Measuring the Option Value of Change: Theory and an Application to Operation Ceasefire

Sylvain Chassang* Meichen Chen Michal Kolesár

Princeton University Yale University Princeton University

March 23, 2026

Abstract

When treatment effects are heterogeneous across units but auto-correlated over time, there is value to dynamic treatment rules that assign treatment status on the basis of past treatment outcomes. Treatment has an option value. The challenge is that correlation in *estimated* treatment effects may be driven by correlation in errors rather than correlation in treatment effects. This paper shows how to estimate the value of dynamic assignment rules in observational data, and applies the methodology to Operation Ceasefire, a widely adopted homicide reduction program. Using a newly constructed data-set of adoption events across the US, we argue that programs inspired by Operation Ceasefire generate statistically insignificant reductions in homicide rates on average, but that their effectiveness is significantly improved by dynamic adoption. A naïve approach would considerably overestimate these benefits.

KEYWORDS: option value, external validity, extrapolation, experimentation, adaptive treatment, operation ceasefire

*Contact: chassang@princeton.edu. We are especially grateful to Jessica Shung for research support in the early stages of the project. We are indebted to Ilyana Kuziemko, Stephen Redding, and seminar participants at Princeton University for feedback.

1 Introduction

This paper is motivated by a concrete program evaluation challenge associated with Operation Ceasefire and its roll-out under various forms to a large number of cities in the US over the last 30 years (Piehl et al., 2000, Kennedy et al., 2001, Kennedy, 2011). Operation Ceasefire was initiated in Boston following a diagnostic that very high homicide rates reduced homicide resolution rates, creating a climate of impunity. It involved multiple steps requiring coordination across stakeholders with different objectives and decision processes: community leaders helped organize call-in meetings in which members of different gangs were brought together; the police demonstrated its ability to tie homicides to gangs, and committed to devote a disproportionate amount of resources to gangs involved in the next few homicides; the district attorney's (DA) office committed to pursue cases brought by the police even if not directly related to homicides.[1] From the perspective of crime reduction, Piehl et al. (2000) shows that the Boston version of the treatment was a success. Operation Ceasefire is also methodologically interesting, as an illustration of a divide and conquer principle important in mechanism design: treating similar behavior asymmetrically can help unravel unattractive equilibria (Abreu and Matsushima, 1992, Winter, 2004, Halac et al., 2019, 2020, Kapon et al., 2024).

Following its success in Boston, versions of Operation Ceasefire were rolled out to a large number of cities (according to our count, 122 between 1996 and 2018). However, as Kennedy (2011) highlights in his detailed account of Operation Ceasefire, correctly implementing the program requires extensive coordination between stakeholders, and one unmotivated or ineffective partner can cause the program to fail, thereby wasting valuable resources. This has several implications: first, treatment effects are likely heterogeneous across implementation events; second, treatment effects are likely correlated overtime, since early successes plausibly encourage commitment to the program; third, outcomes from flawed implementations

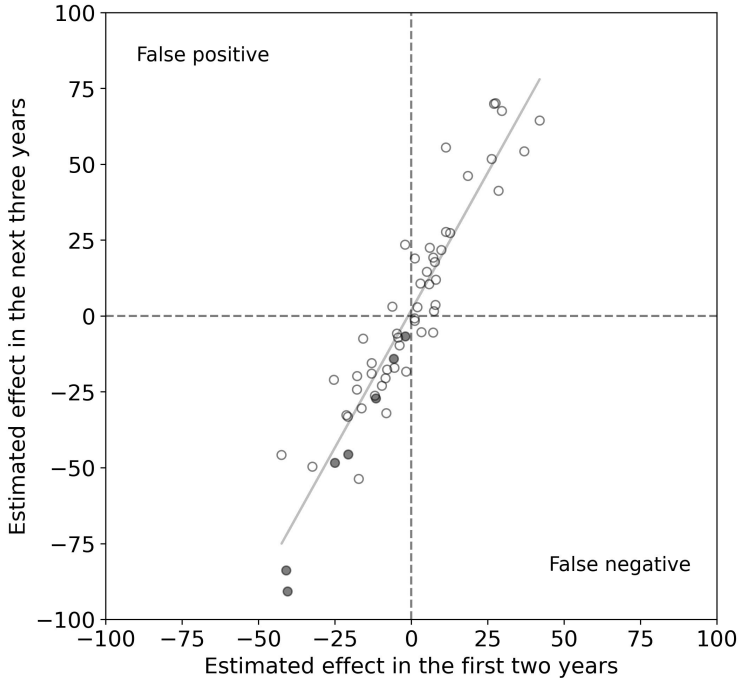
[1] In one case, the DA's office brought charges against a convicted felon tied to a homicide for unlawful possession of a single round of ammunition.

do not inform outcomes from successful implementations.

This presents a challenge for analysts seeking to inform policy-makers on the potential returns to implementation. On the one hand, treatment effects for a curated group of high-quality implementations may be more informative of the returns that motivated policy makers could expect. Correspondingly, the National Network for Safe Communities (NNSC), founded to help disseminate findings as well as support interested communities, reports experiences from a selected group of cities via its website: Chicago, Indianapolis, Stockton, Lowell, Boston, Cincinnati, and New Orleans.[2] On the other hand, as Al-Ubaydli et al. (2017) and Muralidharan and Niehaus (2017) highlight, implementation flaws are a systematic feature of scale-ups, and policy makers are likely unable to determine ahead of time whether or not their implementations will be flawed or successful. In addition, the process of rolling out a program at scale may require compromises and adjustments that change the nature of treatment even if it remains closely related to an original ideal. If that is the case, selecting sites to learn from may bias treatment effect estimates relevant to the policy-maker.

This paper offers a principled and broadly applicable strategy to resolve this tension. Figure 1 plots estimates of unit-specific treatment effects in the first 2 years (X-axis), and next three years (Y-axis) of treatment. As anticipated, there is significant heterogeneity in treatment effect estimates, and cities reported on the NNSC website are selected from the subset of cities for whom treatment effect estimates are negative (i.e. homicide rates decreased). However, they are broadly representative of cities in which the treatment seems to be effective, and treatment effect estimates in the first 2 years after implementation are highly correlated to treatment effect estimates in the next 3 years. From a policy perspective, this suggests that even if success cannot be predicted before the program is implemented, early program outcomes can be used to selectively drop treatment from units where it seems to be doing more harm than good. This permits some selection, but no more than is feasible

[2] `https://nnscommunities.org/impact/impact/`, last accessed January 2026. Braga and Weisburd (2012) and Braga et al. (2019) offer more systematic meta-analyses, but still focus on a restricted set of cities.



Notes: Effects correspond to cumulated reductions in the homicide rate per 100,000 inhabitants. Each solid circle represents a program listed on the NNSC's official website. Hollow circles represent the other programs in our final estimation sample. The line is the best linear fit from OLS.

Figure 1: Individual treatment effects on the intentional homicide rate

given the available information. Unfortunately, this use of treatment effect estimates to target future assignment poses an econometric challenge: if treatment effects are measured with errors, and errors are auto-correlated over time, the auto-correlation in estimated effects need not correspond to auto-correlation in true effects. This paper shows how to deal with this issue.

We study the problem of estimating the impact of dynamic treatment allocation strategies from both experimental and observational data in which there is variation in treatment status across units, but no variation in treatment status within units across time. We work under the potential outcome framework, and allow for treatment effects that are heterogeneous across units and vary over time. We also allow for auto-correlation in errors. A toy example illustrates how naïve plug-in estimators, in which true treatment effects are replaced by

treatment effect estimates that are unbiased within period, are in fact biased in the presence of auto-correlated errors. This suggests a simple powered placebo test that treatment has some impact, based on replicating the test statistic of interest using a randomly selected subset of control units as placebo treatment units.

The paper's first main contribution is to show how to estimate the value of dynamic assignment. We show that a more precise understanding of the time-series process for outcomes can be exploited to correct the bias associated with auto-correlated errors. Our main assumption limits the channels through which past treatment status affects future outcomes, allowing us to extrapolate outcomes for out-of-sample assignment histories in which a unit receives and then drops treatment. It is satisfied in models in which the underlying state is observed and Markov conditional of treatment, as well as models in which the underlying state is unobserved, Markov, and independent of the treatment history. This assumption fails in environments where treatment creates long-term dependencies, for instance, if dropping treatment after receiving it causes harm, as appears to be the case for SRRI anti-depressants in the medical context (Haddad, 1998).

The paper's second main contribution is a reevaluation of Operation Ceasefire and related programs. To this end, we construct a data-set of attempts to replicate key aspects of Operation Ceasefire across the US, and apply our methodology under two data generating processes: a latent state model in which treatment has no long-term effects once it is dropped, and an auto-regressive model in which treatment can have a long term impact even after it is dropped. We find that on average, programs inspired by Operation Ceasefire generate a 0.76 statistically insignificant reduction in homicides rates (on average treatment cities experience 15.74 homicides per 100,000 inhabitants). A dynamic adoption strategy, under which treatment is dropped if estimated treatment effects in the first 2 years are negative, induces a reduction in homicides rates between 1.82 (insignificant) and 2.43 (significant at the 5% level) per treatment year, depending on the model. Under both models, dynamic selection significantly improves treatment effects. Finally, naïve estimators of the value of

dynamic treatment are severely biased, with predicted homicide rate reductions between 3.02 and 4.49 depending on the model.

This paper belongs to the literature on structural treatment effects that leverages a deeper understanding of heterogeneous treatment effects to evaluate counterfactual policies. A large and influential literature identifies heterogeneous treatment effects as an important input to policy design. Heckman and Vytlacil (2005, 2007), Chassang et al. (2012), Mogstad et al. (2018) focus on estimating heterogeneous treatment effects based on underlying participant preferences. This permits counterfactuals under which incentives for treatment adoption vary. Imai and Ratkovic (2013), Wager and Athey (2018), Athey and Wager (2021) develop more general methods to identify relevant heterogeneity that can be used for targeting treatment based on rich observable characteristics. We show that even if ex ante covariates do not predict treatment effects, early outcomes can potentially be used to target treatment continuation.

The point that auto-correlated treatment effects create option value for experimentation is well known. It is the fundamental assumption in the theoretical bandit literature, whether it takes a Bayesian (Gittins, 1979), or frequentist approach (Lai and Robbins, 1985, Auer et al., 2002). Dynamic treatment assignment (often referred to as adaptive treatment strategies) has received significant interest from the clinical trial design literature, leading to the development of sequentially random treatment assignments (Murphy, 2005, Lei et al., 2012), and associated inference methods (Chakraborty and Moodie, 2013, Chakraborty and Murphy, 2014). The evaluation of dynamic treatment strategies has received some interest from economists. In contexts where short term exposure to a costly treatment may lead to long lasting behavioral change, identifying cost-effective exposure strategies is particularly valuable. Dupas (2014) highlights the value of free short-run exposure to new health products (antimalarial bed nets) as a driver of long-term product adoption. Ida et al. (2024) studies the use of dynamic subsidies to improve the adoption of variable price electricity contracts in Japan. This paper deals with the challenge of estimating the value of dynamic assignment

rules using static assignment data.

