Appendix

Statistical Models and Analysis

A statistical model of police shootings

Structural model. The 800,000 police officers in the United States have roughly 80 million reportable contacts with the public and millions of additional public contacts per year (Davis, Whyde, and Langton 2018). A small fraction of interactions involves police officers discharging their firearms. Approximately 1,000 of these discharges result in a fatality per year. In almost all environments and for most police officers, the risk of shooting is exceedingly low. Yet in select moments there are factors in place that increase the risk for a police shooting. Information for distinguishing the officer characteristics that lead to police shootings is embedded in the data on these tens of millions of interactions; in these interactions' individual, situational, organizational, community, and legal factors; and in which officers in which interactions discharged their firearm.

Consider the risk of shooting for an officer with characteristics $\mathbf{x}$ (e.g., age, race, sex, experience, prior involvement in shootings) in an environment $\mathbf{z}$, which captures the shared situational, organizational, community, and legal factors (e.g., time, place, lighting, suspect features, governing policies and laws, community conditions). Let $R = 1$ indicate that the officer discharged their firearm, and $R = 0$, that they did not. A standard statistical model for the probability of shooting is

$$\log \frac{P(R = 1|\mathbf{x}, \mathbf{z})}{P(R = 0|\mathbf{x}, \mathbf{z})} = \alpha'\mathbf{z} + \beta'\mathbf{x} \qquad (1)$$

This logistic regression model separates the log odds of a shooting into an environmental contribution ($\alpha'\mathbf{z}$) and an individual officer contribution ($\beta'\mathbf{x}$). For most values of $\mathbf{z}$, $\alpha'\mathbf{z}$ is a large negative number signaling that almost all environments carry virtually no risk of shooting. Some components of $\mathbf{z}$ greatly increase the risk of a police shooting, such as an armed suspect or officers being fired upon. The coefficients in α associated with those environmental factors that exacerbate the risk of a shooting would be large and positive. $\exp(\beta'\mathbf{x})$ indicates by how many times the odds of shooting increases over the baseline odds $\exp(\alpha'\mathbf{z})$ for an officer with characteristics $\mathbf{x}$. If the environment drives all the risk of shooting and no specific officer characteristic influences the risk of shooting then β, which measures how strongly each officer characteristic influences shooting risk, will equal 0. In other words, when all the correlates of a decision to shoot are environmental, the officer's individual characteristics have no correlation with the decision.

An analogous model for the number of rounds fired can be estimated with different statistical assumptions, using a Poisson regression model that explains differences in how many bullets were shot. This model denotes "R" as recording the number of rounds that an officer with characteristics $\mathbf{x}$ fires in an environment with characteristics $\mathbf{z}$.[i] If any environmental factor is correlated with both an officer characteristic and the odds of shooting (or number of fired rounds), then failing to properly model the environment can introduce bias in the estimate of the effect of officer characteristics on β. This is specifically the statistical problem that generated the cautionary language about confounding in prior research. Failing to adequately measure and model the environmental factors will bias estimates of the β for individual officer factors. And since for estimating the effect of officer characteristics, we have exclusive interest in β, this problem is a nuisance. But they are a nuisance that cannot be ignored.

Decomposing the information in police shooting data. To estimate both the effects of environmental and individual officer characteristics (α and β), we need to randomly select a moment in time and a place involving one or more police officers and record the following facts (in the form of $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}, R_1, \ldots, R_n$) for that instance: the characteristics of each of the *n* officers observed in that moment, the features of their shared environment, and whether each officer fired a gun or the number of rounds each fired. Given the rarity of shootings, except for very large samples, any random sample of instances will likely contain no shooting incidents.

For the moment, assume that we can gather an enormous sample of this type. Then consider the data recorded at one specific time and place, observation i. All the information about the environmental factors and the officer characteristics in this i^{th} instance is encoded in the likelihood function, the probability of observing the shooting outcomes for given values of α and β. The log likelihood function contribution for a single moment i is

$$\log L_i(\alpha, \beta) = \log P(R_1 = r_1, \ldots, R_n = r_n | \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}, \alpha, \beta) \quad (2)$$

The likelihood of combining information from all sampled instances is the product of terms like (2), one for each sampled instance. Virtually all statistical approaches fall into this likelihood-based framework. If we could carry out a large-scale instance sampling operation recording all the relevant individual officer characteristics and environmental factors, then we could estimate α and β with a logistic regression model.

