2018 IEEE International Conference on Systems, Man, and Cybernetics

A penalized likelihood method for balancing accuracy and fairness in predictive policing

George Mohler, Rajeev Raje
Department of Computer and Information Science
Indiana University - Purdue University Indianapolis
{gmohler, rraje}@iupui.edu

Jeremy Carter
School of Public and Environmental Affairs
Indiana University - Purdue University Indianapolis
carterjg@iupui.edu

Matthew Valasik
Department of Sociology
Louisiana State University
mvalasik@lsu.edu

P Jeffrey Brantingham
Department of Anthropology
University of California Los Angeles
pjb@anthro.ucla.edu

***Abstract*—Racial bias of predictive policing algorithms has been the focus of recent research and, in the case of Hawkes processes, feedback loops are possible where biased arrests are amplified through self-excitation, leading to hotspot formation and further arrests of minority populations. In this article we develop a penalized likelihood approach for introducing demographic parity into point process models of crime. In particular, we add a penalty term to the likelihood function that encourages the amount of police patrol received by each of several demographic groups to be proportional to the representation of that group in the total population. We apply our model to historical crime incident data in Indianapolis and measure the fairness and accuracy of the two approaches across several crime categories. We show that fairness can be introduced into point process models of crime so that patrol levels proportionally match demographics, though at a cost of reduced accuracy of the algorithms.**

***Index Terms*—Predictive Policing, Fairness, Hawkes Process, Maximum Penalized Likelihood Estimation, Demographic Parity**

I. Introduction

Crime events cluster in space and time forming "hotspots" where 25-50% of crime may be captured in only a few percent of the land area of a city [1]–[3]. Spatial-temporal predictive policing algorithms [4], [5] attempt to capture the space-time dynamics associated with hotspot formation and direct police patrols in response, which can then lead to crime rate reductions [4]. The predictive policing data cycle is shown in Figure 1, where input may be generated from victim reports or police initiated arrests (for example). This data enters the police database and is then used by a predictive algorithm to inform police activity and patrol. That activity may then influence future suspects and victims in the areas the algorithm selects for patrol, as well as in those areas that do not receive police attention.

Racial bias of predictive policing algorithms has been the focus of several recent research articles [6]–[9]. Lum and Isaac [10] conduct a simulation study of predictive policing focused on drug arrests in Oakland, CA. They show that feedback loops are possible where biased arrests are amplified by a predictive algorithm leading to hotspot formation and further arrests of minority populations (see Figure 1). A similar concern is raised by Ferguson [11], who notes that arrests in a prediction area 'memorializes' that location as 'hot', which guarantees that it will show up again as a prediction area producing further arrests.

In this article we develop a penalized likelihood approach for introducing demographic parity into point process models of crime. Our goal is similar to the one developed in [12] where police patrols should match the "true" crime rate in an area, rather than crime rates that result from biased arrests. To achieve this goal, we add a penalty term to the likelihood function that encourages the amount of police patrol received by each of $1, ..., M$ demographic groups to be proportional to the representation of that group in the total population.

We note upfront that the introduction of a fairness penalty may lead to its own form of bias. Whereas biased arrests may result from interactions between police and suspects, bias may also adversely affect victims of crime [13]. Such bias may arise, for example, if an officer taking a report downgrades a burglary to a lesser crime as a result of the officer's implicit bias towards the victim's race/ethnicity. Fairness algorithms have the potential to operate in a similar manner. Consider the scenario for the predictive policing flow chart in Figure 1 characterized by a fairness algorithm being applied to crime reported by minority victims, leading to reduced hotspots in minority areas and less patrols, which then might lead to further crime rate increases in those areas.

In this paper we do not attempt, given a particular dataset and crime type, to distinguish between these forms of bias that may affect suspects or victims of crime differently. Instead, we explore the accuracy-fairness tradeoff when applying a fairness penalty to maximum likelihood estimation of point process models of crime. We show that patrol rates can be matched to population demographics, but at a cost to the accuracy of the algorithm and a lowering of the maximum crime rate reduction possible by predictive policing. In certain situations this result may be more fair, in others it may be less fair, and subjective

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00421

criteria (outside of the scope of this paper) may be needed to make that determination.