The paper also contributes to the growing literature on external validity. Empirical evidence (Cook et al., 2008, Chetty et al., 2016, Dehejia et al., 2021) shows that external validity is poorly addressed by reweighting heterogeneous treatment effects, even with rich covariates, and the point has been well recognized in the econometric literature (Crump et al., 2009, Athey and Imbens, 2017, Mogstad et al., 2018). Closely related to the motivation of this paper, Rosenzweig and Udry (2020) study the intertemporal persistence of treatment effects within the same units across a variety of high profile policy experiments. They show that treatment effects are imperfectly persistent, but not entirely independent either. Also closely related, Athey et al. (2025) study the problem of identifying short-term surrogate end-points for long-term effects, allowing for faster treatment improvement cycles. In the current paper short-term treatment outcomes are a natural surrogate end-point.

Finally the paper contributes to the evaluation of focused deterrence strategies associated with Operation Ceasefire. Piehl et al. (2000), Kennedy et al. (2001), and Braga et al. (2014) all focus on the Boston experience, while Braga and Weisburd (2012) provide an analysis of outcomes from experiments in selected cities. This paper seeks to include a much broader set of cities that attempted to implement imperfect variants of Operation Ceasefire, and evaluate whether selecting implementation sites on the basis of perceived successful implementation gives a correct representation of the value of experimenting with the program. We find that there are valid grounds for selection, but that the benefits from dynamic adoption are reduced once bias due to auto-correlated estimation errors is corrected.

The paper is organized as follows. Section 2 provides a toy example clarifying the value and challengers associated with the evaluation of dynamic treatment assignment using observational data. Section 3 provides a general framework under which it is possible to estimate the value of dynamic treatment assignments, and establishes that it includes leading data-generating processes of interest. Section 4 describes Operation Ceasefire and our data construction strategy. Section 5 reports our findings. Appendix A reports detailed sample

construction steps. Proofs are contained in Appendix B unless mentioned otherwise.

2 A Motivating Example

This section seeks to clarify the upside, as well as potential challenges, of evaluating the option value of treatments. Anticipating on the application to Operation Ceasefire (Piehl et al., 2000, Kennedy et al., 2001, Braga et al., 2014) developed in Sections 4 and 5, consider the evaluation of a crime reduction intervention taking place across different cities over multiple periods. The effect of treatment on crime is heterogeneous across cities. It depends finely on local conditions, the level of trust and coordination between the police and the District Attorney's office, variations on the initial program inevitably introduced during implementation. Treatment effects cannot be predicted using available covariates but treatment effects in the same location are correlated over time: if the treatment is effective in the first period, it is likely to be effective in the second, and vice versa. This possibility generates an option value for experimenting with treatment: locations that are unsuccessful in the first period can choose not to reconduct an underperforming program in the future.

This section uses a stylized model to articulate the following points:

(i) a treatment may have an average impact on welfare that is zero (or negative) but have positive option value;

(ii) Naïve estimators of option value that plug in estimated treatment effects in each period in an option value expression are biased whenever the outcome process exhibits auto-correlated errors.

(iii) Unbiased option value estimates can be recovered by running a richer dynamic experiment in which some treated units are dropped from treatment in later periods.

(iv) If treatment status is persistent in the data, it is possible to recover option value estimates if we are willing to assume that the process for outcomes is

independent of past treatment status conditional on current treatment status, and past outcomes.

Setup. There are two periods $t \in \{1, 2\}$ and a continuum of experimental units indexed by $i \in [0, 1]$.[3] Let $D_{i,t} \in \{0, 1\}$ denote the assignment of unit i to treatment ($D_{i,t} = 1$) or control ($D_{i,t} = 0$) in period t. Outcomes $Y_{i,t}$ take the following form:

$$Y_{i,1}(D_{i,1}) = a_1 + b_i D_{i,1} + \varepsilon_{i,1} \tag{1}$$

$$Y_{i,2}(D_{i,2}) = a_2 + b_i D_{i,2} + \rho\varepsilon_{i,1} + (1 - \rho)\varepsilon_{i,2} \tag{2}$$

with

$$\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad a_t \sim \mathcal{N}(\mu_a, \sigma_a^2), \qquad b_i \sim N(\mu_b, \sigma_b^2)$$

independently distributed across time t, and units i. Treatment effects b_i are heterogeneous across units, but identical across periods. Error terms $\varepsilon_{i,1}$ and $\rho\varepsilon_{i,1} + (1 - \rho)\varepsilon_{i,2}$ are auto-correlated over time via parameter $\rho > 0$.

Treatment regimes. Under a persistent treatment regime, a unit i is given the same treatment status in both periods 1 and 2:

$$D_{i,1} = D_{i,2} = D_i \in \Delta(\{0, 1\}). \tag{3}$$

A large share of experimental evidence in social sciences, whether it originates from a randomized controlled trial, or a natural experiment, involves persistent treatment assignment. When treatment status D_i is assigned independently across units, such an experiment

[3] Throughout the main text we focus on identification and consider the case of arbitrarily large samples. We report bootstrapped standard errors in Section 5.

allows us to identify the average treatment effect ATE of persistent assignment:

$$ATE = \mathbb{E}[Y_{i,1} + Y_{i,2}|D_{i,1} = D_{i,2} = 1] - \mathbb{E}[Y_{i,1} + Y_{i,2}|D_{i,1} = D_{i,2} = 0].$$

When treatment effects are heterogeneous across units and correlated across time, there is value in using adaptive policies conditioning period 2 treatment $D_{i,2}$ on period 1 outcomes. Let

$$\widehat{TE}_{i,1} = Y_{i,1}[D_{i,1} = 1] - \mathbb{E}[Y_{j,1}|D_{j,1} = 0].$$

$\widehat{TE}_{i,1}$ is an unbiased estimator of the individual treatment effect b_i. Within this motivating section, we are interested in evaluating the value of the following adaptive treatment assignment:

$$D_{i,1} = 1 \quad \text{and} \quad D_{i,2} = \mathbf{1}_{\widehat{TE}_{i,1}>0}. \tag{4}$$

In words, treatment only continues if the effects during the first period are above a minimum threshold.

Dynamic treatment assignment exploits the option value embedded in period 1 experimentation. The value associated with dynamic treatment rule (4) is

$$V \equiv \mathbb{E}\left[Y_{i,1}(D_{i,1} = 1) + Y_{i,2}\left(D_{i,2} = \mathbf{1}_{\widehat{TE}_{i,1}>0}\right)\right] - \mathbb{E}[Y_{i,1}(D_{i,1} = 0) + Y_{i,2}(D_{i,2} = 0)]. \tag{5}$$

The data generated by an experiment with persistent assignment does not allow us to directly estimate value V because the pair of assignment choices $(D_{i,1} = 1, D_{i,2} = 0)$ is not in sample. The remainder of this section establishes key properties of V, and shows how to estimate it using experimental data with persistent assignment.

A treatment with an ATE of 0 can have positive value. A first observation is that a treatment with an ATE of 0 can have positive option value V. More generally, keeping the ATE fixed, option value V increases with the standard deviation of treatment effect σ_b.

Observation 1 (Heterogeneity increases the value of dynamic treatment.)**.** *(i) Assume that $\mu_b = ATE/2 = 0$. Dynamic treatment rule (4) yields a positive value $V > 0$ if and only if the variance of treatment effects is positive: $\sigma_b > 0$.*

(ii) Value V is increasing in μ_b and σ_b.

Proof. Point (i) is immediate. We turn to point (ii). The value V of dynamic treatment can be expressed as

$$V = \mu_B + \mathbb{E}_{b_i,\varepsilon_{i,1}}[b_i \mathbf{1}_{b_i+\varepsilon_{i,1}>0}] = \mu_B + \mathbb{E}_{\varepsilon_{i,1}}\mathbb{E}_{b_i}[b_i \mathbf{1}_{b_i+\varepsilon_{i,1}>0}].$$

For any $\varepsilon_{i,1}$, the function $b_i \mapsto b_i \mathbf{1}_{b_i+\varepsilon_{i,1}>0}$ is increasing and convex in b_i. It follows that it is increasing in both μ_b (a first-order stochastic dominance increase in the distribution of b_i), and in σ_b (a mean preserving spread in the distribution of b_i). □

Naïve option value estimators are biased when errors are correlated. We now turn to the problem of estimating V from data. This would be immediate if we had access to rich experiments such that all pairs of treatment assignments $(D_{i,1}, D_{i,2}) \in \{(0,0),(1,1),(1,0),(0,1)\}$ are in sample and drawn independently across units.

Whenever expectations can be consistently estimated using appropriate sample means from the data, we say that they can be computed with sample data. Throughout the paper we denote by hatted variables $\widehat{\cdot}$ statistics that can be computed using experimental data. When all pairs of treatment assignment are in sample, all expectations in (5) can be computed with sample data.

As we already highlighted, this is no longer possible when data is derived from persistent random assignment. Let

$$\widehat{TE}_{i,2} = Y_{i,2} - \mathbb{E}[Y_{j,2}|D_{j,1} = D_{j,2} = 0].$$

$\widehat{TE}_{i,2}$ is an unbiased estimator of b_i that can be computed using sample data from experiments with persistent assignment. A natural but problematic estimator of V is

$$\widehat{V}_{\text{naïve}} = \mathbb{E}\left[\widehat{TE}_{i,1} + \widehat{TE}_{i,2}\mathbf{1}_{\widehat{TE}_{i,1}>0} \,\middle|\, D_{i,1} = D_{i,2} = 1\right].$$

Estimator $\widehat{V}_{\text{naïve}}$ can be computed even though treatment assignment is constant over time. Unfortunately, it turns out to be biased whenever $\rho \neq 0$, i.e. when error terms are correlated over time.

Observation 2 (Bias)**.** *$\widehat{V}_{\text{naïve}} = V$ if and only if $\rho = 0$. If $\rho \neq 0$, then $\rho \times (\widehat{V}_{\text{naïve}} - V) > 0$.*

Proof. The bias of estimator $\widehat{V}_{\text{naïve}}$ can be expressed as

$$\begin{aligned}\widehat{V}_{\text{naïve}} - V &= \mathbb{E}[(\widehat{TE}_{i,2} - b_i)\mathbf{1}_{\widehat{TE}_{i,1}>0}] = \mathbb{E}[(b_i + \rho\varepsilon_{i,1} + (1-\rho)\varepsilon_{i,2} - b_i)\mathbf{1}_{b_i+\varepsilon_{i,1}>0}] \\ &= \rho\mathbb{E}_{b_i}\underbrace{\mathbb{E}_{\varepsilon_{i,1}}[\varepsilon_{i,1}\mathbf{1}_{\varepsilon_{i,1}+b_i>0}]}_{>0}\end{aligned}$$

where we use the fact that the covariance of monotonic functions of a random variable (here $\varepsilon_{i,1}$) is positive. This concludes the proof. □

In words, estimator $\widehat{V}_{\text{naïve}}$ is consistent if and only if errors are uncorrelated over time. When errors are positively (resp. negatively) correlated, then $\widehat{V}_{\text{naïve}}$ over-estimates (resp. under-estimates) the value of dynamic treatment V.

