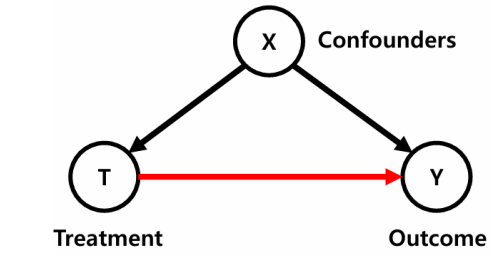
Causal Inference with Spatio-Temporal Data:

Estimating the Effects of Airstrikes on Insurgent Violence in Iraq

Table of contents

Overview on Causal Inference



Study the causal effect of a treatment on an outcome, accounting for confounding

Propensity score

$$e(x) = \Pr(T = 1 | X = x)$$

can be used as a balancing score.

Given a balancing score, treatment is no longer dependent on confounders, i.e.,

$$T \perp\!\!\!\perp X \mid e(X).$$

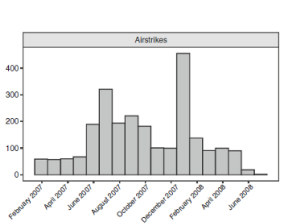
Introduction

- Causal processes may involve **spatial and temporal** dimensions.
  - Spatio-temporal settings encompass infinite number of locations and treatments.
  - Spatial spillover and Temporal carryover
- Standard Causal Inference is not ready for infinite types of treatments and rarely encompass a spatial dimension.

IDEA: Stochastic Intervention on Spatio-temporal Process

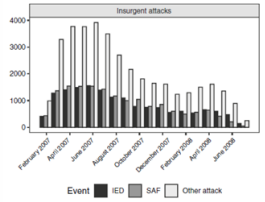
- Stochastic intervention refers to introducing probabilistic changes into a system to observe or predict how these changes influence outcomes.
- It is used in causal inference to account for uncertainty and variability in the effects of interventions; intervention is **not deterministic**.
- By modeling interventions spatio-temporally, it helps estimate causal effects of different intervention strategies(counterfactual scenarios).

Do Airstrikes Increase Insurgent Attacks?



IED
Improvised Explosive Devices

SAF
Small Arms Fire



n = 2,246

n = 68,573

Feb. 23, 2007 ~ Jul. 5, 2008 (469 timeframes)

Figure 1, modified

Motivating Illustration

- Airstrikes are stochastic **treatments**; Insurgence is an **outcome**.

Questions:

1. Effects of different airstrike strategy
   - Increase in airstrikes without change of location
   - Changes in airstrike positions without increasing frequency
2. Speed of insurgence reactions to airstrikes
   - How quickly insurgents react to airstrike strategies?

Potential Confounders; How Are Airstrikes Decided?

- ⋆ Prior patterns of insurgent violence
- ⋆ Prior air operations (as a show of force)
- ⋆ Population size, location and proximity to roads
- ⋆ Reconstruction spending

Causal Inference Framework

Setup 1

- $W_t(s)$: binary treatment variable at location $s \in \Omega$ and time $t \in \{1, 2, \cdots, T\}$
- $W_t \in \mathcal{W}$: treatment pattern at time t
- $\overline{\mathbf{W}}_t = (W_1, W_2, \cdots, W_t)$: history of treatments before t

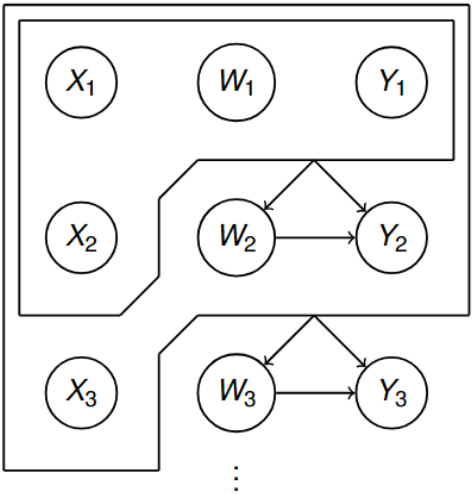
- $Y_t(\overline{\mathbf{w}}_t)$: potential outcome at time t for treatment history $\overline{\mathbf{w}}_t$
- $\overline{\mathcal{Y}}_T = \{Y_t(\overline{\mathbf{w}}_t) : \overline{\mathbf{w}}_t \in \mathcal{W}^t, t \in \{1, 2, \cdots, T\}\}$: collection of potential outcomes
- $\overline{\mathbf{Y}}_t = (Y_1, Y_2, \cdots, Y_t)$: history of outcomes before t

