Prescriptive Machine Learning for Public Policy: The Case of Immigration Enforcement

Dimitris Bertsimas and Mohammad M. Fazel-Zarandi
MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA
{*dbertsim, fazel*}*@mit.edu*

Abstract

We analyze machine learning-based decision-making in the context of immigration enforcement. Despite increasing investments in immigration enforcement in recent years, there is little evidence supporting that existing policies have a significant effect on reducing crime. In this paper, we develop a machine learning algorithm using federal government administrative data, comprised of over 900,000 observations, to predict the risk of criminal recidivism of noncitizens convicted of crimes in the United States. Our objective is to embed these predictions into a decision-making framework to design an immigration policy that prevents crime. Assessing the effectiveness of our algorithm's recommendations in reducing crime is complicated by the presence of unobserved confounders in the data-generating process. We address this issue by exploiting a quasi-experimental design that combines two key observations, namely the classification and allocation of inmates to prisons in the US and ICE officers' tendency to detain and deport immigrants closer to their local offices. These observations create a plausibly exogenous source of variation that allows for an unbiased evaluation of the algorithm's impact. After accounting for unobserved confounders, we show that the implementation of our algorithm reduces severe crimes by 26% and successfully decreases re-offending rates for inmates in both federal and state prison systems. Our results highlight the importance of considering both observed and unobserved factors when utilizing machine learning predictions to make policy decisions. This enhances the reliability of machine learning algorithms in addressing significant social and policy challenges.

1 Introduction

Over the last decade, machine learning algorithms have been successfully applied to a wide range of applications in science, business, and society. Historically, these algorithms focused on making predictions by discovering complex patterns in data. More recently, however, there is increasing attention to utilizing such technologies to make decisions previously reliant on human judgment. This paper develops and evaluates such an algorithm to improve decision-making in a highly controversial and evolving policy area: immigration enforcement.

Much of the debate in the United States on immigration enforcement centers on whether such efforts reduce crime. Despite the federal government's $380 billion spending on immigration enforcement over the past two decades, experts in academia and public policy have argued that many enforcement initiatives have been ineffective in decreasing crime in the US (Miles and Cox (2014); Nixon and Qiu (2018); American Immigration Council (2020); Bertsimas and Fazel-Zarandi (2020)).

In this study, we embed a machine learning algorithm into a decision-making framework to ameliorate a major federal immigration enforcement policy—the Secure Communities Program. This initiative seeks to lower the information cost of identifying noncitizens deportable under US immigration laws by ensuring that all individuals arrested by law enforcement agencies are screened for immigration violations. Under Secure Communities, millions of fingerprints annually submitted to the FBI for criminal background checks are forwarded to the Department of Homeland Security (DHS) to match against immigration records. If DHS identifies an individual in violation of immigration laws, it notifies the Immigration and Customs Enforcement agency (ICE) to analyze the case and decide whether to take custody of the person for removal proceedings. Due to limited resources, ICE uses a priority system to determine which cases to pursue. These priorities are intended to concentrate resources on the detainment and deportation of individuals whose removal promotes public safety and national security.

Despite clear guidelines on who should be a deportation priority, in Bertsimas and Fazel-Zarandi (2020) we show that ICE officers often target low priority immigrants while ignoring some high priority ones. These deviations from the guidelines could be due to inadequate expertise and flawed risk-resource trade-offs by officers. Community norms and personal values may also influence ICE's decisions. In theory, a machine learning algorithm could determine who should be

a priority for deportation. The algorithm can utilize personal characteristics and criminal histories to predict crime risks, which it can then use to recommend who should be detained and deported. This paper develops such an algorithm using a large dataset of individuals identified through Secure Communities. We show that our algorithm is capable of predicting the occurrence of severe crimes (AUC=0.786). This suggests that it can be utilized to complement Secure Communities—a reactive policy that targets individuals based on their prior crimes—with a proactive data-driven alternative, identifying and removing high-risk individuals before they commit further severe crimes.

While developing the machine learning algorithm is a major task on its own, the paper's primary innovation is how we evaluate whether the algorithm's recommendations improve upon ICE's decisions. Such an evaluation is challenging as the data used to train the algorithm is influenced by ICE's past choices, which can rely on information not recorded in the data. To control for unobserved confounders when testing our algorithm, we exploit a quasi-experimental design[1] that combines two key observations specific to our context: (i) the classification and allocation of inmates to prisons by federal and state authorities, and (ii) ICE officers' tendency to detain and deport inmates closer to their local offices. These observations create a plausibly exogenous source of variation in ICE's decisions that allows for an unbiased evaluation of the algorithm's recommendations. After accounting for unobserved confounders, our results show the potential impact of using the algorithm: holding ICE's resources at the current level, implementing our algorithm reduces severe crimes by, on average, 26%. We find that the algorithm decreases re-offending rates for inmates in both federal and state prison systems, and reduces crime across the different crime categories examined.

Taken together, our results highlight the potential impact of utilizing machine learning algorithms to improve decision-making in this public policy domain. We demonstrate that successful implementations of data-driven policies require an approach that carefully connects predictions to decisions by accounting for both observed and unobserved factors that may have influenced prior choices made by humans. Such an approach allows for a more robust evaluation of machine learning algorithms, enabling their more widespread adoption to solve real-world policy problems.

[1]These designs refer to conditions similar to randomized control trials, but differ in that assignments are done through mechanisms other than random assignments by researchers.

The paper is structured as follows. Section 2 summarizes the institutional details of our empirical context and reviews related literature. Section 3 describes the data sources used for the analysis. Section 4 presents our predictive algorithm and details the model training procedure. Section 5 transforms the algorithm's predictions to recommendations and presents the quasi-experimental design used to evaluate these recommendations. The last section concludes the paper.

2 Empirical Context

The United States Immigration and Nationality Act designates various criminal and civil offenses grounds for deportation. These offenses range from aggravated felony convictions to unlawful presence in the US. To facilitate the identification of immigrants (both documented and undocumented) charged or convicted of deportable offenses, DHS introduced the Secure Communities Program in 2008.[2] This program seeks to enhance immigration enforcement through better cooperation between federal immigration authorities and local and state law enforcement agencies.

