention to only cars where both officers had the same tenure. When only one officer was present, he is included in this sample.

Figure A4: Tenure Effects: Dropping COVID Period



This figure replicates Figure 2 dropping data from 2020 and later.

Appendix B: Analysis of Search Skill Model

B.1 Solution to Static Model

Recall that the officer's rule is to search if:

$$\frac{f(\lambda|\theta)}{f(\lambda|0)} \geq \gamma \frac{1-\mu}{\mu}$$

The monotone likelihood property guarantees that this expression will hold with equality for at most one value of λ. If such a λ exists, define it as λ^*. If instead $\frac{f(\lambda|\theta)}{f(\lambda|0)} > \gamma \frac{1-\mu}{\mu}$ $\forall \lambda$, then the officer wishes to search every pedestrian, and we set λ^* to $-\infty$. Conversely, when $\frac{f(\lambda|\theta)}{f(\lambda|0)} < \gamma \frac{1-\mu}{\mu}$ $\forall \lambda$, the officer wishes to search no pedestrians, and λ^* is ∞. Regardless, these conventions ensure that the officer always searches all pedestrians who generate signals $\lambda > \lambda^*$.

A crucial determinant of the officer's behavior is whether μ is greater or less than $\gamma(1-\mu)$. This dictates the officer's preferences in a low skill environment. In the absence of any way to actually differentiate between guilty and innocent pedestrians, the officer must essentially search at random. This leads to a discovery of contraband (reward normalized to 1) in proportion μ of cases and a failure (reward $-\gamma$) in proportion $1-\mu$ of cases. Thus the expected reward from a random search is $\mu - \gamma(1-\mu)$. When this is positive, officers with no further information will prefer to search every pedestrian. When this is negative, they will instead prefer to search no pedestrians. This is clear from the expressions above. If the officer has no skill, then $f(\lambda|\theta) = f(\lambda|0)$, and the likelihood ratio is everywhere equal to 1. When $\mu - \gamma(1-\mu) > 0$, then the likelihood ratio is always greater than $\gamma \frac{1-\mu}{\mu}$, and officers always search. The opposite holds when $\mu - \gamma(1-\mu) < 0$.

B.2 Comparative Statics

Response to Search Costs

Consider the rule that sets the search threshold, λ^*:

$$\frac{f(\lambda^*|\theta)}{f(\lambda^*|0)} = \gamma \frac{1-\mu}{\mu}$$

By the monotone likelihood ratio assumption, the LHS of this equation is increasing in λ^*, and the RHS is clearly increasing in γ. Therefore:

$$\frac{\partial \lambda^*}{\partial \gamma} > 0$$

Next, note that the search rate is given by the fraction of cases searched when $Y_i = 1$ and the fraction of cases

searched when $Y_i = 0$, weighted by the probability of those events:

$$SR(\lambda^*) = \mu(1 - F(\lambda^*|\theta)) + (1 - \mu)(1 - F(\lambda^*|0))$$

which is clearly decreasing in $\lambda*$: $SR^{'}(\lambda^*) < 0$

Finally, the hit rate is given by the fraction of searches made where $Y_i = 1$:

$$HR(\lambda^*) = \frac{\mu(1 - F(\lambda^*|\theta))}{\mu(1 - F(\lambda^*|\theta)) + (1 - \mu)(1 - F(\lambda^*|0))}$$

Refer to Anwar and Fang (2006) footnote 29 for proof that $HR^{'}(\lambda^*) > 0$.

Combining these results, we can conclude that

$$\frac{\partial SR}{\partial \gamma} = SR^{'}(\lambda^*)\frac{\partial \lambda^*}{\partial \gamma} < 0$$

$$\frac{\partial HR}{\partial \gamma} = HR^{'}(\lambda^*)\frac{\partial \lambda^*}{\partial \gamma} > 0$$

Response to Search Skill

The analysis of how search and hit rates respond to changes in search skill, θ, is complicated substantially by the fact that λ^* also responds to changes in θ. Using the officer's choice condition, that relationship is:

$$\frac{\partial \lambda^*}{\partial \theta} = \frac{-\frac{\partial f(\lambda^*|\theta)}{\partial \theta}\frac{1}{f(\lambda^*|0)}}{\frac{\partial}{\partial \lambda^*}\frac{f(\lambda^*|\theta)}{f(\lambda^*|0)}}$$

The denominator of this object is always positive by the monotone likelihood property. The numerator is ambiguous. If $f(\lambda|\theta)$ is single-peaked, and increasing θ simply shifts $f(\lambda|\theta)$ to the right, then it is possible to build some intuition. Both of these conditions hold in the binormal model. $\frac{\partial \lambda^*}{\partial \theta}$ will be negative when λ^* is high because it is on the downward-sloping part of the signal distribution, so shifting that distribution further to the right increases $f(\lambda^*|\theta)$. Conversely, $\frac{\partial \lambda^*}{\partial \theta}$ will be positive when λ^* is low enough to be on the upward-sloping part of the signal distribution.