- $X_t(\overline{\mathbf{w}}_{t-1})$: time-varying confounder; be realized prior to W_t, but after W_{t-1}
- $\overline{\mathcal{X}}_T = \{X_t(\overline{\mathbf{w}}_{t-1}) : \overline{\mathbf{w}}_t \in \mathcal{W}^{t-1}, t \in \{1, 2, \cdots, T\}\}$
- $\overline{\mathbf{X}}_t = (X_1, \cdots, X_t)$: history of covariates before t

- $\overline{H}_t = (\overline{\mathbf{W}}_t, \overline{\mathbf{Y}}_t, \overline{\mathbf{X}}_{t+1})$: all observed history preceding the treatment at time $t+1$
- $\overline{H}_t^* = \{\overline{\mathbf{W}}_t, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}$: observed history of treatment and all of potential outcomes

Setup 2

Note that X_t can be determined by history; we just omit the arrows for simplicity



Causal Estimand 1

Stochastic Intervention specifies the probability distribution that generates the treatment under a potentially counterfactual scenario. In other words, counterfactual outcome is created by intervening the treatment.

Here stochastic interventions F_h are based on Poisson Point Process(PPP) with an intensity function $h : \Omega \to [0, \infty)$. Therefore, the expected number of outcome-active locations for a region B at time t can be defined as below:

$$N_{B,t}(F_h) := \int_{\mathcal{W}} N_B(Y_t(\overline{\mathbf{W}}_{t-1}, w_t))dF_h(w_t)$$

More generally, for M consecutive time periods, it can be extended to

$$N_{B,t}(F_h^M) := \int_{\mathcal{W}^M} N_B(Y_t(\overline{\mathbf{W}}_{t-M}, w_{t-M+1}, \cdots, w_t))dF_h(w_t)dF_h(w_{t-1})\cdots dF_h(w_{t-M+1})$$

. They'll be the predicted value for counterfactual strategy F_h(or F_h^M).

Causal Estimand 2

Average Treatment Effect(ATE) of stochastic intervention $F_{h_1}^M$ versus $F_{h_2}^M$ for a region B at time t is defined by

$$\tau_{B,t}(F_{h_1}^M, F_{h_2}^M) := N_{B,t}(F_{h_2}^M) - N_{B,t}(F_{h_1}^M)$$

. Note that it also depends on an omitted given information: $\overline{\mathbf{W}}_{t-M}$.

To summarize them over the time horizon, they introduce

$$N_B(F_h^M) = \frac{1}{T-M+1} \sum_{t=M}^{T} N_{B,t}(F_h^M)$$

$$\tau_B(F_{h_1}^M, F_{h_2}^M) = N_B(F_{h_2}^M) - N_B(F_{h_1}^M) = \frac{1}{T-M+1} \sum_{t=M}^{T} \tau_{B,t}(F_{h_1}^M, F_{h_2}^M)$$

. As M increases, it goes to be able to describe more slow-responding process.

Estimation and Inference

The Assumption

Assumption 1 (Unconfoundedness)

The treatment assignment at time t does not depend on any potential outcomes and potential confounders conditional on the observed history up to time $t-1$, i.e.,

$$f(W_t|\overline{\mathbf{W}}_{t-1}, \overline{\mathcal{Y}}_t, \overline{\mathcal{X}}_t) = f(W_t|\overline{H}_{t-1})$$

The Assumption

Assumption 1 (Unconfoundedness)

The treatment assignment at time t does not depend on any potential outcomes and potential confounders conditional on the observed history up to time $t-1$, i.e.,

$$f(W_t|\overline{\mathbf{W}}_{t-1}, \overline{\mathcal{Y}}_t, \overline{\mathcal{X}}_t) = f(W_t|\overline{H}_{t-1})$$

Assumption 2 (Bounded relative overlap)

There exists a constant $\delta_W > 0$ such that $e_t(w) > \delta_W \cdot f_h(w)$ for all $w \in \mathcal{W}$, where $e_t(w) = f(W_t = w|\overline{H}_{t-1})$ is a propensity score at time period t; It ensures that all the treatment patterns can be observed. (f_h: density function of F_h)

The Assumption

Assumption 1 (Unconfoundedness)