When law enforcement agencies make an arrest, they submit fingerprints to the FBI for a criminal background check. Under Secure Communities, the federal government also forwards the fingerprints to DHS to match against immigration records. These records contain information on individuals violating US immigration laws, e.g., those who were previously deported or overstayed their visas, as well as lawful immigrants who have been convicted of a deportable crime. Upon finding a matching record, DHS notifies ICE to review the case and issue a "detainer"—a notice of intent to detain—if there is probable cause that the individual is deportable. The detainer requests that the other agency hold the person in custody for forty-eight hours beyond their scheduled release date to permit ICE to assume custody of the individual.[3] The decision to physically take custody of a detainer recipient—which we will refer to as the "detainment decision"— is made by the ICE field office from the arresting jurisdiction (see Figure 1 for the jurisdiction of the 24 ICE

[2]DHS rolled out Secure Communities on a county-by-county basis starting in October 2008. The program was formally activated nationwide in January 2013.

[3]As immigrants convicted of a crime must serve out their criminal sentence before being deported, a detainer could be issued any time between the initial arrest and the date when the immigrant is set to be released from jail/prison.

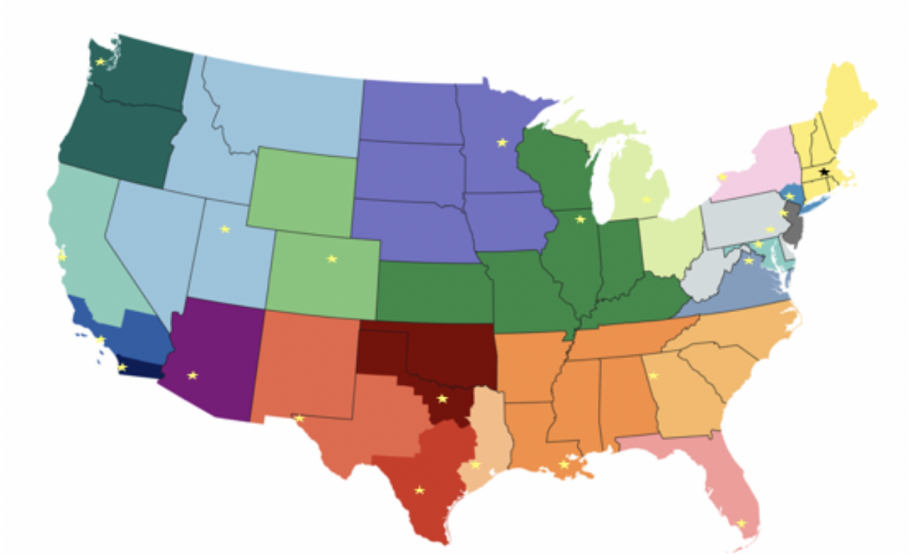

Figure 1: The 24 ICE Enforcement and Removal Operations field offices across the US. Stars mark office locations, and different colors represent the geographic jurisdiction of the offices. The offices are located at: Atlanta, Baltimore, Boston, Buffalo, Chicago, Dallas, Denver, Detroit, El Paso, Houston, Los Angeles, Miami, New Orleans, New York, Newark, Philadelphia, Phoenix, Salt Lake City, San Antonio, San Diego, San Francisco, Seattle, St Paul, and Washington DC.

offices across the United States). For individuals who are taken into custody by ICE, the officers must make a final decision on whether to remove them from the US—we will refer to this as the "deportation decision." Note that if ICE decides not to take custody of an individual, the detainer automatically lapses and the person is released from the criminal justice system upon completion of his/her criminal sentence.

ICE often does not assume custody of detainer recipients due to a myriad of logistical constraints, including the availability of transport vehicles and detention beds, ICE agents' ability to work overtime when transport times are long, and the availability of immigration judges. To better allocate resources, DHS has issued guidelines for ICE officers to utilize when making their detainment and deportation decisions. These guidelines classify detainer recipients into different priority levels; Level 1: those convicted of serious crimes (e.g., homicide, rape, and kidnapping); Level 2: non-aggravated felonies; Level 3: misdemeanors and lesser offenses (see Immigration and

Customs Enforcement (2013) for the complete list of crimes within each priority level).[4] While these priorities mandate ICE to target immigrants convicted of serious crimes, in Bertsimas and Fazel-Zarandi (2020) we show that ICE often deviates from DHS's guidelines. Indeed, we find that non-crime related features predict ICE's detainment decisions more accurately than the severity of the crimes.

As stated earlier, our main objective in this paper is to utilize machine learning to overhaul the Secure Communities Program by complementing it with a proactive data-driven algorithm. Not only does our algorithm shift a backward-looking (reactive) policy into a forward-looking (proactive) one, but it also is more immune to many of the biases and preferences that plague human decisions.

Related Literature

Prior research has primarily been concerned with assessing the causal relationship between immigration and crime (see, e.g., Butcher and Piehl (1998); Sampson (2008); Light and Miller (2018)), where the critical question is whether immigrants increase crime. A corollary question is whether immigration policy reduces crime. Pinotti (2017) exploits a regression discontinuity design to assess the causal effect of immigrant legalization on crime, showing that legalization reduces crime rates of legalized immigrants. Miles and Cox (2014) exploit the sequential rollout of Secure Communities to obtain differences-in-differences estimates of the program's impact on crime rates, finding that it has led to no meaningful reduction in crime.[5] A common thread among the above papers is that they all seek to measure a causal relationship. Our work deviates from this body of research as we focus on a more practical question of whether we can utilize a prediction algorithm to predict the crime risk of immigrants in the US to design an immigration policy that prevents crime. Note that while our focus is on predictions, we also draw insights from causal inference to evaluate our algorithm's recommendations in the presence of unobserved confounders.

Our work is also related to the literature on predicting criminality (see Berk (2012) and the

[4] To better align ICE's detainment practices with the prioritization objectives, DHS replaced Secure Communities with the Priority Enforcement Program in November 2014. Secure Communities, however, was reinstated in January 2017 under the executive order entitled "Enhancing Public Safety in the Interior of the United States".