Concretely, when errors are positively correlated across periods, observing that treated units with high first-period outcomes also perform well in the second period does not constitute compelling evidence that dynamic assignment is valuable.

Placebo bounds on bias. It follows from Observation 2 that consistent estimates of V must control for correlation in error terms. We show how to do so in two steps. We first provide an upper bound on bias based on a placebo value estimate. We then provide a

consistent estimator that exploits more finely the structure of the data-generating process defined by (1) and (2).

We define a placebo treatment effect for untreated subjects by

$$\widehat{TE}^0_{i,t} = Y_{i,t}[D_{i,t} = 0] - \mathbb{E}[Y_{j,t}|D_{j,t} = 0].$$

The associated placebo value of treatment is

$$\widehat{V}_{\mathsf{placebo}} \equiv \mathbb{E}\left[\widehat{TE}^0_{i,1} + \widehat{TE}^0_{i,2}\mathbf{1}_{\widehat{TE}^0_{i,1}>0} \,\middle|\, D_{i,1} = D_{i,2} = 0\right].$$

This statistic can be computed using data from the control group only, and allows us to bound the magnitude of bias $\widehat{V}_{\mathsf{naïve}} - V$.

Observation 3 (Bounding the bias)**.** $|\widehat{V}_{naïve} - V| \leq |\widehat{V}_{placebo}|$.

Proof. Consider the case where $\rho > 0$. We know from the proof of Observation 2 that $\widehat{V}_{\mathsf{naïve}} - V = \rho\mathbb{E}[\varepsilon_{i,1}\mathbf{1}_{\varepsilon_{i,1}+b_i>0}] > 0$. We also have that $\widehat{V}_{\mathsf{placebo}} = \rho\mathbb{E}[\varepsilon_{i,1}\mathbf{1}_{\varepsilon_{i,1}>0}]$.

Observe that for all values of $\varepsilon_{i,1}$,

$$\varepsilon_{i,1}(\mathbf{1}_{\varepsilon_{i,1}>0} - \mathbf{1}_{\varepsilon_{i,1}+b_i>0}) \geq 0.$$

This implies that $0 \leq \widehat{V}_{\mathsf{naïve}} - V \leq \widehat{V}_{\mathsf{placebo}}$. An identical argument applies when $\rho < 0$. □

Identifying the option value of treatment. The bound on V provided by Observation 3 relies on the assumption that the time-series structure of error terms is not affected by treatment. Without additional assumptions, it is possible to precisely identify the option value of treatment by estimating the distribution of error terms and treatment effects.

Observation 4. *Bias term* $\widehat{V}_{naïve} - V = \rho\mathbb{E}[\varepsilon_{i,1}\mathbf{1}_{\varepsilon_{i,1}+b_i>0}]$ *can be estimated using data from a persistent treatment experiment.*

Proof. The distributions of $\varepsilon_{i,1}$ and $b_i+\varepsilon_{i,1}$ can respectively be estimated from the distribution of control and treatment outcomes in period 1. Since b_i and $\varepsilon_{i,1}$ are independent, this means that the distribution of b_i is identified through deconvolution. Finally, we can compute ρ using the covariance between period 1 and 2 outcomes for control units. Altogether, this lets us compute $\rho\mathbb{E}[\varepsilon_{i,1}\mathbf{1}_{\varepsilon_{i,1}+b_i>0}]$. □

3 General Analysis

We now expand on the framework of Section 2 and propose assumptions under which the value of dynamic treatment rules is identified, even when the data exhibits persistent treatment assignment.

3.1 Framework

Outcomes of interest $Y_{i,t}$ are observed over three phases:[4]

$$\mathcal{T}_0 = \{0, \dots, T_0\}, \quad \mathcal{T}_1 = \{T_0 + 1, \dots, T_1\}, \quad \text{and} \quad \mathcal{T}_2 = \{T_1 + 1, \dots, T_2\}.$$

Dates $t \in \mathcal{T}_0$ correspond to a pre-treatment phase, dates $t \in \mathcal{T}_1$ correspond to a first treatment phase, while dates $t \in \mathcal{T}_2$ correspond to a second and final treatment phase. Let $\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2$.

In each treatment phase, units $i \in I$ are assigned to either the control or the treatment group. Treatment assignment for a unit is constant throughout a treatment phase. We denote by (D_i^1, D_i^2) the random variable describing a unit's treatment status in phases 1 and 2. We denote by $(d_i^1, d_i^2) \in \{0, 1\}^2$ its realized values.

Let $\mathbf{Y}_i^0 = (Y_{i,t})_{t\in\mathcal{T}_0}$, $\mathbf{Y}_i^1 = (Y_{i,t})_{t\in\mathcal{T}_1}$, $\mathbf{Y}_i^2 = (Y_{i,t})_{t\in\mathcal{T}_2}$ denote the phase specific outcome

[4] A priori, $Y_{i,t}$ is a scalar, but may be a vector if there are multiple outcomes of interest, e.g. different forms of crime, citizen satisfaction metrics, policing cost and police-force turnover.

vectors. The value associated with outcomes $\mathbf{Y}_i^1$ and $\mathbf{Y}_i^2$ takes the form

$$\left\langle \mathbf{w}_1, \mathbf{Y}_i^1 \right\rangle + \left\langle \mathbf{w}_2, \mathbf{Y}_i^2 \right\rangle,$$

where $\langle \cdot \rangle$ refers to the usual dot product, and $\mathbf{w}_1$ and $\mathbf{w}_2$ are vectors of valuation weights, such as the ones associated with time discounting. At this level of generality, outcomes $Y_{i,t}$ can be multidimensional, for instance including covariates relevant for modeling.[5]

We follow the potential outcome framework, and denote by $\mathbf{Y}_i^1(D_i^1)$ and $\mathbf{Y}_i^2(D_i^1, D_i^2)$ the random outcomes associated with different treatment assignments.

We denote by $\omega_i = (D_i^1, D_i^2, \mathbf{Y}_i^1(0), \mathbf{Y}_i^1(1), \mathbf{Y}_i^2(0,0), \mathbf{Y}_i^2(1,0), \mathbf{Y}_i^2(1,1)) \in \Omega$ the relevant underlying random state, and by Ω the associated state-space. We assume that ω_i is drawn i.i.d. from a distribution $\lambda \in \Lambda \subset \Delta(\Omega)$ where Λ is the set of data-generating processes (DGPs) entertained by the econometrician. We denote by μ the induced distribution of the observed data $(D_i^1, D_i^2, \mathbf{Y}_i^1(D_i^1), \mathbf{Y}_i^2(D_i^1, D_i^2))$.

Definition 1 (identified statistics)**.** *For any function $f : \Lambda \to \mathbb{R}^n$, we say that statistic $f(\lambda)$ is identified if and only if for any feasible distribution μ of the data, the set of values for statistics f consistent with μ, $\{f(\lambda)$ s.t. λ induces $\mu\}$, is a singleton.*

The value of dynamic treatment. Let V_1 denote the effect of permanent treatment on the treated, or ATT:

$$\begin{aligned} V_1 \equiv & \mathbb{E}\left[\left\langle \mathbf{w}_1, \mathbf{Y}_i^1(1) \right\rangle + \left\langle \mathbf{w}_2, \mathbf{Y}_i^2(1,1) \right\rangle | D_i^1 = D_i^2 = 1\right] \\ & - \mathbb{E}\left[\left\langle \mathbf{w}_1, \mathbf{Y}_i^1(0) \right\rangle + \left\langle \mathbf{w}_2, \mathbf{Y}_i^2(0,0) \right\rangle | D_i^1 = D_i^2 = 1\right]. \end{aligned}$$

We are interested in evaluating the impact of a dynamic treatment rule of the form $\delta : \mathbf{Y}_i^0, \mathbf{Y}_i^1 \mapsto \delta_i \in \{0, 1\}$, mapping outcomes available by the end of phase 1 to a treatment

[5] If outcome $Y_{i,t}$ is multidimensional (i.e. $Y_{i,t} = (Y_{i,t,k})_{k \in \{1,\ldots,K\}}$), then each phase outcome vector is the associated flattened matrix, e.g. $\mathbf{Y}_i^0 = (Y_{i,0,1}, Y_{i,0,2}, \ldots, Y_{i,0,K}, Y_{i,1,1}, \ldots, Y_{i,T_0,K})$.

assignment in phase 2:

$$\begin{aligned} V_\delta \equiv & \mathbb{E}\left[\left\langle \mathbf{w}_1, \mathbf{Y}_i^1(1)\right\rangle + \left\langle \mathbf{w}_2, \mathbf{Y}_i^2(1,\delta_i)\right\rangle | D_i^1 = D_i^2 = 1\right] \\ & - \mathbb{E}\left[\left\langle \mathbf{w}_1, \mathbf{Y}_i^1(0)\right\rangle + \left\langle \mathbf{w}_2, \mathbf{Y}_i^2(0,0)\right\rangle | D_i^1 = D_i^2 = 1\right] \end{aligned}$$

Observe that

$$\begin{aligned} V_\delta = & V_1 - \underbrace{\mathbb{E}\left[\left\langle \mathbf{w}_2, \mathbf{Y}_i^2(1,1)\right\rangle \mathbf{1}_{\delta_i=0} | D_i^1 = D_i^2 = 1\right]}_{A} \\ & + \underbrace{\mathbb{E}\left[\left\langle \mathbf{w}_2, \mathbf{Y}_i^2(1,0)\right\rangle \mathbf{1}_{\delta_i=0} | D_i^1 = D_i^2 = 1\right]}_{B}. \end{aligned} \tag{6}$$

Term V_1 is the usual effect of (persistent) treatment on the treated, which can be consistently estimated using standard techniques. In principle, it requires estimating

$$\mathbb{E}[(\mathbf{Y}_i^1(0), \mathbf{Y}_i^1(0,0)) | D_i^1 = D_i^2 = 1],$$

which is not directly in sample. However, estimating the mean of these counterfactual outcomes is standard, and not the focus of this paper. For instance, if $D_i^1 = D_i^2 = D_i$ is randomly assigned so that potential outcomes are the same for treatment and control groups, then we can simply use sample averages in the control group to compute the relevant means. If assignment is not random but treatment and control group satisfy suitable parallel trend conditions, then differences-in-differences methods can be applied.

Option values. The option value of treatment OV_δ, i.e. the value created by selectively dropping treatment in the second phase on the basis of phase 1 outcomes, takes the form

$$OV_\delta \equiv V_\delta - V_1 = B - A.$$

Term A can be computed directly from sample data in a persistent treatment setting. Term B is a problematic term under persistent assignment since it requires evaluating the conditional mean of an unsampled outcome $\mathbf{Y}_i^2(1,0)$.