The outline of the paper is as follows. In Section II we review "self-exciting" or "Hawkes" point process models of crime and introduce our penalized likelihood approach to incorporating fairness. In Section III we apply our model, along with a "neutral" model with no fairness penalty, to Indianapolis crime incident data and measure the fairness and accuracy of the two approaches across several crime categories. In Section IV we discuss the implications of our findings and directions for future research.

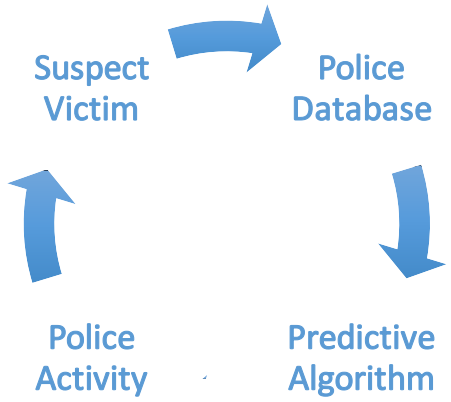

Fig. 1: Flow chart for predictive policing. Input may be generated from police initiated arrests or from victim reports. This data enters the police database and is then used by a predictive algorithm to inform police activity and patrol. That activity may then influence future suspects and victims in the areas the algorithm selects (and indirectly those the algorithm does not select).

II. Methods

A. Hawkes process model of crime

Given a data set of historical crime locations $\vec{x}_i$ and times t_i, we consider a modulated Hawkes process [14] defined in each grid cell g of a grid G with conditional intensity,

$$\lambda_g(t) = \mu_g + \sum_{\substack{t>t_i \\ \vec{x}_i \in g}} \theta\omega \exp(-\omega(t-t_i)). \quad (1)$$

Here the intensity or rate of crime $\lambda_g(t)$ is determined by a background rate, μ_g, that is stationary in time and a dynamic rate that increases after recent criminal activity, reflecting the tendency of certain types of crime to cluster is space and time [3], [15]. When viewed as a branching process, the parameter θ determines the expected number of events triggered by a past event and ω^{-1} is the expected waiting time between parent-daughter events.

To model the background rate, we introduce spatial covariates [14], [16], [17],

$$\mu_g = \exp(\vec{a} \cdot \vec{z}_g), \quad (2)$$

so that the background intensity in grid cell g is log-linear with coefficients $\vec{a}$ and spatial covariates $\vec{z}_g$ in each grid cell g. For the spatial covariates we use block group level demographic variables provided by the American Community Survey (ACS) along with the average historical number of events of the focus crime type modeled by $\lambda_g(t)$.

The demographic variables we use are the population estimates of each of $m = 1, ..., M$ racial groups surveyed in the ACS. The introduction of demographic spatial covariates serves two purposes. On the one hand, the spatial variation of $\vec{z}_g$ may help explain variations in the risk of crime. At the same time, the introduction of the coefficient vector $\vec{a}$ will allow us to control the amount of predictive policing across different racial groups and provide a mechanism for incorporating fairness into the algorithm.

The Hawkes process intensity $\lambda_g(t)$ is used in practice by ranking all grid cells g at a given time t and then directing police patrols to the top k grid cells, called hotspots [4]. We will denote the set of grid cells comprising the top k hotspots at time t as K_t.

B. Maximum likelihood estimation with a fairness-encouraging penalty

The parameters $\vec{a}$, ω, and θ of the Hawkes process can be estimated by maximizing the log-likelihood function,

$$L(\vec{a}, \omega, \theta) = \sum_{i=1}^{N} \log(\lambda_{g_i}(t_i)) - \sum_{g \in G} \int_0^T \lambda_g(t) dt, \quad (3)$$

where g_i is the grid cell of event i and the integral is taken over the time window of observation $[0, T]$. We will refer to Hawkes processes as "neutral" if estimated by maximizing Equation 3.