[5] See Alsan and Yang (2018) and East et al. (2018) for other policy implications of Secure Communities.

references therein). A limitation of this literature is that it has mostly focused on forecasting recidivism, rarely comparing and connecting such forecasts to actual human decisions. There has recently been a small literature on linking predictive models to bail decisions. Jung et al. (2020) use a logistic regression model to predict whether defendants will fail to appear for trial. They then compare their predictions to judges' bail decisions under an unconfoundedness assumption and assess their model's sensitivity to violations of such an assumption.[6] Kleinberg et al. (2017) highlight the potential bias induced by ignoring unobserved confounders. They relax the unconfoundedness assumption by using the random assignment of cases to judges in New York City to analyze how judges' bail decisions compare to their model's predictions. In the context of policing, Mohler et al. (2015) develop an epidemic-type aftershock sequence model to estimate crime hotspot risk. Using randomized controlled trials, they show that their model outperforms a dedicated crime analyst in predicting crime hotspots. Our work differs from these studies in terms of context, methodology, and experimental design.

3 The Data

Our study is based on the universe of detainer orders issued across the United States, spanning from Secure Communities' inception in October 2008 to November 2015. The data, obtained via the Freedom of Information Act, include much of the information used by ICE officers when making their detainment and deportation decisions. We received the data in two separate datasets. The detainer dataset contains the recipients' demographics (age, gender, nationality, place of birth) and location information (the incarceration facility and county/city the detainer was sent to). This dataset is linked through depersonalized IDs to the criminal records dataset, which includes the recipients' criminal histories in the US.

For our analysis, we consider all detainers with complete demographic, criminal records, and location information. To arrive at the final data, we face two obstacles. First, the depersonalized IDs, which allow longitudinal linkages, are accurate within each fiscal year, but may not be fully reliable if numerous detainers were issued for a person serving a multi-year criminal sentence, or

[6] The unconfoundedness assumption (Rosenbaum and Rubin (1983)) entails that all confounding variables that jointly influence the treatment-control assignment and the outcome are contained in the set of *observed* covariates.

if the person was transferred between multiple incarceration facilities. To overcome this issue, we developed an algorithm that uses a combination of personal characteristics and criminal histories to find matching detainers. For each detainer, the algorithm checks if the receipt's personal characteristics and criminal history match any other information in the dataset. The algorithm is highly accurate in detecting duplicates as it leverages the rich information in the criminal records: for each detainer, we have the receipt's complete list of crimes in the US (there are over 450 different types of crime recorded in the data), the status of each of those crimes (charged, convicted, or dismissed), their charge and conviction dates (day, month, and year), as well as the criminal sentence lengths. The algorithm flagged 17.1% of the detainers as duplicates. We only keep the most recent record for these duplicate observations to avoid biasing our predictive algorithm with redundant training data.

A second challenge is that for detainer recipients who are not taken into custody by ICE and who have not been re-arrested for a crime after receiving a detainer, we must control for the requirement that they first serve out their criminal sentence before ICE makes its detainment decisions. Fortunately, we can calculate release dates by combining conviction dates with trial outcomes and sentence lengths. An obstacle in estimating the release dates is that 9.6% have missing sentences. We impute these missing values by developing a machine learning model that predicts criminal sentences based on factors that judges consider when making their sentencing decisions (see Bertsimas and Fazel-Zarandi (2020) for the details of the model). After estimating the projected release dates, 4.7% of detainer recipients are predicted to be in prison by the time the data was compiled, whom we exclude from the analysis.

Our final working sample is comprised of 904,896 detainers and 3,640,599 crimes. Overall, 69.8% of individuals receiving a detainer were physically taken into custody by ICE, and 63.1% of all detainer recipients were eventually deported from the US. Among individuals who were not deported from the US, 13.4% are re-arrested for a Level 1 crime after receiving a detainer. Table 1 presents descriptive statistics of the data. Throughout the analysis, we randomly split the data into a 60% training set, used to train our algorithm, and 40% test set, used to evaluate the algorithm.

Sample size	904,896
Detained	0.698
Deported	0.631
Re-arrested for Level 1 crime	0.134
Demographics	
Age	Mean:33 SD:9
Gender	
Male	0.946
Nationality	
Africa	0.014
Asia	0.028
Caribbean	0.046
Central and South America	0.031
El Salvador	0.047
Europe	0.012
Guatemala	0.057
Honduras	0.053
Mexico	0.707
North America and Oceania	0.005
Location	
State	
California	0.248
Texas	0.216
Arizona	0.059
Florida	0.049
Georgia	0.048
New York	0.041
Other	<0.03
Facility Type	
County	0.624
Federal	0.135
State	0.120
Local	0.093
Other	0.028
Crime	
Category	
General violence	0.244
Fatal violence	0.012
Sexual violence	0.049
Property crime	0.278
Drug crime	0.290
Weapon crime	0.093
Number of crimes ≥ 3	0.396
Number of crimes ≥ 5	0.189
Priority	
Level 1	0.355
Level 2	0.186
Level 3	0.319
Charged but not convicted	0.140

Table 1: Summary statistics of the data.

4 The Prediction Algorithm

In this section, we present our algorithm for predicting crime risks. We model the prediction task as a binary classification problem. The model takes as input n observations of the form $\{(x_i, y_i, d_i)\}_{i=1}^n$, where $x_i \in \mathbb{R}^d$ is the set of features corresponding to observation i's personal characteristics and criminal record, $y_i \in \{0, 1\}$ the outcome indicating whether the person was re-arrested for a Level 1 crime after receiving a detainer (for convenience, we refer to our outcome as crime), and $d_i \in \{0, 1\}$ corresponding to whether ICE assumed custody of the individual upon release from prison to initiate deportation. The objective is to discover the relationship between x_i and y_i, $y_i = f(x_i) + \varepsilon_i$, where ε_i is a random error term.

To estimate f, we train the model only on data from detainer recipients who are not detained and deported (i.e., observation with $d_i = 0$), as we do not observe whether a deported individual would have committed a Level 1 crime if they were not removed from the US. We use a machine learning technique to obtain our estimate of f. Machine learning algorithms are superior to traditional approaches, e.g., linear and logistic regressions, as they impose fewer assumptions about the functional form of f. They are also better suited to capture complex non-linearities, can handle many variables, and are capable of performing implicit variable selection (Hastie et al. (2009); Breiman (2001)).