Note that identifying term B is immediate when assignment (D_i^1, D_i^2) is independent of potential outcomes $(\mathbf{Y}_i^1(0), \mathbf{Y}_i^1(1), \mathbf{Y}_i^2(0,0), \mathbf{Y}_i^2(1,0), \mathbf{Y}_i^2(1,1))$, and there is full support sequential randomization, i.e.

$$\mathsf{Prob}_\lambda(D_i^1 = 1) \in (0,1) \quad \text{and} \quad \mathsf{Prob}_\lambda(D_i^2 = 1 | D_i^1 = 1) \in (0,1).$$

In that case term B can be directly computed using sample means.

Per-period effects. As Hendren and Sprung-Keyser (2020) emphasize, the relevant statistic for adoption is the welfare return on policy costs. When treatment involves significant variable costs, this means that treatment-effect per implementation period, rather than aggregate treatment effect, is the relevant object to identify. Dynamic adoption can further improve efficiency by saving on policy costs. Costs can be implicit, for instance, it could be the shadow cost of police and DA bandwidth assigned to prioritized cases under Operation Ceasefire. Costs can also be financial, as in Dupas (2014) and Ida et al. (2024), which seek to identify the optimal duration of subsidies inducing long-term changes in behavior.

Correspondingly, per-period values,

$$\begin{aligned} V_1^{pP} &\equiv \frac{1}{T_2 - T_0} V_1, \\ V_\delta^{pP} &\equiv \frac{1}{(T_2 - T_1)\mathbb{E}[\delta_i | D_i^1 = 1] + T_1 - T_0} V_\delta, \\ \text{and} \quad OV_\delta^{pP} &\equiv V_\delta^{pP} - V_1^{pP} \end{aligned}$$

are also objects of interest.

3.2 Identification under persistent assignment

We now turn to the main case of interest for this paper, *persistent assignment*:

$$D_i^1 = D_i^2 = D_i.$$

This subsection provides a set of high level assumptions under which V_δ is identified, while the next subsection shows that these assumptions hold for a set of leading DGPs of interest.

As in Section 2, the identification strategy requires understanding the underlying time series process. Some notation is useful. We denote by $t \in \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2$ a time period in which outcome data is available. We denote by $d_{i,t} \in \{0, 1\}$ the treatment status of unit i at time t, and by $\mathbf{d}_i^t = (d_{i,s})_{s \in \{0,\ldots,t\}}$ the history of treatment assignments up to time t.

Assumption 1 (Contemporaneous Impact of Treatment)**.** *There exist known functions Ψ_t, and real-valued processes $\xi_i = (\xi_{i,t})_{t \in \mathcal{T}}$ and $\beta_i = (\beta_{i,t})_{t \in \mathcal{T}}$ such that: ξ_i and β_i are independent of each other, independent across treatment units (but not necessarily independent over time), and*

$$Y_{i,t}(\mathbf{d}_i^t) = \Psi_t(Y_{i,t-1}, [d_i \beta_i]^t, \xi_{i,t}), \tag{7}$$

where $[d_i \beta_i]^t \equiv (d_{i,s}\beta_{i,s})_{s \leq t}$ is the history of experienced treatment effects.

In practice, function ψ_t may have to be estimated either flexibly, or using a parametric family, but the strategies to do so are not novel to this paper. As an example the next section considers affine autoregressive models where coefficients, as well as time fixed-effects are estimated using Arellano and Bond (1991).

Condition (7) expresses an important restriction: treatment $d_{i,t}$ impacts later outcomes only via its contemporaneous impact $\beta_{i,t}$. Note that in principle $Y_{i,t-1}$ can be vector valued, allowing to capture past outcomes or relevant covariates if needed. Condition (7) potentially excludes withdrawal effects of the form associated with antidepressants (Haddad, 1998): going on-and-off anti-depressants can have large negative effects in a way that's not captured

by the effect of treatment on interim depression outcomes.

Let $\xi_i^1 = (\xi_{i,t})_{t \in \mathcal{T}_1}$ and $\xi_i^2 = (\xi_{i,t})_{t \in \mathcal{T}_2}$.

Assumption 2 (Identified Shocks Process)**.** *The joint distribution of* $(\mathbf{Y}_i^0, \xi_i^1, \xi_i^2)$ *conditional on* $D_i^1 = 1$ *is identified.*

Lemma 1. *Under Assumptions 1 and 2, the effect of persistent treatment on the treated,* V_1 *is identified.*

Under Assumption 1, iteratively applying (7) implies that there exists a known vector valued function Ψ^1 such that the following vector equation holds

$$\mathbf{Y}_i^1 = \Psi^1(\beta_i^1, \xi_i^1, \mathbf{Y}_i^0). \tag{8}$$

Assumption 3 (Identified Phase 1 Treatment Parameters)**.** *Given a distribution of* $\mathbf{Y}_i^1$ *and a known distribution of* $(\mathbf{Y}_i^0, \xi_i^1)$ *conditional on* $D_i^1 = 1$*, there exists a unique distribution of* β_i^1 *independent of* ξ_i^1 *such that* (8) *holds.*

Proposition 1 (Identification)**.** *Under Assumptions 1, 2, and 3, the effect of dynamic treatment on the treated* V_δ *is identified.*

Given V_1 and V_δ, it is possible to estimate the option value of treatment OV_δ, as well as value per-period, V_δ^{pP} since the denominator $(T_2 - T_1)\mathbb{E}[\delta_i | D_i^1 = 1] + T_1 - T_0$ can be estimated in sample.

3.3 Application to benchmark DGPs

We now establish that leading frameworks of interest satisfy Assumptions 1, 2, 3. Whenever possible we establish non-parametric identification for error and treatment effect processes. In practice however, limited sample size leads us to use a parametric Gaussian model.

The Markov model. Under a Markov model, transition odds depend only on the current state, current treatment status, and possibly time. Since functions Ψ_t can be flexibly chosen, the distribution of outcomes under such a model can be written as

$$\forall t \in \mathcal{T}, \quad Y_{i,t} \equiv \Psi_t(Y_{i,t-1}, d_{i,t}, \xi_{i,t}) \tag{9}$$

with $(\xi_{i,t})_{t\in\mathcal{T}}$ i.i.d. and uniform. Provided that functions Ψ_t are known or can otherwise be estimated, Assumption 1 holds because the most recent outcome $Y_{i,t-1}$ is a sufficient statistic of past outcomes and the history of treatment. Assumption 2 and 3 hold mechanically since $(\xi_{i,t})_{t\in\mathcal{T}}$ is i.i.d. uniform, and treatment impact parameters $\beta_{i,t}$ are constant and equal to 1.

The auto-regressive parallel trends model. Under the auto-regressive parallel trends model, the process of outcomes satisfies

$$Y_{i,t} = \alpha_i + \lambda_t + \rho Y_{i,t-1} + d_{i,t}\beta_{i,t} + \varepsilon_{i,t} \tag{10}$$

where terms $(\varepsilon_{i,t})_{t\in\mathcal{T}}$ are i.i.d. or MA errors, and α_i has a population mean equal to 0. In this model, like in the Markov model, past treatment has a long-term effect on outcomes via the way it changes past outcomes. In the context of homicides this would be a good fit if, for instance, lower recent homicide rates mean that there is less intrinsic conflict between gangs, which would reduce future violence even if the programs that reduced violence in the first place are removed.

Process (10) takes the form required by Assumption 1,

$$Y_{i,t} = \Psi_t(Y_{i,t-1}, d_{i,t}\beta_{i,t}, \xi_{i,t})$$

with $\xi_{i,t} = \alpha_i + \varepsilon_{i,t}$ and Ψ_t known up to term λ_t and coefficient ρ. Let I represent the set of units in the sample. We know from Arellano and Bond (1991) that ρ is identified from data

$\{(Y_i^0, Y_i^1, Y_i^2)\}_{i \in I}$, which in turn implies that λ_t and ξ_t are identified under the normalization assumption that $\mathbb{E}[\alpha_i] = 0$. This implies that Assumptions 1 and 2 both hold.

We now establish that Assumption 3 also holds. Let

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \rho^{T_1 - T_0 - 1} & \cdots & \rho & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 1 \\ \rho \\ \vdots \\ \rho^{T_1 - T_0 - 1} \end{bmatrix}.$$

We have that

$$\mathbf{Y}_i^1 = \rho Y_{i,T_0} \mathbf{r} + R\left(\lambda^1 + \beta_i^1 + \xi_i^1\right) \tag{11}$$

with

$$\lambda^1 = \begin{bmatrix} \lambda_{T_0+1} \\ \vdots \\ \lambda_{T_1} \end{bmatrix}, \quad \beta_i^1 = \begin{bmatrix} \beta_{i,T_0+1} \\ \vdots \\ \beta_{i,T_1} \end{bmatrix}, \quad \xi_i^1 = \begin{bmatrix} \alpha_i + \varepsilon_{i,T_0+1} \\ \vdots \\ \alpha_i + \varepsilon_{i,T_1} \end{bmatrix}.$$

Observing that since R is invertible, (11) implies that

$$R^{-1}(\mathbf{Y}_i^1 - \rho Y_{i,T_0} \mathbf{r}) - \lambda^1 = \beta_i^1 + \xi_i^1 \tag{12}$$

with β_i^1 and ξ_i^1 independent. Since the left-hand side of (12) is either observed or identified, and the distribution of ξ_i^1 is identified, a multivariate deconvolution argument (see Billingsley, 1995, Chap. 29) implies that the distribution of treatment effects β_i^1 is also identified. This implies that Assumption 3 holds.

In principle, this suggests that non-parametric estimation of the distribution of β_i^1 is possible. However, it is known that the convergence of empirical deconvolution is slow (Fan, 1991). For this reason, our empirical application assumes that β_i^1 follows a multivariate Gaussian distribution.

The latent variable parallel-trend model. Under the latent variable model, the process for outcomes $Y_{i,t}$ is determined by the following pair of equations:

$$Z_{i,t} = \alpha_i + \lambda_t + \rho Z_{i,t-1} + u_{i,t} \tag{13}$$

$$Y_{i,t} = Z_{i,t} + d_{i,t}\beta_{i,t} + v_{i,t} \tag{14}$$

where $u_{i,t}$ and $v_{i,t}$ are independent i.i.d. errors, $Z_{i,t}$ is an unobserved latent state that evolves independently of treatment, and $Y_{i,t}$ depends on the latent state and current treatment.

In contrast to the parallel trends model, under the latent model treatment does not have a persistent effect once it stops. In the context of homicides, this is a suitable model if the underlying causes of violence, such as poverty, or gang competition for territory, are independent of the program being evaluated, and the program only affects how those underlying causes map into violent outcomes.