The treatment assignment at time t does not depend on any potential outcomes and potential confounders conditional on the observed history up to time $t-1$, i.e.,

$$f(W_t|\overline{\mathbf{W}}_{t-1}, \overline{\mathcal{Y}}_t, \overline{\mathcal{X}}_t) = f(W_t|\overline{H}_{t-1})$$

Assumption 2 (Bounded relative overlap)

There exists a constant $\delta_W > 0$ such that $e_t(w) > \delta_W \cdot f_h(w)$ for all $w \in \mathcal{W}$, where $e_t(w) = f(W_t = w|\overline{H}_{t-1})$ is a propensity score at time period t; It ensures that all the treatment patterns can be observed. (f_h: density function of F_h)

Under the two assumptions, the propensity score $e_t(\mathbf{w})$ is a balancing score. i.e.,

$$f(W_t = w|e_t(w), \overline{H}_{t-1}) = f(W_t = w|e_t(w)).$$

The Estimators 1

The first estimator is a standard **Inverse Probability Weighting(IPW)** estimator:

$$\hat{Y}_t(F_h^M;\omega) = \prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} \left[\sum_{s\in S_{Y_t}} K_b(\|\omega - s\|) \right]$$

$$\hat{N}_{B,t}(F_h^M) = \int_B \hat{Y}_t(F_h^M;\omega)d\omega$$

$$\hat{N}_B(F_h^M) = \frac{1}{T-M+1} \sum_{t=M}^{T} \hat{N}_{B,t}(F_h^M)$$

$$\hat{\tau}_B(F_{h_1}^M, F_{h_2}^M) = \hat{N}_B(F_{h_2}^M) - \hat{N}_B(F_{h_1}^M)$$

where $S_{Y_t} = \{\omega \in \Omega : Y_t(\omega) = 1\}$. Assume that the univariate kernel K is continuous and b is an appropriate bandwidth; it helps to smooth the estimator.

Unbiasedness 1

An important property of IPW estimator is **unbiasedness**:

Theorem

Our estimator $\hat{\tau}_B(F^M_{h_1}, F^M_{h_2})$ *is an unbiased estimator of* $\tau_B(F^M_{h_1}, F^M_{h_2})$*. i.e. ,*

$$\mathbb{E}[\hat{\tau}_B(F^M_{h_1}, F^M_{h_2})|\overline{H}^*_{t-M}] = \tau_B(F^M_{h_1}, F^M_{h_2}).$$

Unbiasedness 1

An important property of IPW estimator is **unbiasedness**:

Theorem

Our estimator $\hat{\tau}_B(F^M_{h_1}, F^M_{h_2})$ *is an unbiased estimator of* $\tau_B(F^M_{h_1}, F^M_{h_2})$. *i.e.* ,

$$\mathbb{E}[\hat{\tau}_B(F^M_{h_1}, F^M_{h_2})|\overline{H}^*_{t-M}] = \tau_B(F^M_{h_1}, F^M_{h_2}).$$

Proof.

Since the kernel estimator is unbiased,

$$\mathbb{E}\left[\int_B \sum_{s \in S_{Y_t}} K_b(\|\omega - s\|) d\omega\right] = N_B(Y_t).$$

Unbiasedness 2

(continued).

$$\mathbb{E}\left[\left(\prod_{j=t-M+1}^{t}\frac{f_h(W_j)}{e_j(W_j)}\right)N_B(Y_t)\,\middle|\,\overline{H}^*_{t-M}\right]$$

$$=\int\left(\prod_{j=t-M+1}^{t}\frac{f_h(w_j)}{e_j(w_j)}\right)N_B(Y_t(\overline{\mathbf{W}}_{t-M},w_{(t-M+1):t}))\prod_{j=t-M+1}^{t}f(w_j|\overline{H}^*_{t-M},W_{(t-M+1):(j-1)})dw_{(t-M+1):t}$$

$$=\int\left(\prod_{j=t-M+1}^{t}\frac{f_h(w_j)}{e_j(w_j)}\right)N_B(Y_t(\overline{\mathbf{W}}_{t-M},w_{(t-M+1):t}))\prod_{j=t-M+1}^{t}f(w_j|\overline{H}^*_{j-1})dw_{(t-M+1):t}$$

$$=\int\left(\prod_{j=t-M+1}^{t}\frac{f_h(W_j)}{e_j(W_j)}\right)N_B(Y_t(\overline{\mathbf{W}}_{t-M},w_{(t-M+1):t}))\prod_{j=t-M+1}^{t}e_j(w_j)dw_{(t-M+1):t}$$

$$=\int N_B(Y_t(\overline{\mathbf{W}}_{t-M},w_{(t-M+1):t}))\prod_{j=t-M+1}^{t}f_h(W_j)dw_{(t-M+1):t}=N_{B,t}(F_h^M)$$

Unbiasedness 3

Proof.