Given the high stake nature of our predictions, a major consideration in selecting the machine learning algorithm to train the model is interpretability—the ability to understand the reasoning behind the model's predictions. We trained our model using the Optimal Classification Tree algorithm (Bertsimas and Dunn (2017, 2019)), a state-of-the-art algorithm that is highly accurate and interpretable. This algorithm partitions the data into a set of non-overlapping regions (referred to as leaves) based on the values of independent variables. It then assigns a prediction to all observations that fall within a region. The algorithm outperforms classical decision trees (Breiman et al. (1984)) by leveraging mixed-integer optimization to find the tree with the highest accuracy. The loss function we use to find the optimal tree is:

$$\min_{\mathbb{T}} \sum_{i \in train} [(y_i(1 - \hat{f}(x_i, \mathbb{T}))^2 + (1 - y_i)\hat{f}(x_i, \mathbb{T})^2)] + \alpha \cdot \Gamma(\mathbb{T}).$$

The summation measures the fit of tree $\mathbb{T}$ on the training set, where $\hat{f}(x_i, \mathbb{T})$ is the predicted

probability of crime for observation i. The second term prevents the model from overfitting to the training data by penalizing the tree's complexity, $\Gamma(\mathbb{T})$.

To build the model, we considered observed features that are predictive of crime. This included variables for whether the person had been convicted of the following crimes: property crime, public crime, general violence, sexual violence, fatal violence, drug crime, weapon crime, immigration offense, traffic offense, felony, or misdemeanor. To capture recidivism, we incorporated two variables: one for indicating if the person had at least three convictions and the other for whether the number of convictions was greater than or equal to five. As arrests do not necessarily lead to convictions, we also included an additional set of variables indicating if a person was arrested for one of the above crime categories. Additionally, we considered variables for the age at first arrest and also at first conviction. For demographics, we considered the person's gender, citizenship, country of birth, and age. Finally, the location features we incorporated into the model are the state and type of facility (county, federal, local, state, and others) where the person is incarcerated. We encoded all categorical features through binary variables $x_{ij} \in \{0, 1\}$, where $x_{ij} = 1$ if condition j holds for person i.

To assess the predictive performance of the trained model, we measured the out-of-sample area under the curve (AUC). This measure, which lies between 0.5 and 1, quantifies our model's ability to identify individuals who will be re-arrested for a Level 1 crime. The AUC of our model is 0.786, indicating that it can proactively detect serious crimes with high accuracy. To check the robustness of our results across various crime categories, we ran separate models with outcomes indicating whether an individual was re-arrested for one of the following crime types: fatal violence, general violence, sexual violence, and major drug offenses. These models had AUC values above 0.7, illustrating that our algorithm's predictive power cuts across crime categories.

5 From Predictions to Prescriptions

Our objective is not only to predict crime risks, but also to use the algorithm to make detainment and deportation recommendations. To connect the algorithm's predictions to decisions, we sort individuals based on their predicted risk scores. We then transform the scores into decisions $a_i : \hat{f}(x_i) \to \{0, 1\}$ (where $a_i = 1$ if the algorithm recommends deportation) by choosing a

threshold that takes into account ICE's operational constraints; i.e., the algorithm recommends deportation for all individuals with a crime score above the specified threshold.

Evaluating whether our algorithm's recommendations dominate ICE's decisions is challenging as it requires accurate estimation of counterfactual outcomes; computing the number of crimes prevented by the algorithm depends on the baseline rate of crimes that would have occurred in the absence of the current ICE/DHS policy. Assuming a fixed detainment capacity, $\sum_i a_i = \sum_i d_i$, the reduction in Level 1 crimes resulting from the algorithm's recommendations can be formally expressed by:

$$\underbrace{E[y_i|a_i=0] - E[y_i|d_i=0]}_{\text{Change in Crime}} \propto \underbrace{E[y_i|d_i=0, a_i=1]}_{\substack{\textbf{Observed} \\ \text{Reduction in Crime}}} - \underbrace{E[y_i|d_i=1, a_i=0]}_{\substack{\textbf{Unobserved} \\ \text{Increase in Crime}}}.$$

The above equation illustrates the difficulty in evaluating the recommendations: for cases where the detainer recipient is deported, we do not observe whether they would have been re-arrested for a Level 1 crime if they were instead not removed from the US. An approach often used to estimate these missing outcomes is covariate matching, in which deported individuals are matched to non-deported ones with similar observable characteristics. The missing outcomes are then set equal to the average crime rate of the matched individuals with known outcomes:

$$E[y_i|x_i, d_i=1, a_i=0] = E[y_i|x_i, d_i=0].$$

A potential limitation of this approach is that it relies on an unconfoundedness assumption, which implies that outcomes are missing at random (Kleinberg et al. (2017)). In our context, however, the outcomes could be missing in a non-random way, as the availability of outcome data depends on ICE's prior decisions, i.e., the algorithm was trained only on data from individuals who were not detained and deported. If ICE officers rely on other informative features that are not recorded in the data, the imputed outcomes could be misleading and could bias the results in support of the algorithm.

5.1 The Quasi-Experimental Design

To overcome the complication in evaluating our algorithm's recommendations, we exploit a quasi-experimental structure that combines two key observations specific to our setting. The first observation is related to how defendants are assigned to federal and state prisons after being convicted of a crime. Upon entering the prison system, inmates go through a classification process that assesses their risk, custody level, and treatment needs. The Federal Bureau of Prisons carries out this process for those convicted of federal crimes, while each state's Department of Corrections is responsible for individuals in state courts. During this process, inmates go through a series of evaluations and interviews. They are then assigned security classification scores based on the nature of their crime, criminal record, health, and behavioral attributes. These scores are used to assign each prisoner to a facility with an appropriate security level, e.g., low, medium, or high security. The classification scores are periodically reviewed and adjusted as more information about inmates' behavior becomes available during incarceration.

The second component of our quasi-experimental design is our discovery that the proximity between local ICE offices and incarceration facilities is a strong predictor of who ICE detains and deports. We observed this association by training a machine learning model that maps physical distances to ICE's detainment decisions. The AUC of this model is 0.701, suggesting a strong association between proximity and detainment. We find that this negative correlation is valid for the majority of states and ICE offices.