By substituting (13) into (14), we obtain that

$$\text{for } t = T_0 + 1, \quad Y_{i,t} = \lambda_t + \rho Y_{i,t-1} + \beta_{i,t} d_{i,t} + \xi_{i,t} \tag{15}$$

$$\text{for } t > T_0 + 1, \quad Y_{i,t} = \lambda_t + \rho Y_{i,t-1} + \beta_{i,t} d_{i,t} - \rho\beta_{i,t-1} d_{i,t-1} + \xi_{i,t} \tag{16}$$

with $\xi_{i,t} = \alpha_i + u_{i,t} + v_{i,t} - \rho v_{i,t-1}$.

As in the auto-regressive parallel trends model, we know from Arellano and Bond (1991) that λ_t, ρ and the distribution of ξ_i can still be identified from $\{(Y_i^0, Y_i^1, Y_i^2)\}_{i \in I}$. This implies that Assumptions 1 and 2 hold.

Assumption 3 follows from a deconvolution argument. With matrix R and vector $\mathbf{r}$ the same as above, we have that

$$\mathbf{Y}_i^1 = \rho Y_{i,T_0} \mathbf{r} + R\left(\lambda^1 + \xi_i^1\right) + \beta_i^1. \tag{17}$$

Since β_i^1 and $\rho Y_{i,T_0}\mathbf{r} + R\left(\lambda^1 + \xi_i^1\right)$ are independent, the distribution of β_i^1 can be recovered

from the estimated distribution of $\rho Y_{i,T_0}\mathbf{r} + R(\lambda^1 + \xi_i^1)$ and the observed distribution of $\mathbf{Y}_i^1$ through deconvolution.

4 Evaluating Operation Ceasefire

4.1 Program design

Operation Ceasefire was initiated in Boston in 1995, largely because homicide rates that had risen rapidly during the crack epidemic of the late 1980s were not declining fast enough. The program was led by David Kennedy, Anne Piehl and Anthony Braga, in partnership between the Boston Police Department, the Harvard Kennedy School, and numerous community stakeholders, including parole officers, community outreach organizations, and members of the District Attorney's office. Policies developed during this initial effort have since spread to several dozen cities across the US. Kennedy (2011) provides a vivid description of the program, and the highs and lows of its adoption.

The program was predicated on the following observations. First, the Boston police had fairly confident guesses of who had committed each homicide. Out of 155 victims under the age of 21, 125 (i.e. 80%) had a known or associated killer. Second, a relatively small number of organized groups were responsible for the majority of murders: 60% of homicides were assigned to one of 61 gangs operating in Boston at the time. Finally, gang members made up the majority of both offenders and victims.

In spite of this information, police investigation and judicial pursuits were not an effective deterrent. Cases infrequently led to significant jail time. One difficulty is that generating convictions that stick requires extensive work. Without sufficient evidence, a District, or US Attorney is unlikely to take on a case, let alone generate a conviction leading to real prison time. Successful convictions often require coordination between local police departments, federal law enforcement, as well as local and federal district attorneys. Building up such

cases requires resources that become stretched thin in periods of high crime. As a result, gang members operate under perceived impunity, or worse, under the impression that law enforcement and the community simply do not care about gang on gang violence.

Operation Ceasefire consisted of the following steps:

1. Members of different gangs were brought together for a "call-in", often at the behest of trusted parties, including family members, parole officers, and community leaders.[6]

2. Police publicly established that they were able to associate crimes with gangs.

3. Police and District Attorneys committed to a plausible promise: that they would allocate a disproportionately large amount of resources to getting convictions against gangs responsible for the next several homicides.

In his account, Kennedy (2011) relates a comment by then lieutenant Gary French highlighting the prioritization of police resources following the call-in:

> "I remember saying to you six or seven months ago, we're responding all the time, we may not be able to stay in one area for long. [...] But now that we're not just putting out fires all the time, the guys really want something to do, we can do it right."

4.2 Program adoption

There are two difficulties in constructing our sample of municipalities experimenting with Operation Ceasefire: first, there is no pre-existing centralized database of adoption events; second, even if homicide reduction programs are explicitly modeled after Operation Ceasefire,

[6]The working team of the original Boston Operation Ceasefire also met some particular group members individually and delivered individualized warnings after call-ins (Kennedy and Friedrich, 2014). Such individual meetings are commonly referred to as custom notifications. A few replications of Operation Ceasefire, for example, the Minneapolis HEALS Initiative in Minneapolis, Minnesota, relied solely on custom notifications, while the others only used them as a supplementary communication method for call-ins.

there is variation in which aspects of Operation Ceasefire are kept and which are adjusted. We view this as an unavoidable aspect of scaling.

Constructing a timeline of adoption events. We base our data collection effort on three sources:

- A list of jurisdictions members of the National Network for Safe Communities.
- Kennedy (2011), the articles it cites, and the articles cited in this paper.[7]
- Systematic internet searches described below.

We compile a comprehensive list of programs implemented between 1996 and 2018 that satisfy at least one of the following two criteria: (1) the program is explicitly reported as being modeled on Boston Operation Ceasefire; (2) the program conducted call-ins (group meetings) or custom notifications (individual meetings), in which deterrence messages were delivered to targeted individuals. We exclude programs that did not focus on homicides, or gang violence even when they share similar designs.[8] We ignore programs from Florida because the associated homicide data from our data source is not complete.

We start from an initial list of places and search for programs implemented in them. The initial list of places consists of listed NNSC member jurisdictions and cities that implemented a program of interest based on Kennedy (2011), NNSC official websites, Kennedy and Friedrich (2014), National Network for Safe Communities (2015), National Network for Safe Communities (2016), and the first page Google search results using "operation ceasefire", "group violence intervention", or "David Kennedy" as the search term.[9]

[7]All references are available at `https://www.dropbox.com/scl/fi/9od8h62y3xdi39gpohru3/references.zip?rlkey=oy15k71gvs867wv9ftnw2vwk8&st=i4zrz69z&dl=0`.

[8]This includes Drug Market Intervention (DMI) programs modeled on Boston Operation Ceasefire. DMI also implements call-ins and custom notifications. However, National Network for Safe Communities (2015) describes DMIs as designed specifically to close overt drug markets and suggests cities suffering from serious violence follow the Group Violence Intervention, rather than the DMI, model. Other NNSC programs such as Intimate Partner Violence Intervention are also excluded.

[9]See National Network for Safe Communities (2016), `https://web.archive.org/web/2020062223`

For each place on the initial list, we conduct three preliminary Google searches, pairing the place name with "operation ceasefire", "group violence intervention", or "David Kennedy" as the search terms. We review every accessible reference on the first page of each search.

For programs that appear to be of our interest but do not have call-in news, custom notification news, or an official program website in all first-page search results, we conduct supplementary searches for such references. To find official program websites, we combine the place name and program name as the search term and review all search results on the first page. The process is the same for news reports of call-ins and custom notifications, except that we add "call-in" or "custom notification" in the search term. When references from preliminary searches refer to call-ins or custom notifications with other local names, we use the local names in the supplementary search.

During this search process, whenever we encounter evidence that an additional city implemented a relevant program, we add it to the search list and conduct to the same search process as the one described above.

The resulting program list includes 122 programs in 88 cities across 26 states.[10] We attempt to match each city with a local law enforcement agency based on its county and name and end up with 87 matched cities. For each program, we hand-code its characteristics based on the program-specific references from both preliminary searches and supplementary searches.

Since the homicide data reporting system in the United States transitioned from the Summary Reporting System to the National Incident-Based Reporting System in 2021 and FBI identified 2018 as the pivotal transition year, we restrict our homicide data to 2018 to maintain consistent variable definitions.[11] Programs started in 2014 are thus the latest

4248/https://nnscommunities.org/impact/cities/, and http://johnjay.jjay.cuny.edu/newsroom/2666.php for lists of NNSC member jurisdictions. See https://web.archive.org/web/20200518121147/https://nnscommunities.org/strategies/group-violence-intervention/ and https://web.archive.org/web/20190331174953/https://nnscommunities.org/our-work/strategy/group-violence-intervention for official NNSC websites.

[10]We started to work on the dataset in February 2020 and finalized it in August 2022.

[11]See, for example, https://www.fbi.gov/file-repository/ucr/fbi-letter-on-nibrs-transitio

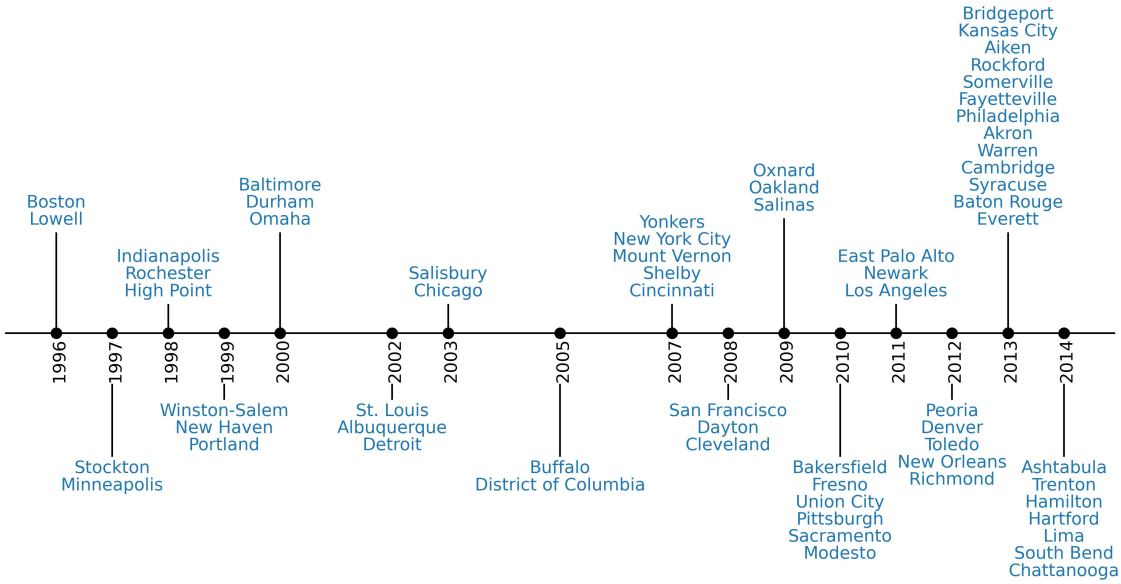


Figure 2: Timeline of initial call-in or custom notification by location

programs for which we can observe the treatment effects in the first five treatment years. For our main results, we define treatment units as the 65 cities, or equivalently the corresponding 65 local law enforcement agencies given our matched pairs, that have at least one program between 1996 and 2014 with call-ins or custom notifications.

For each treatment unit, we define its first treatment year as the year in which its first recorded program with call-ins and/or custom notifications started. Figure 2 plots the treatment units along with their observed first treatment year.

Program variants. The original Boston Operation Ceasefire emphasizes (1) direct communication with targeted individuals, (2) common knowledge of enforcement priorities, (3) group-based enforcement, and (4) collaboration among law enforcement agencies, community members, and social service providers. In practice, replications do not always include all four components. With our hand-coded characteristics, we summarize differences across programs underlying Figure 2 as follows:

`n-071018.pdf` for a description about the transition. Last accessed on February 22, 2026.