Next we will incorporate a fairness penalty into Equation 3. As defined above, we will let K_t denote the top k grid cells at time t, or the grid cells with the highest k values of $\lambda_g(t)$. We let z_g^m represent the population count of racial group m in grid cell g. Assuming that each of the grid cells in K_t receives the same amount of patrol, then the amount of patrol a particular racial group receives, p_m, per individual of that group per day is,

$$p_m = \frac{T^{-1} \sum_{t=1}^{T} \sum_{g \in K_t} z_g^m}{\sum_{g \in G} z_g^m}. \quad (4)$$

Here we are assuming that the hotspots are defined once per day and we therefore use a discrete sum over days. In the numerator, for each day t, we sum the number of individuals of group m residing in the hotspots K_t and in the denominator we count the total number of individuals in group m in the city.

We then define a measure of fairness, F, by comparing the patrol statistics p_m between pairs of groups,

$$F(\vec{a}, \omega, \theta) = \sum_{m > m'} (p_m - p_{m'})^2. \quad (5)$$

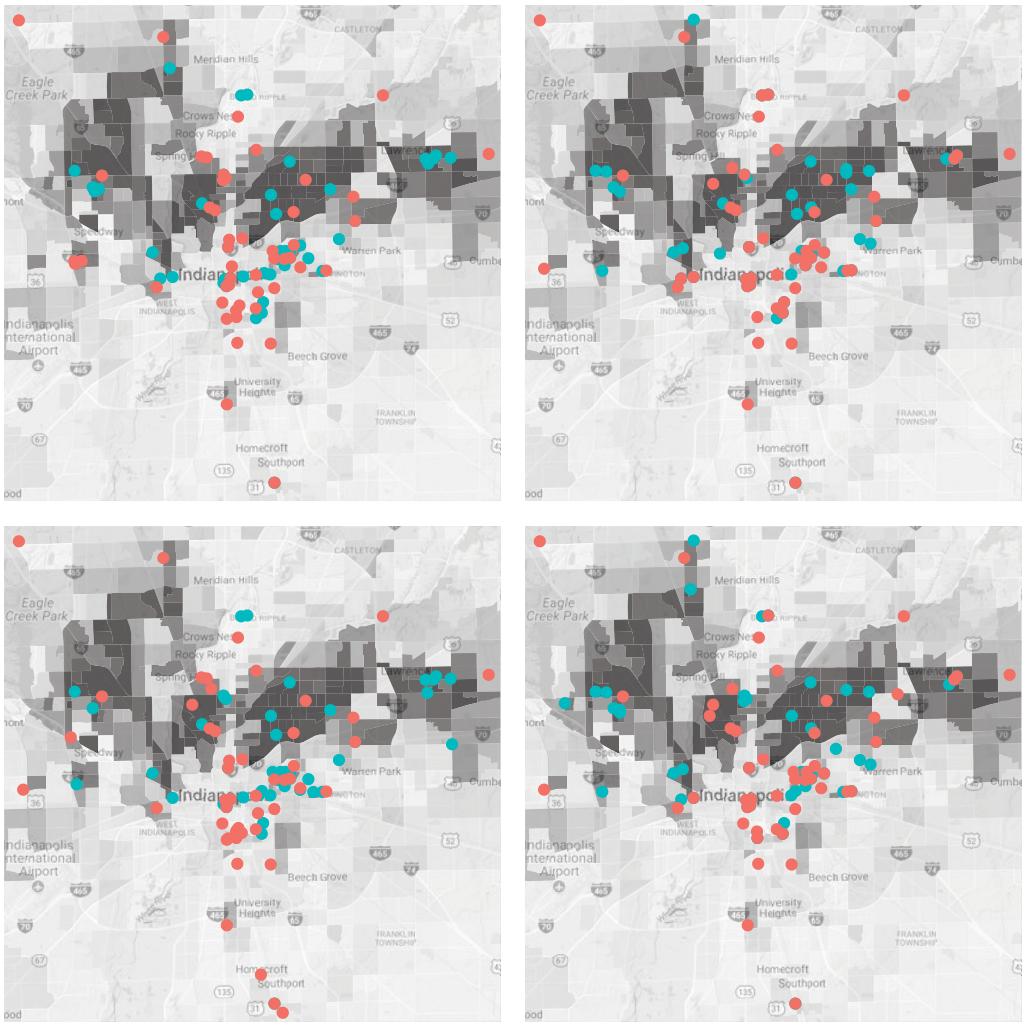

Fig. 2: Top 50 hotspot locations on a given day for the neutral (blue) and fair (red) Hawkes processes overlaid on a population density map of combined Black and Hispanic populations. From top left clockwise the maps show aggravated assault, robbery, motor vehicle theft and burglary hotspots.

When $F = 0$, then each group m receives the same amount of patrol per individual. Finally, we add F to the log-likelihood as a fairness penalty and maximize,

$$\sum_{i=1}^{N} \log(\lambda_{g_i}(t_i)) - \sum_{g \in G} \int_0^T \lambda_g(t) dt - \chi F, \tag{6}$$