By the tower property of iterated expectation,

$$\mathbb{E}\left[\hat{N}_{B,t}(F_h^M)|\bar{H}_{t-M}^*\right] = N_{B,t}(F_h^M)$$

and

$$\mathbb{E}[\hat{\tau}_B(F_{h_1}^M, F_{h_2}^M)|\overline{H}_{t-M}^*] = \tau_B(F_{h_1}^M, F_{h_2}^M)$$

consequently. ■

Asymptotic Normality

Furthermore, it has asymptotic normality under some regularity conditions.

Theorem

If the bandwith $b_T \to 0$ and as $T \to \infty$, we have that

$$\sqrt{T}(\hat{N}_B(F_h^M) - N_B(F_h^M)) \xrightarrow{d} \mathcal{N}(0, \nu)$$

where ν represents the probability limit of $\frac{1}{T-M+1}\sum_{t=M}^{T} \nu_t$ as $T \to \infty$ with

$$\nu_t = Var\left(\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} N_B(Y_t) \,\middle|\, \bar{H}_{t-M}^* \right)$$

. Although the estimated propensity scores(from a correctly specified parametric model) are used, it works and yields more efficient estimates.

The Estimators 2

The second estimator is a stabilized version of Horvitz-Thompson(HT) IPW estimator; **Hájek estimator** defined by

$$\tilde{N}_B(F_h^M) = \frac{\sum_{t=M}^{T} \hat{N}_{B,t}(F_h^M)}{\sum_{t=M}^{T} \left(\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} \right)}$$

$$\tilde{\tau}_B(F_{h_1}^M, F_{h_2}^M) = \tilde{N}_B(F_{h_2}^M) - \tilde{N}_B(F_{h_1}^M)$$

. Note that the denominator is a sum of all weights in the weighted average $\hat{Y}_t(F_h^M; \omega)$.

It is also unbiased and asymptotically normal. As is well known, this estimator is more stable compared to HT estimators. The authors find that Hájek estimator outperforms the HT estimator in simulation with finite samples.

Sensitivity Analysis 1

We can weaken the unconfoundedness condition with unmeasured, time-varying confounder U_t, i.e.,

$$f(W_t|\overline{\mathbf{W}}_{t-1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T, \overline{\mathcal{U}}_T) = f(W_t|\overline{H}_{t-1}, \overline{U}_t)$$

where $\overline{\mathcal{U}}_T$ represents the collection of all potential values of U_t. In this case, the true score $e_t^*(w) = f(W_t = w|\overline{H}_{t-1}, \overline{U}_t)$ can be different to the misspecified $e_t(w)$.

The discrepancy of them is described by a value $\Gamma \geq 1$. By allowing

$$\frac{1}{\Gamma} \leq \rho_t = \frac{e_t(W_t)}{e_t^*(W_t)} \leq \Gamma$$

for $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_T) \in [\Gamma^{-1}, \Gamma]^T$, we can bound the estimator of average treatment effect. A procedure like this is called **Sensitivity Analysis**.

Sensitivity Analysis 2

Define

$$\hat{N}_B^{\boldsymbol{\rho}}(F_h) = \frac{\sum_{t=1}^{T} \rho_t \frac{f_h(W_j)}{e_j(W_j)} \int_B \sum_{s \in S_{Y_t}} K_b(\|\omega - s\|) d\omega}{\sum_{t=1}^{T} \rho_t \frac{f_h(W_j)}{e_j(W_j)}}$$

$$\hat{\tau}_B^{\boldsymbol{\rho}}(F_{h_1}, F_{h_2}) = \hat{N}_B^{\boldsymbol{\rho}}(F_{h_2}) - \hat{N}_B^{\boldsymbol{\rho}}(F_{h_1})$$