Combining the above two observations creates a mechanism that allows us to evaluate our algorithm in a meaningful manner. Specifically, the first observation suggests that conditional on prison characteristics (i.e., location, type, and security level), ICE officers draw from the same distribution of defendants when making their detainment and deportation decisions. The second observation implies variation in detainment rates across different prisons based on their proximity to ICE offices. It is this variation that we exploit below to control for unobserved confounders when testing our algorithm. Two pieces of evidence supporting the plausibility of the exogenous source of variation are that the prison classification process applies to all inmates in federal and state prison systems regardless of legal status, and the observation that most ICE field offices were located prior to the inception of Secure Communities.[7]

[7] Note that while inmates are not randomly assigned to prisons, nor are ICE offices randomly located, as an

5.2 Prescription Tests

To assess the algorithm's prescriptions, we restrict our analysis to inmates in federal prisons and those incarcerated in state prisons in California and Texas.[8] We excluded detainees in county and local jails because of the dynamic nature of their inmate population. While a federal or state prison typically holds inmates with similar characteristics, jails manage a much more diverse cohort of people as the majority of their population are those awaiting trial or sentencing for a broad range of offenses. By excluding county and local jails, we ensure that conditional on prison characteristics, average inmate characteristics are not systematically linked to ICE's decisions.[9][10]

For our experiments, we assign inmates to nine bins based on the type of their incarceration facility (federal prison, state prison in California, or state prison in Texas) and the security level of the facility (low, medium, or high). Each bin contains an average of 14988.89 inmates incarcerated in 12.88 prisons. As a robustness check, we test whether the distribution of observed characteristics and offense types are similar across prisons in each of the nine bins, and find no systematic differences in their inmate populations.

Figure 2 plots the detainment rate against the quartile of physical distance between prisons and ICE offices for different security levels. The figure shows a strong negative relationship: detainment rates decrease as proximity increases for all security levels. On average, ICE detains 9.91 percentage points fewer inmates from prisons in the farthest quartile than those in the third distance quartile. This difference increases to 14.78 and 25.48 percentage points compared to the second and first quartiles, respectively. In the following two tests, we utilize this variation in ICE's detainment rates together with the assumption that inmates in different prisons with the same

empirical matter, inmates in a federal/state prison with a specific security level located far from an ICE office are, on average, similar to inmates in another prison with the same security level but near an ICE office.

[8] We limit our analysis of state prisons to California and Texas as they account for more than half of all state prisoners in our dataset. Other states either have small sample sizes or lack sufficient heterogeneity in prison security levels or distances for an accurate and meaningful evaluation.

[9] We also exclude individuals in administrative facilities and those in US Marshals custody as these facilities hold inmates when they are in transit to prison or before and during court proceedings.

[10] Focusing on prisons rather than jails also alleviates concerns regarding selection issues due to the anticipation of cooperation by local authorities—even jurisdictions that adopted sanctuary-type policies continued their cooperation for inmates incarcerated in prisons.

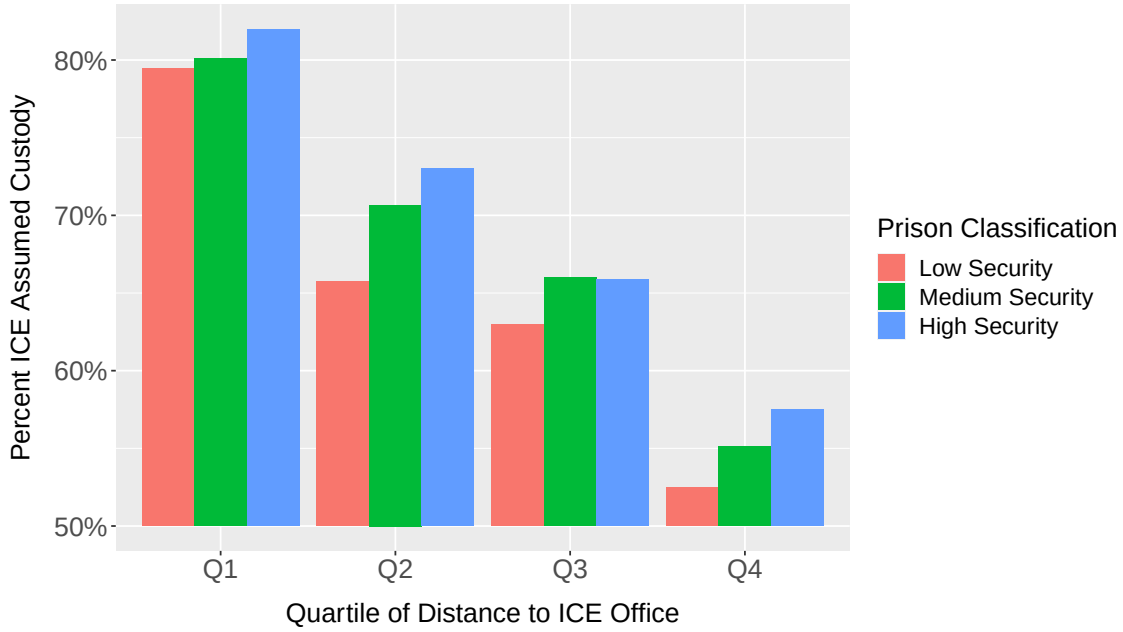

Figure 2: The relationship between detainment rates and proximity of prisons to ICE offices for different security levels. The figure illustrates the average detainment rate across federal prisons and state prisons in California and Texas for each security level and distance quartile. For federal prisons, the quartile cutoff distances are 98, 179, 276 miles; for states prisons, the cutoffs are 54, 105, and 151 miles.

security level are, on average, similar (as suggested by the inmate classification and allocation process described above) to evaluate the algorithm's prescriptions.

One-Sided Test: Expanding Resources

The first test for evaluating the algorithm's recommendations is set up to circumvent the problem of estimating unknown counterfactual outcomes. It does so by combining the quasi-experimental setup with the fact that we know the counterfactual outcomes for individuals who were not deported: they would not have committed another crime in the US if they were instead deported. As suggested by our experimental design, this test relies on the assumption that observations in data bin b are, on average, similar.

The test proceeds as follows. For each of the nine data bins, we focus on the prisons in the farthest distance quartile, Q_4. We use the algorithm to predict the crime risks of inmates in this quartile who were not taken into custody by ICE. These risk scores are subsequently used

to detain and deport additional inmates. We select additional individuals to detain until we attain the detainment rate of prisons in the third quartile, Q_3, and calculate the crime rate of the remaining, non-deported, individuals. We then continue deporting additional inmates from Q_4 up to the detainment rates of prisons in Q_2 and Q_1, calculating the resulting crime rates. If the algorithm's recommendations are to improve upon ICE's decisions, it must be that the crime rates produced by the above procedure are lower than the observed crime rates of inmates in Q_1, Q_2, and Q_3.