with respect to $\vec{a}$, ω, and θ. We will refer to Hawkes processes as "fair" if estimated by maximizing Equation 6. By increasing the penalty parameter χ, one can control the balance between accuracy and fairness in the point process model.

We remark that the loss function determined by Equation 6 is non-differentiable due to the sort and threshold required for defining the top k hotspots K_t for each day t. However, a Nelder-Mead simplex method [18] can be used to find a local maximum and in the next section we show that such a method works well in practice.

III. Experimental Results

We estimate the neutral ($\chi = 0$) and fair ($\chi = 10^8$) Hawkes process models from crime incident data from the city of Indianapolis, Indiana for the years 2012 and 2013. The four crime types we focus on include aggravated assault, robbery, motor vehicle theft, and burglary. Each Hawkes process is defined on a grid with cell size 300m x 300m and incidents are each assigned to a grid cell. Crime incident locations are also assigned to a block group polygon corresponding to the American Community Survey for which we have white, Black, and Hispanic population density estimates from the survey. Every grid cell is then assigned population density estimates according to the mode taken over all incidents in each cell (to handle the edge cases of a grid cell overlapping two or more block group polygons). The reason for the choice of using grid cell based predictions rather than block group level predictions is that block groups are larger than typical predictive policing

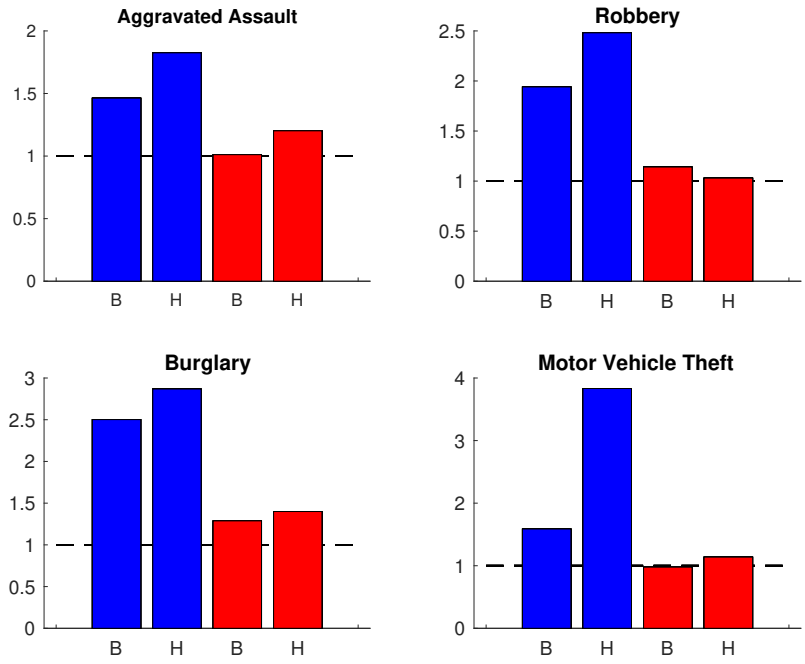

Fig. 3: Patrol levels per individual of Black and Hispanic populations relative to white populations (dashed line) for the neutral (blue) and fair (red) Hawkes processes.