When faced with a likelihood involving nuisance parameters, such as the coefficients for the environmental factors, a commonly used strategy is to identify a sufficient statistic for the nuisance parameters, condition on that statistic, and base inference for the parameter of interest on a conditional likelihood (Kalbfleisch and Sprott 1970, 1973). That is, we need a statistic computable from the data, $S(R_1, \dots, R_n)$, such that when we condition on it, the likelihood factors into a component that involves β but not α as shown in (3).

$$\begin{aligned}\log L_i(\alpha, \beta) = \log\ & P(S(R_1, \dots, R_n)|\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}, \alpha, \beta) + \\ & \log P(R_1 = r_1, \dots, R_n = r_n|S(R_1, \dots, R_n), \mathbf{x}_1, \dots, \mathbf{x}_n, \beta) \\ = & \text{ collective group contribution} + \\ & \text{individual officer contribution}\end{aligned} \tag{3}$$

A key statistical result for logistic regression models and Poisson regression models is that setting $S(R_1, \dots, R_n) = R_1 + \cdots + R_n$, the number of shooters or the total rounds fired, satisfies this factorization property (McFadden 1973; Breslow et al. 1978). The number of shooters or the number of rounds fired is a signal of how much risk a particular incident's environment has. We do not need a good estimate of that risk, just a sufficient statistic for that risk.

This factorization shows the information about the model parameters is separable into different sources. The first term in (3) describes the scene as a whole, how many officers shot or how many total rounds did the officers collectively discharge. The second term in (3) describes the specific individual officers' actions on the scene, which specific officers shot or how many rounds each specific officer discharged. Information on the role of environmental factors is only available in the collective group term. Information on the role of officers' characteristics is split between the individual officers' actions (individual officer term) and their collective actions (collective contribution term).

The individual officer term in (3) is known as the *conditional* likelihood model.[ii] What makes conditional likelihood so valuable for studying police shootings is that knowledge of the environmental factors, $\mathbf{z}$, has no impact on the estimate of β. The conditional likelihood approach solves the confounding problem by providing a consistent estimate of β without ever needing to specify a model for $\mathbf{z}$ (Manski and Lerman 1977; Prentice and Pyke 1979).

For studying the decision to shoot versus not shoot, the individual officer contribution in (3) is

$$P(R_1 = r_1, \dots, R_n = r_n | R_1 + \cdots + R_n, \mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}, \alpha, \beta) \propto \frac{e^{r_1\beta'\mathbf{x}_1} \dots e^{r_n\beta'\mathbf{x}_n}}{\sum_{\rho_i\in\{0,1\},\sum\rho_i=\sum r_i} e^{\rho_1\beta'\mathbf{x}_1} \dots e^{\rho_n\beta'\mathbf{x}_n}} \tag{4}$$

The numerator only involves the officer characteristics and the actual observed configuration of shooters and nonshooters. The denominator sums over all possible configurations of shooters and nonshooters such that the counts match what was actually observed. Values of β that make (4) large are those that among all configurations of shooters and nonshooters make the observed configuration of shooters most likely.[iii]

For studying the number of rounds fired, the individual officer term is almost identical to (4), but in this case the denominator sums over all combinations of *n* non-negative integers that sum to the total number of rounds actually fired.

$$P(R_1 = r_1, \dots, R_n = r_n | R_1 + \cdots + R_n, \mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}, \alpha, \beta) \propto \frac{e^{r_1\beta'\mathbf{x}_1} \dots e^{r_n\beta'\mathbf{x}_n}}{\sum_{\sum\rho_i=\sum r_i} \frac{1}{\rho_1!\cdots\rho_n!} e^{\rho_1\beta'\mathbf{x}_1} \dots e^{\rho_n\beta'\mathbf{x}_n}} \tag{5}$$

Importantly, the individual officer contributions encoded in the expressions in (4) and (5) do not involve $\alpha'\mathbf{z}$ in any respect. There is no need to model or even record information on the environment to compute these terms.

The only moments in time and the only places in space that provide information about β through the individual officer term are those instances in which multiple officers are present for a police shooting. Even if we exhaustively documented millions of instances, for those instances in which no officers shot both (4) and (5) reduce to 1. Also for instances involving a single officer, both (4) and (5) reduce to 1. These points are critical for advancing research on individual-level analysis of police shootings. Those rare multi-officer shootings are special instances containing substantial information about the role of officer characteristics and, fortunately, extracting that information requires no measurement or modeling of environmental factors.

An illustration from a single shooting. Consider the data from a single shooting shown in Table A1. These two officers are nearly identical in all respects, except that the first officer was 24 years old when he joined the police and the second officer was 25 years old. At a particular moment and particular place, the first officer shot three rounds and the second officer shot four rounds.