. Then We can find $\boldsymbol{\rho}_{\max}(h) = \text{argmax}_{\boldsymbol{\rho}} \hat{N}_B^{\boldsymbol{\rho}}(F_h)$ and $\boldsymbol{\rho}_{\min}(h) = \text{argmin}_{\boldsymbol{\rho}} \hat{N}_B^{\boldsymbol{\rho}}(F_h)$ via

$$\begin{cases} \text{maximize/minimize} & \sum_{t=1}^{T} \rho_t^* \frac{f_h(W_j)}{e_j(W_j)} \int_B \sum_{s \in S_{Y_t}} K_b(\|\omega - s\|) d\omega \\ \text{subject to} & \frac{\kappa}{\Gamma} \le \rho_t^* \le \Gamma\kappa, \ \sum_{t=1}^{T} \rho_t^* \frac{f_h(W_j)}{e_j(W_j)} = 1, \ \text{and } \kappa = \rho_t^*/\rho_t \ge 0 \end{cases}$$

, which is LP. Then the bounds for the causal estimates are obtained as below.

$$\hat{N}_B^{\boldsymbol{\rho}_{\min}(h_2)}(F_{h_2}) - \hat{N}_B^{\boldsymbol{\rho}_{\max}(h_1)}(F_{h_1}) \le \hat{\tau}(F_{h_1}, F_{h_2}) \le \hat{N}_B^{\boldsymbol{\rho}_{\max}(h_2)}(F_{h_2}) - \hat{N}_B^{\boldsymbol{\rho}_{\min}(h_1)}(F_{h_1})$$

Simulation & Empirical Analysis

Simulation Setup 1

- $T \in \{200, 400, 500\}$: length of time series
- Four confounders $X_t = (X_t^1, X_t^2, X_t^3, X_t^4)$; two are time-varying.
- All of confounders, treatment patterns, and outcomes are generated by latent variable D_t: minimal distance from airstrikes, country border, Baghdad, etc.
- F_h is the inhomogeneous PPP with intensity $h = c\varphi$ (φ: observed density).
- Adjusting c from 3 to 8 increases the intensity of airstrikes.
- Choosing M from $\{1, 3, 7, 30\}$ determines the intervention duration.
- Region of interest(B) is either the whole country(Ω), Baghdad administrative unit, or the town of Mosul (North of Baghdad).
- To examine lagged interventions over three time periods, define intervention strategy $F_{\mathbf{h}} = F_{h_1} \times F_{h_2} \times F_{h_3}$; $h_1 = h_2 = 5\varphi$, $h_3 = c\varphi$.

Simulation Setup 2

$X^1(\omega)$

(a)

Fixed covariate reflecting the distance from the closest road and country border

$X^3(\omega)$

(b)

Time-varying covariate reflecting the observed airstrike and attack patterns

Log Density of φ

(c)

Estimated log-density of treatment patterns

Simulation Results

$B = \Omega$

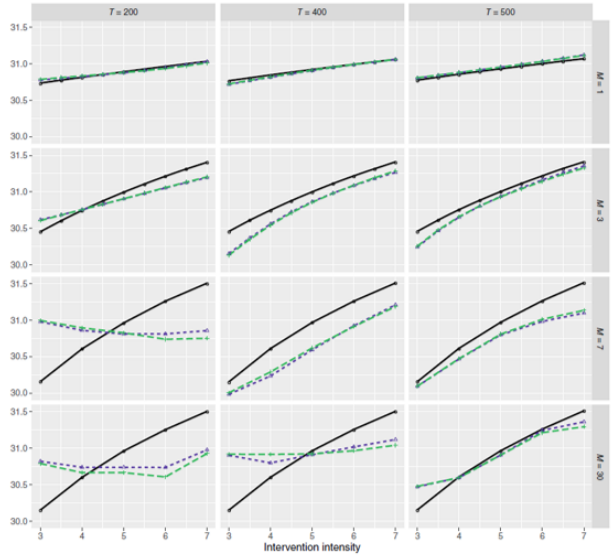
x axis: Intervention Intensity c
(Outer: $T = 200, 400, 500$)

y axis: Average Potential Outcomes
(Outer: $M = 1, 3, 7, 30$)