Figure 3 illustrates the results of the test. The red dots show the reduction in Level 1 crimes among inmates in the farthest distance quartile as a result of using the algorithm to expand detainment rates of prisons in Q_4 up to the rates of prisons in Q_3, Q_2, and Q_1, respectively. The blue dots show the percent difference in observed crimes for prisons in each of the first three distance quartiles compared to the crimes of Q_4; these differences are a consequence of the influence of proximity on ICE's detainment decisions. As ICE officers often deviate from DHS's priorities, we repeat the above process, but instead of using the algorithm, we use DHS's guidelines to detain additional inmates. The green dots show the reduction in crime due to using these guidelines. Finally, as a benchmark, the purple dots show the result of detaining additional inmates randomly.

As shown in Figure 3, the algorithm's recommendations dominate ICE's decisions across all distance quartiles. While both ICE's decisions and DHS's prioritization policy are inferior to the algorithm in reducing severe crimes, they outperform random selection. An interesting and surprising observation is that ICE officers have better performance compared to DHS's policy. This suggests that officers may be deviating from the priorities set forth by DHS as they are paying attention to factors not included in the (one-size-fits-all) guidelines. Note that while the figure shows the average reduction in crime across the nine bins, we observe a similar trend within each bin.

The above approach also allows us to calculate the reduction in Level 1 crimes if detainment resources are expanded. This is shown in Table 2. Using the algorithm to increase the detainment rate of prisons far away from ICE offices by 9.9 percentage points reduces crime by, on average, 16.7%. If the algorithm's detainment rates are increased by 14.8 and 25.5 percentage points, the number of severe crimes is reduced by 26.1% and 41.2%, respectively. These reductions in crime

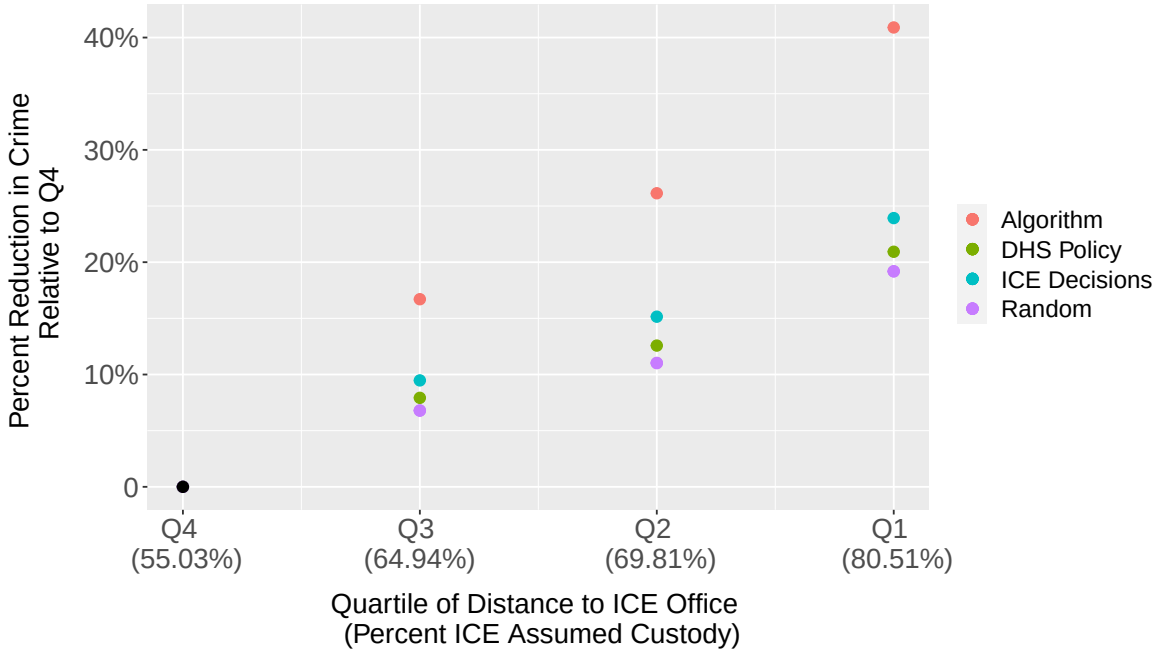

Figure 3: The reduction in Level 1 crimes among prisons in the farthest distance quartile as a result of increasing their detainment rates up to the rates of prisons in Q_3, Q_2, and Q_1. The red dots correspond to the algorithm's reduction in crime for prisons in Q_3, Q_2, and Q_1; the blue dots show the decrease in observed crimes as we move from Q_4 to the first three quartiles; the green dots correspond to the crime reduction as a result of using DHS's guidelines to determine additional deportees; the purple dots show the result of detaining additional inmates randomly.

would decrease to 9.5%, 15.1%, and 23.9% if ICE officers were making the additional detainments, and further lessened to 7.9%, 12.6%, and 20.8% if the DHS policy was fully implemented. These results clearly demonstrate the potential value of utilizing the algorithm's recommendations to make detainment and deportation decisions.

Two-Sided Test: Fixed Resources

This test seeks to evaluate the algorithm not just for individuals that ICE decided not to take into custody, but also for those who were detained and deported. As stated earlier, this is a challenging task as it requires estimating counterfactual outcomes. Fortunately, we can utilize the quasi-experimental design to overcome this obstacle. Unlike the one-sided test, where we expanded the detainment resources, this test evaluates the algorithm on the test set under the

	Level 1 Crime Reduction		
	Q_3	Q_2	Q_1
	Δdetainment=0.099	Δdetainment=0.148	Δdetainment=0.255
Algorithm	0.167	0.261	0.412
ICE Officers	0.095	0.151	0.239
DHS Policy	0.079	0.126	0.208

Table 2: Reductions in Level 1 crimes through expansion of ICE's detainment resources. The first row shows the reductions in crime due to using the algorithm to make additional detainments; the second row shows crime reductions if ICE officers expanded their resources; the last row shows the declines in crime if DHS's policy was used to increase deportations.

condition that resources are fixed.