Table A1

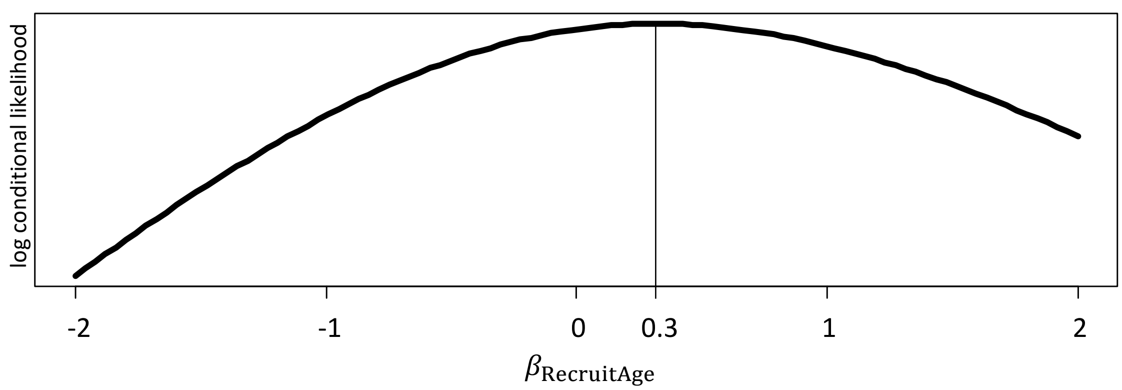
Records for an Example Shooting Involving Two Officers who Differ Only in age at Recruitment

OIS ID	*Rounds*	*Recruit age*	*Years on job*	*Sex*	*Race*	*Prior OIS #*	*Force complaints*	*Rank*	*Assignment*	*Gun type*	*Caliber*
2	3	24	4	Male	White	0	0	Officer	Special unit	Pistol	9 mm
2	4	25	4	Male	White	0	0	Officer	Special unit	Pistol	9 mm

While evidence from a single shooting is not persuasive about its causation, this hypothetical shooting suggests that one additional year of recruit age is associated with 1.3 times as many rounds fired (4/3 = 1.3). To formally use these data to learn about the relationship between these officers' characteristics and the number of rounds fired, we apply the conditional likelihood term from (5) to this shooting example. Since these officers are identical on almost all characteristics, all of the officer characteristics except age when recruited drop out of the conditional likelihood. Figure A1 plots the logarithm of the conditional log likelihood for a range of values for $\beta_{\text{RecruitAge}}$.[iv] This curve describes the relative likelihood of observing the younger officer shoot three rounds and the older officer shoot four rounds for different choices of $\beta_{\text{RecruitAge}}$. Consistent with the fact that the shooter who discharged one additional round was also one year older at the time of recruitment, this hypothetical shooting's likelihood is largest for positive values of $\beta_{\text{RecruitAge}}$. The value of $\beta_{\text{RecruitAge}}$ that maximizes the curve in Figure A1 is 0.3, suggesting that one additional year of age at recruitment increases the shooting rate exp(0.3) = 1.3 times. Indeed, the older officer shot 1.3 times the number of rounds that the younger officer discharged.

Figure A1

Contribution of the Example Shooting to the Log Conditional Likelihood Function



Since this example involves only one officer characteristic, tracing out the form of the conditional likelihood is simple. The curve peaks at a value that we can easily verify is consistent with the observed configuration of officers' ages at recruitment and rounds fired. To combine information from all shootings, we compute the product of (4), for a decision to shoot analysis, and the product of (5), for a number of rounds fired analysis, across all observed police shooting incidents. Though harder to visualize when combining data from numerous, more complex shootings, the relationships estimated from the full dataset using the conditional likelihood offer the best explanation for the observed police shooting patterns.

Weighing the information lost. Equation (3) explicitly shows that information about environmental factors is available only in data on officers' collective action (i.e., total number of shooters, total number of rounds fired) while information about officer characteristics is split between officers' individual actions (i.e. did the specific officer shoot, how many rounds did the specific officer fire) and the officers' collective action. The illustration in Figure A1 showed that we can extract information about the officer characteristics by examining only officers' individual actions.

While using only the individual officer contribution to the likelihood greatly simplifies the analysis of officer characteristics, it ignores any information about officer characteristics that we could extract from the collective group contribution to the likelihood. The problem with the first term in (6), which shows the complete log-likelihood for a dataset with *m* incidents with incident *i* involving n_i officers, as described in Kalbfleisch and Sprott (1973, pg. 314), "Information about [β] is tied up with information about the unknown [α] and it is difficult to quantify or ascertain what is lost."