Next we derive, in response to a change in θ, the change in the fraction of potential searches where the officer makes a search and finds contraband (i.e. the numerator of the hit rate):

$$\frac{\partial}{\partial \theta}\mu(1 - F(\lambda^*|\theta)) = -\mu\frac{\partial F(\lambda^*|\theta)}{\partial \theta} - \mu f(\lambda^*|\theta)\frac{\partial \lambda^*}{\partial \theta}$$

The first term is positive and reflects the fact that, holding the search threshold constant, officers can identify more high-value searches when they have greater skill. The second term captures the effect of a change in search threshold

due to a change in skill.

Likewise, we can derive the change in the fraction of potential searches where the officer makes a search and does not find contraband (i.e. the remainder of the search rate):

$$\frac{\partial}{\partial\theta}\mu(1-F(\lambda^*|\theta)) = -(1-\mu)f(\lambda^*|0)\frac{\partial\lambda^*}{\partial\theta}$$

Here, only the change in search threshold matters. The distribution of signals for specious searches is fixed, so this change in search threshold is the only way for the officer to adjust on this margin.

The sum of these two terms is the change in the search rate due to an increase in θ. When $\frac{\partial\lambda^*}{\partial\theta} < 0$, this is unambiguously positive. The officer's main response to the increase in skill is to conduct more searches that he would have otherwise had a hard time identifying as high-value. Conversely, when $\frac{\partial\lambda^*}{\partial\theta} > 0$, the search rates may decline in response to an increase in θ. The officer's main response to the increase in skill is to stop searching some low-value pedestrians.

A sufficient condition for an increase in θ to increase hit rates is:

$$-\mu f(\lambda^*|\theta)\frac{\partial\lambda^*}{\partial\theta} \geq -(1-\mu)f(\lambda^*|0)\frac{\partial\lambda^*}{\partial\theta}$$

Using the officer's decision rule for setting λ^*, this can be reduced to:

$$-\gamma(1-\mu)f(\lambda^*|0)\frac{\partial\lambda^*}{\partial\theta} \geq -(1-\mu)f(\lambda^*|0)\frac{\partial\lambda^*}{\partial\theta}$$

which in turn reduces to

$$\gamma frac\partial\lambda^*\partial\theta \leq \frac{\partial\lambda^*}{\partial\theta}$$

A useful baseline to consider is $\gamma = 1$, in which case the expression above is certain to hold. In this case, the officer's utility is such that his objective is to maximize the hit rate. He will therefore not make any change to λ^* that causes a reduction in the hit rate.

When $\gamma \neq 1$, the officer's objective is no longer directly to maximize the hit rate, and the direction of $\frac{\partial\lambda^*}{\partial\theta}$ must be considered. When $\frac{\partial\lambda^*}{\partial\theta} > 0$, then the condition above holds when $\gamma < 1$. These two events are likely to coincide because a low searching cost will incline officers to make a lot of searches, putting them on the downward sloping part of $f(\lambda^*|\theta)$ where $\frac{\partial\lambda^*}{\partial\theta} > 0$. A similar argument holds for the reverse case. However, with the level of flexibility currently allowed to $f(\lambda^*|\theta)$ and the way it can respond to changes in θ, I do not offer a formal proof that increasing search skill must in all cases increase the observed hit rate.

B.3 Solution to Endogenous Model

We now consider the officer's endogenous choice of skill investment, a. One set of degenerate solutions to this problem is for the officer to make no meaningful effort to distinguish innocent pedestrians from guilty. He either searches all pedestrians he is legally allowed to, or he searches none. In this event, he has no use for skill, so his optimal choice is to invest nothing in it. Formally, solutions of this kind arise when the likelihood ratio without investment, $\frac{f(\lambda|\theta_0)}{f(\lambda|0)}$, never intersects with $\gamma\frac{1-\mu}{\mu}$. λ^* is driven to either ∞ or $-\infty$, $\frac{\partial}{\partial\theta}F(\lambda^*|\theta_0)$ goes to 0, and this equals the marginal cost of investment at $a = 0$.