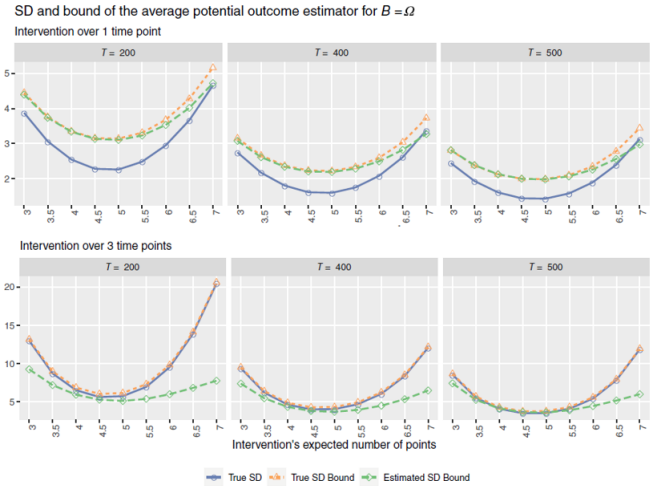
Black: Truth
Blue/Green: Predicted with
true/estimated propensity score

As M increases,
predictions get **worse**.
(Especially for small T)

Whether propensity scores are
estimated or true is **not important**.



Simulation Results 2



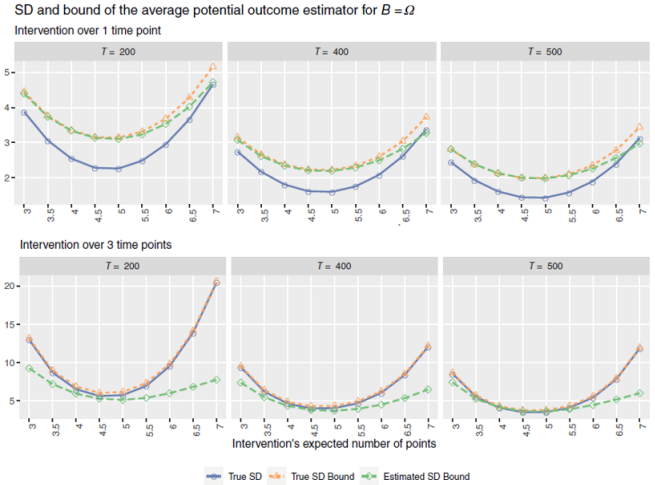
$B = \Omega$

x axis: Intervention's expected number of points $\hat{c}$
(Outer: $T = 200, 400, 500$)

y axis: SD or SD Bound (Outer: $M = 1, 3$)

Blue: True SD
Orange/Green: True/Estimated Bound

Simulation Results 2



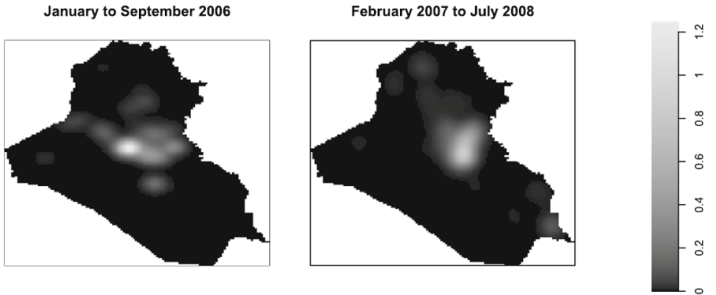
In the low uncertainty scenarios (small M or $\hat{c} \approx 5$),

- true variance bound is larger than theoretical variance.
- estimated variance bound follows true variance bound.

In the high uncertainty scenarios (large M or $\|\hat{c} - 5\| \gg 1$),

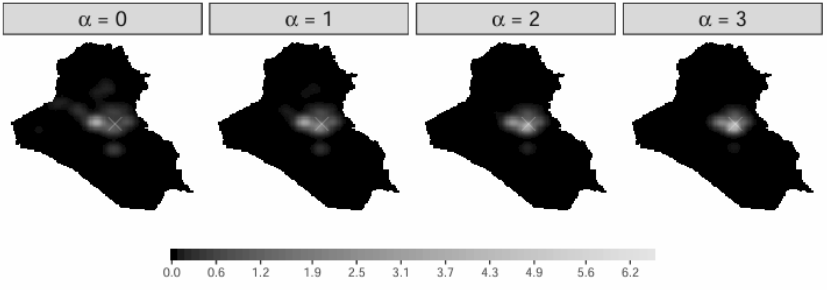
- estimated variance bound underestimates true variance bound.

Stochastic Intervention



- Intensity is modeled by $h = c\varphi_0$, where φ_0 is a baseline airstrike density(left).
- An increase in c reflects the increase in airstrike density.