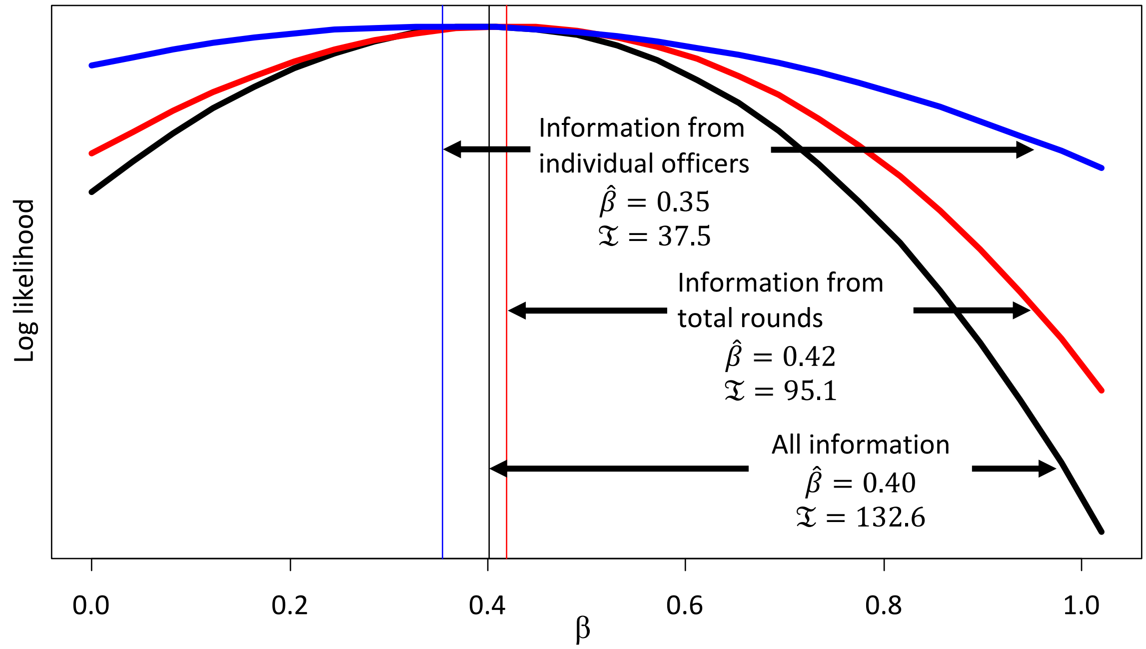
$$\log L(\alpha, \beta) = \sum_{i=1}^{m} \log P\left(R_{i1} + \cdots + R_{in_i} \middle| \mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}, \mathbf{z}_i, \alpha, \beta\right) + \sum_{i=1}^{m} \log P\left(R_{i1} = r_{i1}, \dots, R_{in_s} = r_{in_i} \middle| R_{i1} + \cdots + R_{in_i}, \mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}, \beta\right) \quad (6)$$

To explore this question, I simulated data on five million interactions, each involving a random number of police officers. For almost all shootings, I simulated such that shootings would occur in about 1 in 8,000 incidents, but for about 600 incidents the risk rises to 1 in 7 (to simulate the presence of an armed offender). For 10 percent of the officers I assigned them a risk factor that increases their risk of shooting by 50 percent, equivalent to $\beta = \log(1.5) = 0.4$. I then simulated the number of rounds fired by each officer in each incident. With this design, I can determine how much information is contained in each term in (6). The simulated dataset contained a total of five million incidents, 1,082 shooting incidents, 559 shooting incidents involving multiple officers, and 157 shooting incidents involving multiple officers that had variation in their officer-level risk factor. These latter 157 incidents are the only type of incident that have information through the conditional likelihood.

Figure A2 shows three log likelihood curves. The highest one is the log conditional likelihood using only information on individual officer actions, the second term in (6). The middle curve is the log likelihood for the collective group contribution, the first term in (6), using only data on the total number of rounds fired.[v] The lowest curve is the complete likelihood using all information, the sum of the other two curves.

Figure A2

Components of the Log Likelihood for Simulated Data



Notes: $\hat{\beta}$ is the maximum likelihood estimate from each source (true value was 0.4). The vertical lines mark each $\hat{\beta}$. $\mathfrak{I}$ is the observed Fisher information from each source

These curves express the amount of information about the officer characteristics contained in each data source. All three give similar values of $\hat{\beta}$, the strength of association between officer characteristics and shooting behavior, but the tightness of the curvature around the estimate indicates the strength of information. The "observed Fisher information" formally quantifies how much information is in the likelihood, computed as the negative second derivative of the log likelihood at its maximum. For this model, the observed Fisher information is additive. The total observed Fisher information in the full dataset equals the sum of the observed Fisher information from the individual officer contribution and the observed Fisher information from the collective group contribution.