To exploit the quasi-experimental design, we define $c_i \in \{0, 1\}$ as an indicator of whether inmate i is incarcerated in a prison close to an ICE office.[11] Moreover, let $L(\mathbb{T})$ denote the set of leaves (regions of the feature space) of our optimal tree algorithm. The test begins by sorting the leaves $l \in L(\mathbb{T})$ in descending order of predicted crime scores. It then increases inmates' detainment in prisons near an ICE office ($c_i = 1$) in each of the nine bins in the leaf with the highest score. For each extra deportation in a given bin, we decrease the deportation of inmates who belong to a similar bin and are incarcerated in prisons close to an ICE office ($c_i = 1$), but who fall in the leaf with the lowest crime risk. This ensures that the number and location of deportations are the same as those of ICE. Once the prisons near ICE offices in bins in the highest scoring leaf hit a 100% deportation rate, we move to the second-highest leaf and deport individuals from it in a similar manner. At the same time, for each bin in the leaf with the lowest score, we decrease the number of deportations until the deportation rate of prisons close to an ICE office ($c_i = 1$) is the same as the rate for prisons farther away ($c_i = 0$); we then move to the leaf with the second-lowest crime score. We continue this procedure until the two processes meet.

We quantify the change in Level 1 crime rates by measuring the difference between the number of reduced crimes resulting from increasing deportations in the leaves with higher predicted crime

[11] Assuming that the distance variable takes two values, near or far, is for ease of analysis. The analysis can be extended to allow for a multi-valued distance variable with some added complexity.

scores, denoted by Δ_{decrease}, and the number of increased crimes in the leaves with lower predicted scores, Δ_{increase}. The former quantity is easy to calculate: we observe the crimes of individuals the algorithm decides to deport (whom ICE had not deported) and know that their counterfactual crime rate would be zero. Estimating the latter quantity, the crime rate of those deported by ICE whom the algorithm decides not to deport, requires additional structure.

Let $I_{l,b} = \{i \in I_{test} | x_i \in \{l \cap b\}\}$ denote the observations in leaf l and bin b for which our procedure decreases the number of deportations, and $n_{l,b}$ the number of decreased deportations. We estimate the increase in the number of crime in bin b of leaf l by:

$$n_{l,b} \cdot \delta(l,b) = n_{l,b} \cdot \left| \frac{E[y_i | i \in I_{l,b}, c_i = 1] - E[y_i | i \in I_{l,b}, c_i = 0]}{E[d_i | i \in I_{l,b}, c_i = 1] - E[d_i | i \in I_{l,b}, c_i = 0]} \right|,$$

where the sample counterparts of the expected values (probabilities) are:

$$E[y_i | i \in I_{l,b}, c_i = c] = \frac{\sum_{i \in I_{l,b}, c_i = c} y_i}{|i \in I_{l,b}, c_i = c|},$$

$$E[d_i | i \in I_{l,b}, c_i = c] = \frac{\sum_{i \in I_{l,b}, c_i = c} d_i}{|i \in I_{l,b}, c_i = c|}.$$

The numerator of δ estimates the difference in the average crime rates between near and far prisons (within bin b of leaf l). The denominator scales this difference by the proportion of incarcerated individuals near an ICE office who are taken into custody and deported, but who would not have been detained if they were incarcerated far away.[12] We estimate the algorithm's increase in crime by:

$$\Delta_{\text{increase}} = \sum_{l \in \tilde{L}} \sum_{b \in \tilde{B}} n_{l,b} \cdot \delta(l,b),$$

where $\tilde{L}$ is the set of leaves and $\tilde{B}$ the set of bins within those leaves in which we decrease the number of detainees.

For the algorithm to be effective in reducing crime, it must be that $\Delta_{decrease} > \Delta_{increase}$.[13]

[12] Under the assumptions explicated by Angrist et al. (1996), δ is an estimator for the Local Average Treatment Effect.

[13] Note that this is a necessary indicator of the effectiveness of the algorithm, as we can only measure the counterfactuals at an aggregate level, and not at the individual level. It would also be sufficient under unconfoundedness.

	Federal Prisons	State Prisons
ICE/DHS Policy	0.1076	0.1537
Algorithm	0.0776	0.1173
Crime Reduction	27.82%	23.63%

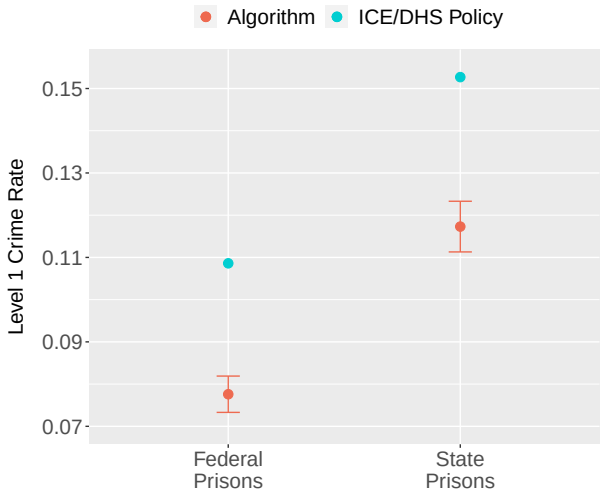

Figure 4: The first row of the table shows Level 1 crime rates under current ICE decisions/DHS policy for federal and state prisons; the second row shows the corresponding crime rates under the algorithm's decisions; the last row highlights the percentage reduction in crime as a result of using the algorithm. The figure depicts the table's results and shows 95% confidence intervals for the algorithm's crime rates (to calculate the asymptotic confidence intervals, we adapted the procedure described in Imbens and Rubin (2015)).

The change in crime rate can be calculated by:

$$\Delta_{\text{crime rate}} = \frac{\Delta_{\text{decrease}} - \Delta_{\text{increase}}}{\text{ICE Capacity}}.$$

To empirically evaluate our algorithm, we carry out the above procedure for federal and state prisons separately. We use the median distance to classify a prison as either near ($c_i = 1$) or far ($c_i = 0$) from an ICE office (the median distances for federal and state prisons are 179 and 105 miles, respectively).

The table and plot in Figure 4 report the results of our analysis. The results indicate large potential benefits from using the algorithm to make detainment and deportation decisions: the algorithm reduces the Level 1 crime rate for those incarcerated in federal prisons from 0.1076 to 0.0776, a 27.82 percent decline. We observe a similar pattern for those incarcerated in state prisons, where the algorithm reduces the crime rate by 23.63 percent, from 0.1537 to 0.1173. Note that these are conservative estimates of the algorithm's impact in reducing crime since

we only applied the algorithm to the proportion of the population for which we could control for unobserved confounders. The gains are potentially higher if the algorithm is applied to our study's entire population.[14]

6 Discussion and Implications

The results of the two prescription tests clearly demonstrate the potential societal impact of utilizing our framework to improve immigration enforcement decisions. Naturally, the 26% reduction in crime from employing the algorithm is specific to the population of immigrants in federal and state prisons, under the condition that distance creates a plausibly exogenous source of variation in detainment rates for this subpopulation. Our algorithm's impact may vary across other subpopulations (e.g., those incarcerated in local and county jails).