hotspots [4].

We train the models on 2012 data using the Matlab Nelder-Mead simplex optimization function. For testing, we calculate the intensity in each grid cell for each day in 2013. We rank the grid cells according to the intensity each day and define hotspots as the top 50 grid cells. To measure accuracy, we use the Predictive Accuracy Index (PAI) [2], [19], which measures the percentage of incidents captured in the top k grid cells flagged for patrol. The PAI is area normalized (by the area of the k cells) so that a PAI of 1 corresponds to random predictions. We also measure the amount of patrol received per day per individual of Black and Hispanic populations relative to white populations in the top 50 hotspots.

In Figure 2, we plot the top 50 hotspot locations on a given day for the neutral (blue) and fair (red) Hawkes processes overlaid on a population density map of combined Black and Hispanic populations. The hotspots generated by the fair Hawkes process are shifted more towards the central "downtown" region of Indianapolis, as well as to northern and southern areas with predominantly white populations.

In Figure 3, we plot patrol levels per individual of Black and Hispanic populations relative to white populations (dashed line) for the neutral (blue) and fair (red) Hawkes processes. In the case of the neutral model, for all four crime types Hispanic populations in the hotspots receive from 200% to 400% the amount of patrol as white populations and Black populations receive 150% to 250% the amount of patrol compared to white populations. In the case of the fair Hawkes process, patrol levels per individual across all 3 groups are close to even, with the exception of aggravated assault and burglary, where Hispanic and Black population patrol levels are 20-35% higher (due to the fixed value for χ chosen apriori).

In Figure 4, we plot PAI values for the neutral (blue) and fair (red) Hawkes processes. Here we observe that PAI values are 30% to 250% higher in the case of the neutral model compared to the fair Hawkes process. Thus there is a significant cost in accuracy measured by PAI associated with the introduction of the fairness penalty into the likelihood. It still should be pointed out that the PAI values of the fair Hawkes process range from 6 to over 15, which are significantly higher than random and potentially more accurate than human analyst accuracy (demonstrated to have PAI values below 3 in [4]).

IV. Discussion

Research has demonstrated that racial bias exists in traditional enforcement strategies, from the racial profiling of vehicles [20] or pedestrian stops [21], [22], traffic tickets [23], drug enforcement and arrests [24], [25], and use of force [26]. The common thread connecting all of these studies is that racial bias is being observed in situations where police are being proactive and exercising the most discretion. Furthermore, when police engage in aggressive, zero-tolerance policing practices they tend to be focused in particular areas and are not widespread across a city.

On the other hand, impoverished urban areas with a high proportion of Black and Hispanic residents may experience under-policing [22], [25], [27]–[29]. Law enforcement may be under-resourced in these areas, where officers must travel