The richer set of solutions to this problem are values of a such that some finite λ solves:

$$\frac{f(\lambda|\theta_0 + a)}{f(\lambda|0)} = \gamma\frac{1-\mu}{\mu}$$

This ensures an interior solution for $\lambda^*(\theta_0 + a)$, and through the Envelope Theorem, we can ignore any direct effects of changing a on $\lambda^*(\theta_0 + a)$. The first order condition for this set of solutions is therefore:

$$-\mu N[\frac{\partial}{\partial\theta}F(\lambda^*(\theta_0 + a^*)|\theta_0 + a^*)] = c'(a^*)$$

where $\frac{\partial}{\partial\theta}$ should be understood to mean the derivative with respect to the second argument in $F(\lambda|\theta)$, ignoring any direct change in λ due to changes in θ.

Here, the left hand side captures the marginal benefit of additional investment in search skill. First, hold constant the search threshold, λ, which also holds constant the number of unsuccessful searches. Then, an officer with more skill will discover contraband more often. This is because greater skill shifts more of the signalling distribution for guilty pedestrians beyond his search threshold. This effect is moderated both by the number of opportunities to search, N, and the frequency with which a random search opportunity can be expected to yield contraband, μ. The right hand side directly captures the marginal cost of investment in search skill.

We are guaranteed at least one local maximum. First notice that:

$$-\frac{\partial}{\partial\theta}F(\lambda^*(\theta_0)|\theta_0) \geq 0 \ = c'(0)$$

so marginal benefit begins at or above marginal cost.

Second, the standard bounds on a CDF ensure that:

$$\lim_{a\to\infty} -\frac{\partial}{\partial\theta}F(\lambda^*(\theta_0 + a)|\theta + a) < \infty = \lim_{a\to\infty} c'(a)$$

so marginal benefit ends below marginal cost. Leaving aside those cases where $-\frac{\partial}{\partial\theta}F(\lambda^*(\theta_0)|\theta_0) = 0$—these are the

cases with degenerate solutions of no investment—marginal cost will overtake marginal benefit from below at a local maximum.

Appendix C: Data Details

C.1 Overview

Ultimately, this sample generating process leads to four separate samples based on two levels of observation, shifts and searches, and two data configurations, long and wide. Shift-level data are used in regressions where the dependent variable is stops or searches per shift. Search-level data are used in regressions where the dependent variable is the probability of finding contraband. Long data contain an observation per officer per level of observation, e.g. if two officers are involved in a single search, the search will appear in the dataset twice. Wide data are organized by car, as officers typically share a squad car with a partner while on a shift. This topic will be revisited in greater detail below.

C.2 Raw Data

All four samples combine data from two sources: Investigatory Stop Reports (ISRs) and attendance data. ISRs are publicly available as a result of a consent decree, and may be downloaded here. Each report describes an investigatory stop in detail, including whether any searches were performed and whether those searches uncovered any contraband. The ISRs also include information about protective pat downs (PPDs) performed by officers, which are less thorough than a search and do not require probable cause. We will distinguish between searches and PPDs in our analysis, placing most of our attention on formal searches. Each ISR also records up to two officers who participated in the stop. Attendance data, which describe the time, location, duties, and other details associated with police officers' shifts were obtained via a FOIA request.

C.3 Extraction from Raw Spreadsheets

The first sweep of the data is done in Stata, using programs originally written for another project: clean_stops.do and clean_attendance.do. The main tasks are to pull the data in from across multiple spreadsheets and convert string fields into more tractable numerical fields. The only observations dropped at this step are attendance records associated with administrative staff who are not sworn officers. The sample sizes at the end of this step are as follows.

- **Stops:** 732,212
- **Attendance:** 38,372,817

C.4 Initial Cleaning

From here on, all data work is performed within R. At this stage, each data source is treated separately by make_stops.R and make_attendance.R, each of which converts variables (mostly dates and times) into more R-friendly formats and

constructs a variety of new variables used in our analysis. Make_stops.R does not drop any observations in this step. Make_attendance drops any shifts that officers were not present for (the dataset records both scheduled and unscheduled absences), reducing the sample size to 29,108,286.

C.5 Calculating Tenure

Make_attendance.R also performs the important task of calculating officers' length of tenure in years on each of their shifts. An officer with CPD typically spends 6 months in police academy, then a year as a trainee, during which they are always assigned alongside a more experienced officer, before becoming a full officer. The goal is to start counting from their first day as a trainee (meaning a tenure value of less than 1 indicates the officer is a trainee). However, there is evidence that in some cases this typical schedule is not followed (e.g. officers who started police academy in late 2019 waited longer than 6 months to join the force due to low staffing demands during lockdown).