- All 65 programs communicated deterrence messages to targeted individuals, either through call-ins or custom notifications. Call-ins are more common than custom notifications: 63 programs conducted call-ins, whereas 16 conducted custom notifications.

- At call-ins, 8 programs communicated prioritized enforcement against the next or most violent individual or group after the call-ins, as in the original Boston Operation Ceasefire. Among other programs, 33 communicated prioritized enforcement only against the individuals or groups in attendance, without extending the commitment to the next or most violent group. In total, 41 programs communicated common-knowledge enforcement priorities at call-ins.

- 56 programs explicitly targeted group violence. However, only 31 of these programs included group-based enforcement, under which sanctions were directed at the entire group when one or more members committed crimes.

- 58 programs involved law enforcement agencies, community members, and social service providers. In 42 such programs, representatives from all sides delivered anti-crime messages at call-ins.

Taken together, 25 programs adopted all four core components. 21 programs adopted at least three core components, while 19 adopted two or fewer.

Finally, it is worth noting that among the 65 programs, 37 programs explicitly targeted homicides and 59 programs explicitly targeted firearm violence. All programs explicitly targeted at least one among homicides, firearm violence, and group violence.

4.3 Homicide data

Homicide data from 1990 to 2018 is obtained from the Supplementary Homicide Reports (SHR) database. Each observation corresponds to a reported homicide incident defined by the Uniform Crime Reporting (UCR) Program. The reporting agencies, each identified by

a unique seven-digit originating agency identifier (ORI) code, include agencies such as the local police departments and the sheriffs offices.

We are interested in the effect of treatment on the yearly intentional homicide rate per 100,000 people. To calculate it, we use the recorded population corresponding to each ORI in the observations. For each calender year t and ORI i, we divide the yearly total number of reported intentional homicides with the recorded population, and multiply it by 100,000 to obtain the corresponding yearly intentional homicide rate Y_{it} per 100,000 inhabitants.

We restrict the sample to ORIs representing large local police departments, as defined in Bureau of Justice Statistics (2018), that have population larger than 10,000 more often than not between 1990 and 2018. We also exclude control ORIs that did not report at least one homicide, either intentional or unintentional, in every 60 consecutive months between 1990 and 2018, as well as treatment ORIs that did not do so in the two years immediately before their treatment.[12] Let $\Delta Y_{i,t} \equiv Y_{i,t} - Y_{i,t-1}$. To ensure we focus on data for which parallel trend assumptions hold, we exclude control ORIs for which the pair $(\Delta Y_{i,t-1}, \Delta Y_{i,t})$ is an outlier between 1992 and 2018, as well as treatment ORIs for which $(\Delta Y_{i,t-1}, \Delta Y_{i,t})$ is an outlier in the year immediately before their treatment. Finally, we exclude control ORIs that appear in the list of municipalities evaluated for inclusion in Section 4.2 but are not treatment ORIs.[13] Such ORIs potentially had other programs related to Operation Ceasefire and thus are not valid controls.

In the end, there are 634 control ORIs and 58 treatment ORIs in the sample.[14] The sample construction process is described in greater detail in Appendix A.

Table 1 contains the summary statistics for the control and treatment units respectively.

[12]For treatment ORIs, only the two years immediately before their treatment matters for calculating counterfactuals.

[13]These are ORIs corresponding to listed member jurisdictions of NNSC but did not implement programs satisfying our inclusion criteria.

[14]The changes in the numbers of treatment and control ORIs from the data cleaning process are consistent with data errors in the UCR data reported in earlier research such as Evans and Owens (2007), Mello (2019), and Weisburst (2019).

		N	Mean	SD
Population	Treatment	851	483913	1062636
	Control	13948	104992	167290
Intentional homicide rate (Y_{it})	Treatment	851	15.74	11.77
	Control	13948	4.43	4.69
Change between adjacent years (ΔY_{it})	Treatment	851	-0.19	4.73
	Control	13948	-0.12	4.13

Notes: Each control unit contributes 22 yearly observations, each corresponding to a different year between 1992 and 2013. Each treatment unit contributes a potentially different number of yearly observations, each corresponding to a different pre-treatment year for which $(\Delta Y_{i,t-1}, \Delta Y_{it})$ is not an outlier.

Table 1: Summary statistics.

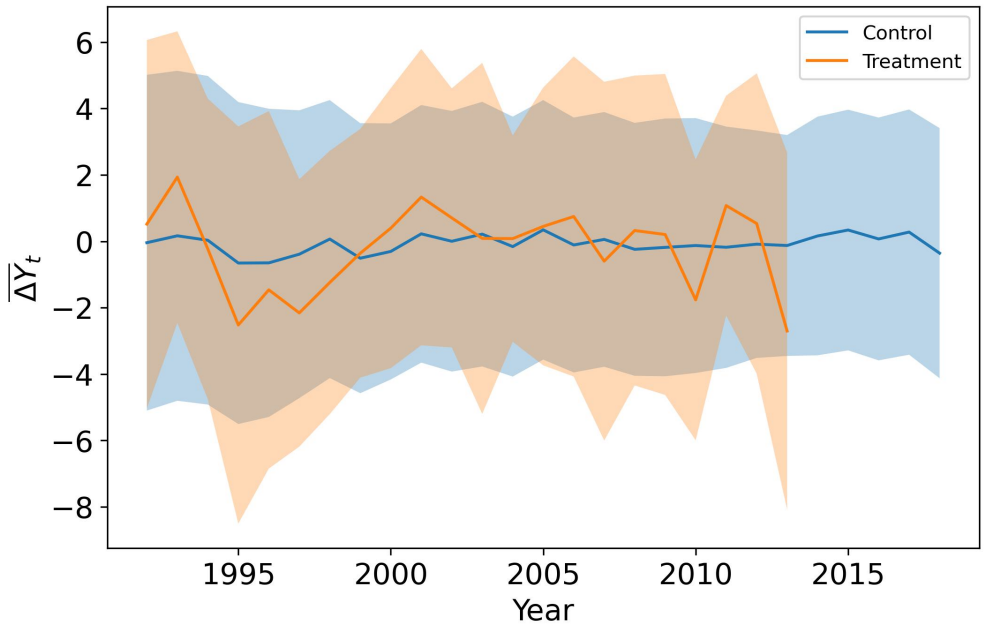
Compared to control ORIs, treatment ORIs have larger average reported intentional homicide rates and recorded population. This suggests that the decision to implement a program along the lines of Operation Ceasefire was not random.

Table 1 shows that the difference between Y_{it} and $Y_{i,t-1}$, represented by ΔY_{it}, for treatment units is comparable to year-on-year differences for control units in terms of their sample mean and standard deviation. Figure 3 further confirms the comparability by plotting ΔY_{it} over time with the same set of observations as Table 1. We thus attribute the distributional differences in the intentional homicide rate between the treatment and control ORIs to distributional differences in ORI fixed effects. Accordingly, in Section 5, we estimate a dynamic panel data model with control ORIs and then apply the estimates to treatment ORIs.

5 Findings

We apply the framework of Section 3 as follows:

- Phase 1 corresponds to the first 2 years of treatment, while phase 2 corresponds to the next three years of treatment.



Notes: Lines show the average ΔY_{it} over time separately for the control and treatment units. Shaded regions indicate ± 1 standard deviation. In calculating yearly averages and standard deviations, each control unit contributes one observation in every year, whereas each treatment unit contributes one observation only in pre-treatment years for which $(\Delta Y_{i,t-1}, \Delta Y_{it})$ is not an outlier.

Figure 3: The average ΔY_{it} over time by treatment status

- Sequences of outcomes are valued using constant weights: $\mathbf{w}_1 = (1, 1)$ and $\mathbf{w}_2 = (1, 1, 1)$.

- Treatment effects are evaluated under both the auto-regressive and the latent-variable parallel-trend DGPs described in Section 3.

- Under dynamic adoption, the program is maintained in phase 2 provided that treatment effect estimates in phase 1 are negative. Note that this treatment continuation decision turns out to be the same under the two candidate DGPs.

5.1 Treatment effect estimates under persistent assignment

Let us define phase 1 and phase 2 treatment effects under persistent assignment as

$$TE_i^1 \equiv \left\langle \mathbf{w}_1, \mathbf{Y}_i^1(1) - \mathbf{Y}_i^1(0) \right\rangle \quad \text{and} \quad TE_i^2 \equiv \left\langle \mathbf{w}_2, \mathbf{Y}_i^2(1,1) - \mathbf{Y}_i^2(0,0) \right\rangle .$$

Given a constant treatment regime $D_i^1 = D_i^2$, it turns out that stochastic processes for outcomes $(Y_{i,t})_{i\in I, t\in\mathcal{T}}$ generated under the latent-variable model can also be generated by the auto-regressive with different rationalizations in terms of coefficients $(\beta_{i,t})_{i\in I, t\in\mathcal{T}}$, and error terms $\epsilon_{i,t}$ following an auto-correlated MA process. It follows that estimates of TE_i^1 and TE_i^2 coincide under the two models. Only estimates following assignment sequences $D_i^1 = 1$ and $D_i^2 = 0$ differ.

For this reason this section describes how we estimate TE_i^1 and TE_i^2 under the auto-regressive model. Estimates under the latent variable model are identical. To apply Arellano and Bond (1991) we take first differences on (10):

$$\Delta Y_{i,t} = \Delta\lambda_t + \rho\Delta Y_{i,t-1} + \beta_{i,t}d_{i,t} - \beta_{i,t-1}d_{i,t-1} + \Delta\epsilon_{i,t}. \tag{18}$$

We estimate $\Delta\lambda_t$ and ρ using a randomly selected set of 508 control units (80% of the control sample) and use an GMM estimator as in Arellano and Bond (1991), adjusting instruments to allow for ϵ_{it} following i.i.d., MA(1), MA(2), and MA(3) processes.[15]

To calculate standard errors , we perform a block bootstrap procedure with 500 rounds. In each round, we first resample 508 control ORIs from the 508 training ORIs and 58 treatment ORIs from the 58 treatment ORIs with replacement. We then rerun the whole estimation procedure on the panel data corresponding to the resampled ORIs.

Table 2 reports estimates for ρ. Though there is a large difference between $\hat{\rho}$ estimated under i.i.d. errors and that estimated under MA(1) errors, $\hat{\rho}$ is largely stable as we allow for

[15]When ϵ_{it} is i.i.d., $(Y_{i,s})_{s\leq t-2}$ are valid instruments. When it instead follows an MA(q) process for $q = 1, 2, \ldots$, only $(Y_{i,s})_{s\leq t-2-q}$ are valid.

	i.i.d. errors	MA(1) errors	MA(2) errors	MA(3) errors
$\hat{\rho}$	0.2***	0.71***	0.65***	0.6***
	(0.03)	(0.03)	(0.03)	(0.04)
	[0.0]	[0.0]	[0.0]	[0.0]

Notes: The table shows estimates under different assumptions on the error term, ϵ_{it}. Clustered standard errors at the ORI level are in parentheses. P-values are in brackets. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 2: Estimates for ρ.

more persistent intertemporal correlations in the errors. It is always positively significant and changes only from 0.71 to 0.6, providing supportive evidence for ϵ_{it} being MA(1). From now on, we thus report estimates under both the auto-regressive model and the latent variable model assuming that ϵ_{it} is MA(1).