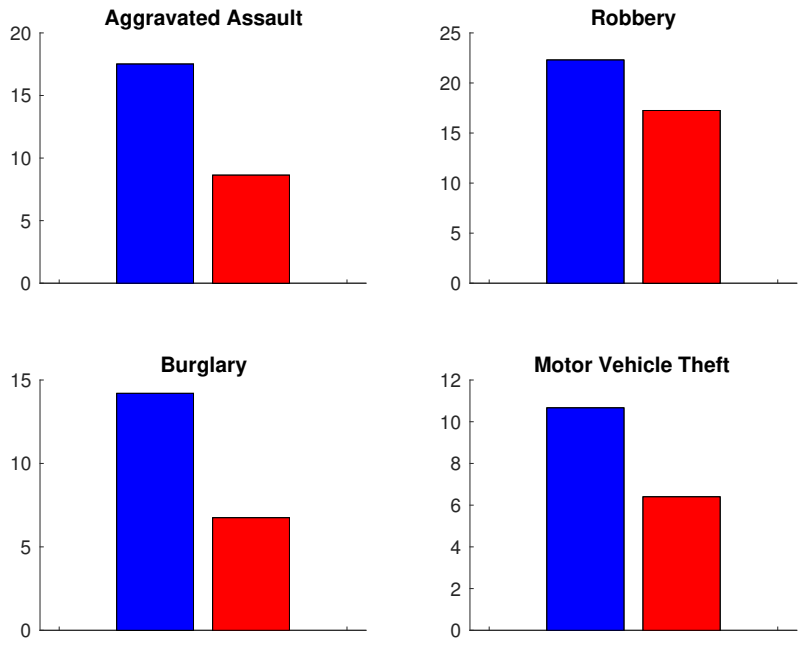

Fig. 4: Predictive accuracy index (PAI) for the neutral (blue) and fair (red) Hawkes processes.

from call to call and are rarely able to engage in proactive enforcement strategies [25], [29], [30]. As such, residents within these communities can become frustrated that police are not present enough in their communities to respond quickly to calls for service [27], [30]. Over time this perception that law enforcement is ineffective contributes to greater levels of cynicism among residents [31]–[33]. Because crime detection is reactive and dependent upon local citizens to report crime incidents [29], [34], these areas may then experience under-reporting that leads to fairness issues for predictive policing concerning victimization, separate from those predictive policing issues of bias existing around arrests.

The goal of this paper was to examine how utilizing a Hawkes process that incorporates fairness into the algorithm, where the amount of police patrols received in an area is proportional to the size of a particular demographic group's share of the population, is able to accurately predict future criminal incidents. This paper shows that through a penalized likelihood approach we are able to include fairness into the Hawkes process, but at a cost to the accuracy of the algorithm. Referring back to the two scenarios in the introduction, in the case when the elevated crime rate for minority populations is due to biased arrests, then the accuracy of the neutral Hawkes process may be viewed as an artificial consequence of bias. In this case the fair Hawkes process may be used to remove bias while still yielding a PAI value well above that of random patrol. There may be other situations where areas with minority populations have a higher rate of crime victimization. In this scenario the fair Hawkes process would divert police resources away from these areas, introducing its own form of bias, and the neutral Hawkes process is the preferred model. Distinguishing between these two scenarios depends on the context surrounding the city where predictive policing is deployed and the data that is input into the model. Making such a distinction is outside of the scope of this paper and should be the focus of future research.

V. Acknowledgements

This work was supported in part by NSF grants SCC-1737585, SES-1343123, ATD-1737996 and ATD-1737770. P.J.B. and G.M. are co-founders and serve on the board of PredPol, a predictive policing company.

References

[1] D. Weisburd, "The law of crime concentration and the criminology of place," *Criminology*, vol. 53, no. 2, pp. 133–157, 2015.
[2] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, no. 1, pp. 4–28, 2008.
[3] G. Mohler, M. Short, P. J. Brantingham, F. Schoenberg, and G. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.
[4] G. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham, "Randomized controlled field trials of predictive policing," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1399–1411, 2015.
[5] G. Ridgeway, "Policing in the era of big data," 2018.
[6] P. J. Brantingham, M. Valasik, and G. O. Mohler, "Does predictive policing lead to biased arrests? a randomized controlled trial," *Statistics and Public Policy*, 2018.
[7] S. Brayne, "Big data surveillance: The case of policing," *American Sociological Review*, vol. 82, no. 5, pp. 977–1008, 2017.
[8] A. G. Ferguson, "Policing predictive policing," *Washington University Law Review*, vol. 94, in press.