All officers have a recorded start date indicating when they started police academy. The raw attendance data extend back to the beginning of 2008. For officers with start dates in 2008 or later, tenure is calculated beginning on the first day we observe a shift worked that is not associated with training in the academy. For officers with start dates before 2008, tenure is calculated beginning on the day 6 months after their start date.

C.6 Joining

From here on, the remaining tasks are performed in join_stops_attendance.R.

The next step is to join these two data sources. The attendance data run through 2020; restricting the stop data to this end date reduces the number of observations to 594,790. The stop data begin in 2016; restricting the attendance data to this start date reduces the sample size to 11,785,362.

The first step is to associate each stop with the associated shift(s). Each ISR records up to two officers involved in the stop. Melting the stop data to have an observation per officer per stop yields 1,082,094 observations. These observations are then matched with the attendance data via a rolling join: each officer-stop is associated with the latest shift worked by the same officer that began before the time of the stop. There is evidence of some imperfect joins: about 7% of stops are associated with shifts that were reported to have ended more than 2 hours before the time of the stop, and so are dropped, reducing the sample to 1,003,721 officer-stops. This is the basis of the search-level long dataset described in section C.1 (though at this point it includes stops on which searches did not occur).

Next, these data can be summarized at the shift level, summing numbers of stops, searches, found contraband, etc. to create the basis for the shift-level long dataset described in section C.1. No shifts are dropped, leaving the sample size at 11,785,362.

C.7 Identifying Cars

The most complex aspect of our sample generation is the identification of officers who are grouped together into a single car during their shifts to form the wide versions of our datasets. A typical squad car contains two officers along with, potentially, one or more trainees.

We start with the attendance data as it stood after the date restriction at the beginning of section C.6. Two restrictions are made to narrow our attention to shifts with relatively standard structures. Our analysis ultimately is based only on CPD's district units, the units who provide typical beat cops for each of Chicago's 22 districts. Dropping all other shifts reduces the sample to 6,800,309 observations. Then, shifts associated with abnormal car numbers are dropped. This includes both those that are indicative of officers not being in a position to make any stops at all (e.g. "ADMIN") as well as some that simply indicate shift duties that do not follow the typical structure of 1 or 2 officers in a squad car (e.g. "BIKE"). This restriction reduces the sample to 5,207,714 shifts.

Next, shifts are determined to share a car if they have the same start time (including date), watch number (CPD's system for ensuring that shifts start at different times of day while overlapping to avoid low coverage during transitions), car number, and beat number. Some of the resulting cars still fall outside of the typical structure: dropping shifts associated with cars containing more than 2 senior officers or 3 trainees reduces the sample to 5,046,265 shifts. Reshaping the sample to a car per observation results in a sample of 3,365,942 observations.

Then, these cars must be joined, as previously done with individual shifts, to stop data. Restricting the stop-level long dataset from section C.6 to stops made by officers on district units reduces the sample from 1,003,721 to 842,530 officer-stops. These have already been joined with shift data, so all that is left to do is see if these shifts have cars associated with them, which 776,378 of them do. Recall that this sample contains an observation per stop per officer. Reshaping it to an observation per stop per car (i.e. a stop involving 2 cars will appear twice but a stop listing 2 officers from the same car will appear only once) results in a sample with 512,855 observations. This is the basis for the search-level wide sample described in section C.1.

These stop data can then again be summarized at the car level for numbers of stops, searches, found contraband, etc. per car. This is the basis for the shift-level wide sample described in section C.1. No observations are dropped, leaving the dataset at 3,365,942 rows.

C.8 Final Restrictions

Finally, a few additional restrictions are made to narrow the focus of the analysis on stops and shifts of relatively similar natures. We describe each restriction and its result on the sample size for each of our 4 organizational structures. Note that some of the restrictions now applied to the long datasets were already applied to the wide datasets during their initial constructions.

- Shift-level, long
  - Starting size: 11,785,362
  - District units: 6,800,309
  - Title of “police officer”: 5,960,507
  - Typical watch number (1-4): 5,828,043
  - Typical beat number: 5,616,164
- Shift-level, wide
  - Starting size: 3,365,942
  - Both officers with title of “police officer”: 2,862,786
  - Typical watch number (1-4): 2,803,835
  - Typical beat number: 2,792,804
- Search-level, long
  - Starting size: 1,003,721
  - District units: 842,530
  - Search or PPD conducted: 373,627
  - Title of “police officer”: 367,711
  - Typical watch number (1-4): 366,350
  - Typical beat number: 365,307
- Search-level, wide
  - Starting size: 512,855
  - Search or PPD conducted: 229,057
  - Both officers with title of “police officer”: 223,733
  - Typical watch number (1-4): 222,937
  - Typical beat number: 222,488