Given estimators $\widehat{\Delta\lambda}_t$ and $\widehat{\rho}$, we estimate coefficients $\beta_{i,t}$ with

$$\hat{\beta}_{i,t} = \Delta Y_{i,t} - \widehat{\Delta\lambda}_t - \hat{\rho}\Delta Y_{i,t-1} \text{ for } t = 1 \text{ and}$$
$$\hat{\beta}_{i,t} = \Delta Y_{it} - \widehat{\Delta\lambda}_t - \hat{\rho}\Delta Y_{i,t-1} + \hat{\beta}_{i,t-1} \text{ for } t > 1.$$

This yields treatment effect estimators

$$\widehat{TE_i^1} = \left\langle \mathbf{w}_1, \widehat{R}_1\widehat{\beta}_i^1 \right\rangle \quad \text{and} \quad \widehat{TE_i^2} = \left\langle \mathbf{w}_2, \hat{R}_{12}\hat{\beta}_i^1 + \widehat{R}_2\widehat{\beta}_i^2 \right\rangle$$

with

$$\widehat{R}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \widehat{\rho} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \widehat{\rho}^{T_1-T_0-1} & \cdots & \widehat{\rho} & 1 \end{bmatrix}, \widehat{R}_{12} = \begin{bmatrix} \hat{\rho}^{T_1-T_0} & \hat{\rho}^{T_1-T_0-1} & \cdots & \hat{\rho} \\ \hat{\rho}^{T_1-T_0+1} & \hat{\rho}^{T_1-T_0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\rho}^{T_2-T_1-1} \\ \hat{\rho}^{T_2-T_0-1} & \cdots & \hat{\rho}^{T_2-T_1+1} & \hat{\rho}^{T_2-T_1} \end{bmatrix},$$

	$\widehat{ATT}$	$\widehat{ATTpP}$
MA(1) errors	-3.81	-0.76
	(5.65)	(1.13)
	[0.5]	[0.5]

Notes: The table shows estimates under the assumption that the error term, ϵ_{it}, is MA(1). Bootstrap standard errors are in parentheses. Two-sided p-values computed with the standard normal distribution are in brackets. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3: The ATT estimates

and

$$\widehat{R}_2 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \widehat{\rho} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \widehat{\rho}^{T_2-T_1-1} & \cdots & \widehat{\rho} & 1 \end{bmatrix}.$$

ATTs. These treatment effect estimates allow us to estimate aggregate average treatment effects on the treated (ATTs), by averaging $\widehat{TE_i^1} + \widehat{TE_i^2}$ over treated units. ATTs per treated period are recovered by dividing ATT estimates by $T_2 - T_0 = 5$.

Table 3 reports that the estimate for the ATT common to the two models is equal to -3.81, which translates to a decrease of -0.76 per 100,000 inhabitants per treatment year. This point estimate represents a 5% reduction in homicide rates in treated units, and the estimates are insignificant. The impact of programs modeled on Operation Ceasefire is on average ambiguous.

Heterogeneous treatment effects. As we argue in Section 2, if treatment effects are heterogeneous, programs that have limited value on average can become quite valuable if

they are applied to suitably selected units.

One possibility is to base selection on ex ante observed covariates. It turns out that basic unit-level characteristics X_i such as the population and region are not highly correlated with the estimated effect $\widehat{TE}_i^1 + \widehat{TE}_i^2$.

In contrast, as highlighted in Figure 1, the raw correlation between treatment effect estimate $\widehat{TE}_i^1$ and $\widehat{TE}_i^2$ is positive and large, equal to 0.94. This suggests a dynamic adoption strategy taking the form

$$\delta\left(\widehat{TE}_i^1\right) = \mathbf{1}_{\widehat{TE}_i^1 < 0}.$$

Though the high correlations may partially result from the inherent auto-correlation in $\Delta\epsilon_{it}$, it seems that the dynamic treatment regime that assigns the long-run treatment only to treated units with negative estimated cumulative treatment effects in the short run is a promising candidate treatment regime for generating a negative and significant option value. We now investigate this possibility.

5.2 Option value estimates

Naïve values. Let I_1 denote the set of treated units. We define the naïve value of dynamic treatment $\widehat{V}^{\mathsf{naïve}}$ by

$$\widehat{V}^{\mathsf{naïve}} \equiv \frac{1}{\mathsf{card} I_1} \sum_{i \in I_1} \widehat{TE}_i^1 + \mathbf{1}_{\widehat{TE}_i^1 > 0} \widehat{TE}_i^2.$$

In turn we compute associated option value $\widehat{OV}^{\mathsf{naïve}} \equiv \widehat{V}^{\mathsf{naïve}} - \widehat{ATT}$ and per-period metrics $\widehat{V}_{pP}^{\mathsf{naïve}}$, and $\widehat{OV}_{pP}^{\mathsf{naïve}} \equiv \widehat{V}_{pP}^{\mathsf{naïve}} - \widehat{ATT}_{pP}$. Estimates are reported in Table 4.

These naïve value estimates now suggest a large impact of dynamic treatment, corresponding to statistically significant a 28.5% reduction in the homicide rate. Of course, correlation in estimation over time likely biases these estimates. The question is how much.

	$\widehat{V}^{\text{naïve}}$	$\widehat{OV}^{\text{naïve}}$	$\widehat{V}_{pP}^{\text{naïve}}$	$\widehat{OV}_{pP}^{\text{naïve}}$
Estimate	-16.18***	-12.38***	-4.49***	-3.73***
s.e.	(4.2)	(2.36)	(1.02)	(0.5)
p-value	[0.0]	[0.0]	[0.0]	[0.0]

Notes: Bootstrap standard errors are in parentheses. Two-sided p-values are in brackets. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 4: Naïve value estimates

Placebo test. A placebo test, simulating the distribution of $\widehat{OV}^{\text{naïve}}$ with pseudo treatment units drawn from the control group suggests that bias is high, but that dynamic treatment may be effective. As Table 5, the estimated $\widehat{OV}^{\text{naïve}}$ is below the 10^{th} percentiles of its placebo distribution, but not below the 5^{th} percentile.

Quantile	5^{th}	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}	95^{th}
	-12.85	-11.79	-10.55	-9.04	-7.59	-6.42	-6.02

Table 5: Placebo distribution of $\widehat{OV}^{\text{naïve}}$

Debiased values. Debiased estimates of the value of dynamic treatment $\widehat{V}$, of the associated option value $\widehat{OV}$, and per-period counterparts are reported in Table 6. We use the approach described in Section 3, under both the auto-regressive and the latent variable models, under the added assumption that parameters $(\beta_{i,t})_{t\in\mathcal{T}_1\cup\mathcal{T}_2}$ are jointly Gaussian and that error terms $(\xi_{i,t})_{t\in\mathcal{T}}$ are also jointly Gaussian. This makes the deconvolution steps used in Section 3 much less data intensive.

	Auto-regressive model	Latent variable model
$\widehat{V}$	-6.56	-8.75**
	(4.75)	(4.44)
	[0.17]	[0.05]
$\widehat{OV}$	-2.75*	-4.94**
	(1.58)	(2.32)

	[0.08]	[0.03]
$\widehat{V}_{pP}$	-1.82	-2.43**
	(1.24)	(1.15)
	[0.14]	[0.03]
$\widehat{OV}_{pP}$	-1.06***	-1.67***
	(0.11)	(0.12)
	[0.0]	[0.0]

Notes: The table shows estimates under the assumption that the error term, ϵ_{it}, is MA(1). Bootstrap standard errors are in parentheses. Two-sided p-values are in brackets. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 6: Debiased value estimates

Estimated effects now depend on the underlying model. The impact from early treatment — good or bad — is causally persistent in the auto-regressive model, but not in the latent variable model. As a result, adaptive treatment has lower value added under the auto regressive model than under the latent variable model.

Under both models, however, option value estimates $\widehat{OV}$ and $\widehat{OV}_{pP}$ are at least marginally significant, suggesting that dynamic treatment offers a meaningful improvement over persistent treatment. Estimates $\widehat{V}_{pP}$ of treatment effect per-treated period suggests that the dynamic adoption of programs modeled on Operation ceasefire yields a reduction in the homicide rate of 12% to 15% per period of treatment. This is a meaningful improvement over ATT estimates, but remains well short of the improvements suggested by naïve value estimates.

6 Conclusion

When treatment effects are heterogeneous across units but correlated over time, it is beneficial to condition continued adoption on early signals that treatment is effective. This is especially useful if other available covariates are poor predictors of treatment effects.

While naïve estimators of the value of dynamic adoption are biased when errors in treatment effect estimates are correlated, we provide assumptions under which it is possible to obtain consistent estimates even under persistent assignment. We illustrate the methodology in the context of homicide reduction programs modeled on Operation Ceasefire. On average, those programs led to a modest statistically insignificant reduction in homicide rates, but dynamic treatment assignment significantly improves treatment effects by targeting units in which treatment appears effective. This both explains why selectively reporting outcomes from successful treatment units is not entirely misleading, while also providing a more transparent alternative: report the effect from dynamic adoption.

Appendix

A Sample Construction

The procedure aims at identifying large local police departments that reported homicide incidents consistently between 1990 and 2018. To start with, we limit the sample to the 7,641 ORIs that represent local police departments as defined in Bureau of Justice Statistics (2018). We then drop 70 ORIs with zero recorded population in at least one associated observation. Such police departments are likely not the primary police departments for their service areas.[16] For the remaining ORIs, we impute population counts for years that have no recorded homicide and thus no recorded population by backwards/forwards filling and linear interpolation. Given the population by year for each ORI, we then only retain 3,351 police departments with population larger than 10,000 more often than not between 1990-2018. Our choice for large local police departments is consistent with previous literature that studies effects from new programs on crime rates with UCR data. Examples include Evans and Owens (2007), Mello (2019), and Weisburst (2019).

We then remove ORIs with inconsistently reported homicide incidents. We begin with excluding 2,274 control ORIs that did not report at least one homicide, either intentional or unintentional, in every 60 consecutive months between 1990 and 2018. For treatment ORIs, only the two years immediate before the treatment matters for calculating counterfactuals. We thus only consider months in these two years and drop 4 treatment ORIs that did not consecutively report for 60 months prior to at least one such month. We next drop ORIs with extreme changes in the intentional homicide rates across adjacent years. More specifically, we define the set of relevant units in year t, I_t, as the union of the set of control units and the set of treatment units that initiated programs in years after year t. For year t and unit i, we then define ΔY_{it} as the change in the homicide rate between year $t-1$ and t.

[16] See Loftin et al. (2008) for a detailed discussion on the map between ORIs and their corresponding service areas.