While crime prevention is one of the purposes of immigration enforcement, there are also other purposes, and there are likely trade-offs between those other purposes (e.g., removing those violating immigration laws) and crime prevention. To fully measure the impact of utilizing our algorithm, it is necessary to incorporate other goals of immigration enforcement into our framework. Nevertheless, in this paper, we focus on the crime prediction and minimization aspects of immigration enforcement because: (1) the federal government has power over those who have already been arrested by local law enforcement agencies; and (2) the greatest possible collateral benefit can be achieved through crime prevention. If the federal government is going to spend a budget of a certain amount of deportations, it would presumably prefer to achieve the most crime reduction possible with that expenditure.

Our framework could also have broader implications concerning the cooperation between federal immigration enforcement and local law enforcement agencies. In recent years, many jurisdictions across the US have adopted sanctuary policies, limiting their cooperation with immigration authorities, due to concerns ranging from opaque and inconsistent detainment decisions by ICE to infringements on civil liberties. Given the more transparent and objective nature of our machine learning algorithm, it has the potential to alleviate some of these concerns.

[14] While our analysis controls for both the number and location (near vs. far prisons) of deportations, ICE may face other logistical constraints. Thus, the 26% (average) reduction in crime can be interpreted as what would have happened if ICE could logistically follow our algorithm's prioritization perfectly.

Any use of machine learning to make detainment and deportation recommendations must be preceded by an understanding of the ethical and legal issues inherent in deploying algorithms to make such consequential decisions. The use of predictive analytics in the criminal justice system has become immensely controversial in recent years (Chouldechova (2017)). The most common view among the opponents is that algorithms are biased by disproportionally affecting one group over another. In our view, the issue is more complicated as there are multiple definitions of disproportionate, and no process can avoid all of them simultaneously. Moreover, even when algorithms are unjust, it is not clear that they are worse than the humans they are augmenting. Of course, before implementing our algorithm, it is necessary to make sure that it does not target specific subpopulations (e.g., based on race, nationality, and gender).[15]

7 Conclusion

This paper develops and evaluates a machine learning algorithm to improve decision-making in the controversial context of immigration enforcement. Our approach exploits a quasi-experimental design to deal with the empirical challenge that unobserved confounders can bias the comparison between the algorithm's recommendations and human decisions. The design combines two key observations specific to our context, namely the classification and allocation of inmates to prisons in the US and ICE officers' tendency to detain and deport inmates closer to their local offices, to allow for a meaningful evaluation of the algorithm's prescriptions. After accounting for unobserved confounders, we document that our algorithm reduces severe crimes by 25%, and is successful in decreasing re-offending rates for inmates in both federal and state prison systems. Combining our results with the one million observed Level 1 crimes in the dataset highlights the potential societal impact of using machine learning algorithms to improve national immigration policies.

More broadly, we demonstrate that successful implementations of data-driven policies entail going beyond machine learning predictions by carefully connecting such predictions to decisions. We show that a crucial component of the prediction-decision interaction is to account for both observed and unobserved factors that may have influenced the data used to train the algorithms.

[15] By comparing the descriptive statistics of individuals whom our algorithm switches from detained to not-detained with the complementary individuals who are switched from not-detained to detained, we do not observe any practically significant demographic differences.

By enhancing the credibility of machine learning-based decision-making, we hope to encourage more extensive use of algorithms to tackle significant social and policy challenges.

References

Marcella Alsan and Crystal Yang. Fear and the safety net: Evidence from secure communities. Technical report, National Bureau of Economic Research, 2018.

American Immigration Council. The Cost of Immigration Enforcement and Border Security (American Immigration Council, Washington, DC, 2020), 2020.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

Richard Berk. *Criminal justice forecasts of risk: A machine learning approach*. Springer Science & Business Media, 2012.

Dimitris Bertsimas and Jack Dunn. Optimal classification trees. *Machine Learning*, 106(7): 1039–1082, 2017.

Dimitris Bertsimas and Jack Dunn. *Machine Learning Under Modern Optimization Lens*. Dynamic Ideas, Belmont, MA, 2019.

Dimitris Bertsimas and Mohammad M. Fazel-Zarandi. Interpretable machine learning for analyzing national immigration policy. *MIT Working Paper*, 2020.

L Breiman, J Friedman, R Olshen, and C Stone. *Classification and Regression Trees*. Wadsworth, Monterey, CA, 1984.

Leo Breiman. Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2001.

Kristin F Butcher and Anne Morrison Piehl. Cross-city evidence on the relationship between immigration and crime. *Journal of Policy Analysis and Management*, 17(3):457–493, 1998.

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163, 2017.

Chloe East, Philip Luck, Hani Mansour, and Andrea Velasquez. The labor market effects of immigration enforcement. 2018.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Immigration and Customs Enforcement. ICE Criminal Offense Levels Business Rules, 2013.

Jongbin Jung, Connor Concannon, Ravi Shroff, Sharad Goel, and Daniel G Goldstein. Simple rules to guide expert classifications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2020.

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237–293, 2017.

Michael T Light and Ty Miller. Does undocumented immigration increase violent crime? *Criminology*, 56:370–401, 2018.

Thomas J Miles and Adam B Cox. Does immigration enforcement reduce crime? *The Journal of Law and Economics*, 57(4):937–973, 2014.

George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. Randomized controlled field trials of predictive policing. *Journal of the American statistical association*, 110(512):1399–1411, 2015.

R Nixon and L Qiu. What is ICE and why do critics want to abolish it? *The New York Times*, 3 July 2018. https://www.nytimes.com/2018/07/03/us/politics/fact-check-ice-immigration-abolish.html., 2018.

Paolo Pinotti. Clicking on heaven's door: The effect of immigrant legalization on crime. *American Economic Review*, 107(1):138–168, 2017.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Robert J Sampson. Rethinking crime and immigration. *Contexts*, 7(1):28–33, 2008.