[9] B. J. Jefferson, "Predictable policing: Predictive crime mapping and geographies of policing and race," *Annals of the American Association of Geographers*, 2017.

[10] K. Lum and W. Isaac, "To predict and serve?" *Significance*, vol. 13, no. 5, pp. 14–19, 2016.

[11] A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: New York University Press, 2017.

[12] D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian, "Runaway feedback loops in predictive policing," *arXiv preprint arXiv:1706.09847*, 2017.

[13] P. Brantingham, "The logic of data bias and its impact on place- based predictive policing," *Ohio State Law Review*, 2018.

[14] G. Mohler, J. Carter, and R. Raje, "Improving social harm indices with a modulated hawkes process," *International Journal of Forecasting*, 2018.

[15] M. Short, M. D??Orsogna, P. Brantingham, and G. Tita, "Measuring and modeling repeat and near-repeat burglary effects," *Journal of Quantitative Criminology*, vol. 25, no. 3, pp. 325–339, 2009.

[16] R. Waagepetersen, "Estimating functions for inhomogeneous spatial point processes with incomplete covariate data," *Biometrika*, pp. 351–363, 2008.

[17] D. Cox, "The statistical analysis of dependencies in point processes," *Stochastic point processes*, vol. 55, p. 66, 1972.

[18] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[19] G. O. Mohler and M. Porter, "Rotational grid, pai-maximizing crime forecasts," *preprint*, 2017.

[20] F. R. Baumgartner, D. A. Epp, K. Shoub, and B. Love, "Targeting young men of color for search and arrest during traffic stops: evidence from north carolina, 2002-2013," *Politics, Groups, and Identities*, pp. 1–25, 2016.

[21] J. Fagan, A. A. Braga, R. K. Brunson, and A. Pattavina, "Stops and stares: Street stops, surveillance and race in the new policing," 2015.

[22] V. M. Rios, *Punished: Policing the lives of Black and Latino boys*. NYU Press, 2011.

[23] R. A. Dunn, "Measuring racial disparities in traffic ticketing within large urban jurisdictions," *Public Performance & Management Review*, vol. 32, no. 4, pp. 537–561, 2009.

[24] M. Lynch, M. Omori, A. Roussell, and M. Valasik, "Policing the progressive city: The racialized geography of drug law enforcement," *Theoretical Criminology*, vol. 17, no. 3, pp. 335–357, 2013.

[25] D. Simon and E. Burns, *The Corner: A Year in the Life of an Inner-City Neighborhood*. New York: Broadway Books, 1997.

[26] J. Legewie, "Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination," *American Journal of Sociology*, vol. 122, no. 2, pp. 379–424, 2016.

[27] E. Anderson, *Code of the Street: Decency Violence and the Moral Life of the Inner City*. New York: Norton and Company, 1999.

[28] J. Leovy, *Ghettoside: A True Story of Murder in America*. New York: Spiegel and Grau, 2015.

[29] P. Moskos, *Cop in the Hood: My Year Policing Baltimore's Eastern District*. Princeton: Princeton Univeristy Press, 2008.

[30] Z. Canepari, D. Cooper, and J. Dimmocl, "Flint town," Los Gatos, 2018.

[31] M. Desmond, A. V. Papachristos, and D. S. Kirk, "Police violence and citizen crime reporting in the black community," *American Sociological Review*, vol. 81, no. 5, pp. 857–876, 2016.

[32] D. S. Kirk and A. V. Papachristos, "Cultural mechanisms and the persistence of neighborhood violence," *American Journal of Sociology*, vol. 116, no. 4, pp. 1190–1233, 2011.

[33] L. A. Slocum, "The effect of prior police contact on victimization reporting: Results from the policepublic contact and national crime victimization surveys," *Journal of Quantitative Criminology*, 2018.

[34] D. J. Black, "Production of crime rates," *American Sociological Review*, vol. 35, no. 4, pp. 733–748, 1970.