$\Delta\bar{Y}_t$ is accordingly ΔY_{it} averaged across all units in I_t. We mark $(\Delta Y_{i,t-1}, \Delta Y_{it})$ as an outlier whenever the Mahalanobis distance between $(\Delta Y_{i,t-1}, \Delta Y_{it})$ and $(\Delta\bar{Y}_{t-1}, \Delta\bar{Y}_t)$ is greater than the 95 percentile of the chi-squared distribution. We next drop 351 control units with at least one $(\Delta Y_{i,t-1}, \Delta Y_{it})$ defined as an outlier between $t = 1992$ and $t = 2018$ and 3 treatment units with $(\Delta Y_{i,t-1}, \Delta Y_{it})$ defined as an outlier in year t immediate before treatment.

Finally, we exclude 30 ORIs that appear in our program list defined in Section 4.2 but are not treatment ORIs defined for the main results, leaving 634 control ORIs and 58 treatment ORIs in the sample.

B Proofs

Proof of Lemma 1. We need to identify

$$\mathbb{E}_\lambda \left[(\mathbf{Y}_i^1(0), \mathbf{Y}_i^2(0,0)) | D_i^1 = D_i^2 = 1 \right].$$

By Assumption 2 the joint distribution of $(\mathbf{Y}_i^0, \xi_i^1, \xi_i^2)$ conditional on $D_i^1 = D_i^2 = 1$ is identified from data. You can generate an associated distribution for $(\mathbf{Y}_i^1(0), \mathbf{Y}_i^2(0,0))$ conditional on $D_i^1 = D_i^2 = 1$ as follows.

Draw $(\mathbf{Y}_i^0, \xi_i^1, \xi_i^2)$ from its distribution conditional on $D_i^1 = D_i^2 = 1$. By iteratively applying Assumption 1, given no treatment, generate a time series sequence of outcomes $Y_{i,t}$ according to

$$Y_{i,t} = \Psi_t(Y_{i,t-1}, 0^t, \xi_{i,t}),$$

where 0^t is the vector of 0s, associated with no treatment, up to and including time t.

The simulated out $(\mathbf{Y}_i^1(0), \mathbf{Y}_i^2(0,0))$ have the required distribution. ■

Proof of Proposition 1. Given decomposition (6) and Lemma 1, we only need to show that $\mathbb{E}\left[\mathbf{Y}_i^2(1,0) | \mathbf{Y}_i^1, D_i^1 = 1\right]$ is identified from data.

By Assumption 2 the joint distribution of $(\mathbf{Y}_i^0, \xi_i^1)$ conditional on $D_i^1 = 1$ is known.

From (8) and Assumption 3 it follows that the observed distribution of $\mathbf{Y}_i^1(1)$ conditional on $D_i^1 = 1$ lets us infer the distribution of β_i^1 conditional on $D_i^1 = 1$.

Given (8), this implies that the joint distribution of $(\xi_i^1, \beta_i^1, \mathbf{Y}_i^1(1))$ is known. This implies that the conditional distribution

$$\mathsf{Prob}_\lambda(\xi_i^1, \beta_i^1 | Y_i^1(1), D_i^1)$$

is identified. In other terms, it is possible to estimate the joint distribution of phase 1 shocks and phase 1 treatment effects given outcomes and treatment status.

In turn, given that the joint distribution of (ξ_i^1, ξ_i^2) conditional on $D_i^1 = 1$ is known by Assumption 2, it follows that we can identify the conditional distribution of phase 2 shocks and phase 1 treatment effects

$$\mathsf{Prob}(\xi_i^2, \beta_i^1 | \mathbf{Y}_i^1(1), D_i^1(1)).$$

By sampling from this distribution and iteratively applying Assumption 1, given no treatment in phase 2, generate a time series sequence of outcomes $Y_{i,t}$ for $t \in \mathcal{T}_2$ according to

$$Y_{i,t} = \Psi_t(Y_{i,t-1}, [\beta_i^1, 0]^t, \xi_{i,t})$$

where the $[\beta_i^1, 0]^t$ is the vector

$$(\beta_{i,T_0+1}, \ldots, \beta_{i,T_1}, 0, \ldots, 0)$$

of length $t - T_0$.

The simulated out $\mathbf{Y}_i^2(1, 0)$ has the required distribution. ■

References

Abreu, D. and H. Matsushima (1992): "Virtual implementation in iteratively undominated strategies: complete information," *Econometrica: Journal of the Econometric Society*, 993–1008.

Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017): "What can we learn from experiments? Understanding the threats to the scalability of experimental results," *American Economic Review*, 107, 282–286.

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 58, 277–297.

Athey, S., R. Chetty, G. W. Imbens, and H. Kang (2025): "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely," *Review of Economic Studies*, rdaf087.

Athey, S. and G. W. Imbens (2017): "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic perspectives*, 31, 3–32.

Athey, S. and S. Wager (2021): "Policy learning with observational data," *Econometrica*, 89, 133–161.

Auer, P., N. Cesa-Bianchi, and P. Fischer (2002): "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, 47, 235–256.

Billingsley, P. (1995): *Probability and Measure*, Wiley, 3 ed.

Braga, A. A., D. M. Hureau, and A. V. Papachristos (2014): "Deterring gang-involved gun violence: measuring the impact of Boston's Operation Ceasefire on street gang behavior," *Journal of Quantitative Criminology*, 30, 113–139.

Braga, A. A., D. Weisburd, and B. Turchan (2019): "Focused deterrence strategies effects on crime: A systematic review," *Campbell Systematic Reviews*, 15, e1051.

Braga, A. A. and D. L. Weisburd (2012): "The Effects of Focused Deterrence Strategies on Crime: A Systematic Review and Meta-Analysis of the Empirical Evidence," *Journal of Research in Crime and Delinquency*, 49, 323–358.

Bureau of Justice Statistics (2018): "Law Enforcement Agency Identifiers Crosswalk, United States, 2012," .

Chakraborty, B. and E. E. Moodie (2013): *Statistical Methods for Dynamic Treatment Regimes*, Springer.

Chakraborty, B. and S. A. Murphy (2014): "Dynamic Treatment Regimes," *Annual Review of Statistics and Its Application*, 1, 447–464.

Chassang, S. and S. Kapon (2022): "Designing randomized controlled trials with external validity in mind," Tech. rep., National Bureau of Economic Research.

Chassang, S., G. Padró i Miquel, and E. Snowberg (2012): "Selective trials: A principal-agent approach to randomized controlled experiments," *American Economic Review*, 102, 1279–1309.

Chetty, R., N. Hendren, and L. F. Katz (2016): "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 106, 855–902.

Cook, T. D., W. R. Shadish, and V. C. Wong (2008): "Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons," *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 27, 724–750.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009): "Dealing with limited overlap in estimation of average treatment effects," *Biometrika*, 96, 187–199.

Dehejia, R., C. Pop-Eleches, and C. Samii (2021): "From local to global: External validity in a fertility natural experiment," *Journal of Business & Economic Statistics*, 39, 217–243.

Dupas, P. (2014): "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," *Econometrica*, 82, 197–228.

Evans, W. N. and E. G. Owens (2007): "COPS and Crime," *Journal of public Economics*, 91, 181–201.

Fan, J. (1991): "On the optimal rates of convergence for nonparametric deconvolution problems," *The Annals of Statistics*, 1257–1272.

Gittins, J. C. (1979): "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41, 148–164.

Haddad, P. (1998): "The SSRI discontinuation syndrome," *Journal of psychopharmacology*, 12, 305–313.

Halac, M., I. Kremer, and E. Winter (2019): "Raising capital from heterogeneous investors," *American Economic Review*.

Halac, M., E. Lipnowski, and D. Rappoport (2020): "Rank Uncertainty in Organizations," *Available at SSRN 3553935*.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73, 669–738.

Heckman, J. J. and E. J. Vytlacil (2007): "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," *Handbook of econometrics*, 6, 4779–4874.

Hendren, N. and B. Sprung-Keyser (2020): "A unified welfare analysis of government policies," *The Quarterly journal of economics*, 135, 1209–1318.

Ida, T., T. Ishihara, K. Ito, D. Kido, T. Kitagawa, S. Sakaguchi, and S. Sasaki (2024): "Dynamic targeting: Experimental evidence from energy rebate programs," Tech. rep., National Bureau of Economic Research.

Imai, K. and M. Ratkovic (2013): "Estimating treatment effect heterogeneity in randomized program evaluation," *The Annals of Applied Statistics*, 443–470.

Kapon, S., L. Del Carpio, and S. Chassang (2024): "Using divide-and-conquer to improve tax collection," *The Quarterly Journal of Economics*, 139, 2475–2523.

Kennedy, D. M. (2011): *Don't shoot: one man, a street fellowship, and the end of violence in inner-city America*, Bloomsbury Publishing USA.

Kennedy, D. M., A. A. Braga, A. M. Piehl, and E. J. Waring (2001): *Reducing Gun Violence: the Boston Gun Project's Operation Ceasefire*, US Department of Justice Office of Justice Programs.

Kennedy, D. M. and M. A. Friedrich (2014): *Custom Notifications: Individualized Communication in the Group Violence Intervention*, Office of Community Oriented Policing Services.

Lai, T. L. and H. Robbins (1985): "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, 6, 4–22.

Lei, H., I. Nahum-Shani, K. Lynch, D. Oslin, and S. A. Murphy (2012): "A" SMART" design for building individualized treatment sequences," *Annual review of clinical psychology*, 8, 21–48.

Loftin, C., D. McDowall, and M. Xie (2008): "Likely errors when linking supplementary homicide report records for large US cities," *Homicide Studies*, 12, 234–248.

Mello, S. (2019): "More COPS, less crime," *Journal of public economics*, 172, 174–200.

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619.

Muralidharan, K. and P. Niehaus (2017): "Experimentation at scale," *Journal of Economic Perspectives*, 31, 103–124.

Murphy, S. A. (2005): "An experimental design for the development of adaptive treatment strategies," *Statistics in medicine*, 24, 1455–1481.

National Network for Safe Communities (2015): *Drug Market Intervention: An Implementation Guide*, Office of Community Oriented Policing Services.

——— (2016): *Group Violence Intervention: An Implementation Guide*, Office of Community Oriented Policing Services.

Piehl, A. M., D. M. Kennedy, and A. A. Braga (2000): "Problem Solving and Youth Violence: An Evaluation of the Boston Gun Project," *American Law and Economics Review*, 2, 58–106.

Rosenzweig, M. R. and C. Udry (2020): "External validity in a stochastic world: Evidence from low-income countries," *The Review of Economic Studies*, 87, 343–381.

Wager, S. and S. Athey (2018): "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 113, 1228–1242.

Weisburst, E. K. (2019): "Safety in police numbers: Evidence of police effectiveness from federal COPS grant applications," *American Law and Economics Review*, 21, 81–109.

Winter, E. (2004): "Incentives and discrimination," *American Economic Review*, 94, 764–773.