In this simulation, the 157 interactions involving a police shooting with multiple officers with variation in their risk factors contain 28 percent of the total observed Fisher information in the full dataset of five million interactions. While representing 0.003 percent of the interactions, these moments contain a disproportionate amount of the information. To access the remaining 72 percent of information, we would need comprehensive data collection on the other 4,999,843 police interactions including an exhaustive inventory of the environmental factors in each instance. Perhaps someday a cost-effective, automated data collection process may be possible using data passively collected through video, audio, and sensors. In the meantime, we should forgo using that information since the cost of collecting the data and the risk of inadequately documenting and modeling the environmental factors are too great. Going back four years to gain a four-fold increase in the number of multi-officer shooting incidents would achieve the same level of precision as an exhaustive examination of 5 million interactions in one year, but without the risk of confounding.

Contagion, anti-contagion, and inference. The derivation of the conditional likelihood terms in (4) and (5) required a key independence assumption, that given the officer characteristics and the features of their shared environment, the officer outcomes are independent. The validity of this assumption depends on whether officer actions are contagious (an officer shoots because other officers are shooting) or anti-contagious (an officer does not shoot because another officer shot first). The presence of contagion in police shooting is debated (White and Klinger 2012).

Even in the presence of contagion or anti-contagion, use of the conditional likelihood will still result in estimates of $\hat{\beta}$ with the correct sign. The officers with the highest risk of shooting or the highest expected number of rounds discharged will still be most likely to shoot and likely fire the most rounds. The magnitude of the effect can be amplified with strong anti-contagion, if an officer with high risk factors for shooting abstains from shooting because another high-risk officer shoots first. The magnitude of the effect can be muted with strong contagion, if one officer shooting prompts other low risk officers to shoot. For the decision to shoot analysis, incidents in which all officers shoot provide no information to the conditional likelihood. Strong contagion could greatly reduce power by eliminating cases from the conditional likelihood.

Although violations of independence will not affect our conclusions on whether an officer characteristic increases or decreases the shoot risk or expected number of rounds fired, the traditional asymptotic standard errors will be incorrect. Permutation tests avoid relying on asymptotic assumptions. A permutation test randomly permutes the observed data in a manner that is consistent with the null hypothesis, computes new estimates based on the permuted data, and repeats this process numerous times (Good 1994). The collection of parameter estimates based on the permutated data provide a nonparametric reference distribution from which we can compute p-values. The null hypothesis is that the officer characteristics are unrelated to an officer's shooting decisions. Therefore, generating a permuted dataset entails randomly shuffling the shooting status or number of rounds among the officers within each shooting. Each police shooting incident still will have the same collection of values of $(r_1, r_2, \dots, r_n)$, but they will be randomly assigned to the n officers involved in the shooting. In these permuted datasets, the r_i are independent of the officer characteristics, but will retain any dependence from contagion or anti-contagion. The collection of estimates of β based on these permuted datasets yield a valid reference distribution, the distribution of $\hat{\beta}$ under the hypothesis of no relationship between officer characteristics and number of rounds fired. After generating several thousand permuted datasets and computing the associated parameter estimates $\hat{\beta}_{(1)}, \hat{\beta}_{(2)}$, ... the permutation p-value for the *k*th coefficient is the fraction of estimates from the permuted data that were less than or equal to the estimate computed on the original data.

Notes

---

[i] The Poisson regression model has the form $\log P(R = r|\mathbf{x}, \mathbf{z}) = r(\alpha'\mathbf{z} + \beta'\mathbf{x}) - \exp(\beta'\mathbf{x}) - \log r!$

[ii] Technically it is an *approximate* conditional likelihood for β since some information on β remains in the second likelihood term.

[iii] For example, if a shooting involved four officers where the first two officers shot and the others did not, then we would denote r = (1,1,0,0), which is what the numerator would evaluate. The denominator would compare that observed configuration to all of the six possible configurations of two shooting and two nonshooting officers, (1,1,0,0), (1,0,1,0), (1,0,0,1), (0,1,1,0), (0,1,0,1), (0,0,1,1). A good estimate of β would be one that makes the observed configuration (1,1,0,0) the most probable among all the configurations.

[iv] The log conditional likelihood for this example is $-\log \sum_{\rho_2=1}^{6} \frac{1}{(7-\rho_2)!\rho_2!} \exp\left((\rho_2 - 4)\beta_{\text{RecruitAge}}\right)$.

[v] A Poisson distribution with mean $\sum \exp(\alpha_0 + \alpha_1 z_i + \beta x_i)$.