Stochastic Intervention



- To see the effect of change in focal location, $h_\alpha = c\varphi_0 d_\alpha$ is introduced.
- $d_\alpha(\omega) = d(\omega)^\alpha / \int_\Omega d(\omega)^\alpha dw$, where s_f is a mode and $d(\omega) \propto d(\omega, s_f)^{-1}$.
- precision parameter $\alpha \uparrow \Rightarrow$ airstrikes are concentrated on Baghdad

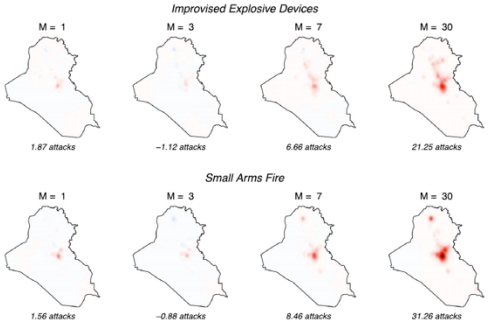
Propensity Score Model

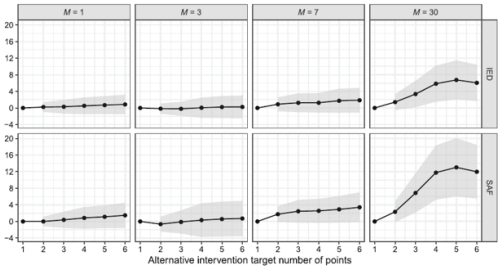
- The propensity score model is a inhomogeneous PPP with intensity

$$\lambda_t(\omega) = \exp\{\boldsymbol{\beta}^\top X_t(\omega)\}$$

  where X_t includes an intercept, temporal splines, and 32 spatial surfaces including all the covariates.
  - 6 surfaces: airstrikes and insurgence during the last day, week, and month
  - 3 surfaces: shows-of-force during the last day, week, and month
  - 1 surface: amount of aid spent past month
  - 4 surfaces: distance from cities, roads, rivers, and populations
  - 18 surfaces: seperate predictors for distances from local settlements
- Note that the intervention strategy F_h is not influenced by these covariates for simplicity; we don't consider adapted(and complex) strategy.

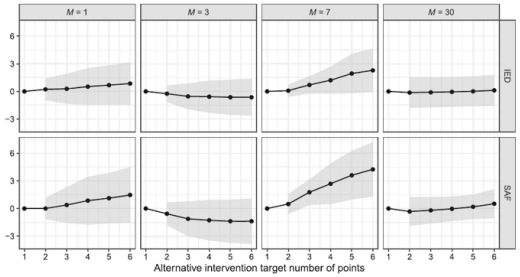
Empirical Results 1





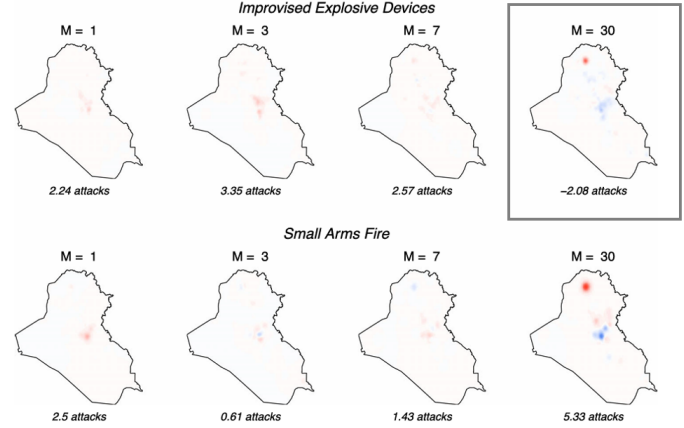
- Increased bombings for a long period($M \geq 7$) increases insurgent violence.

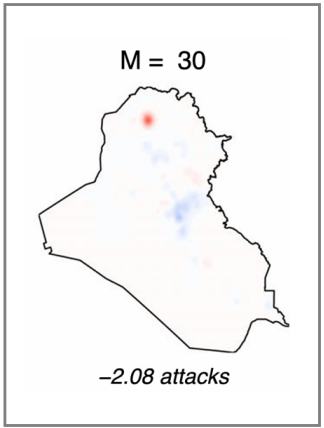
Empirical Results 2



- When the increasing of airstrikes happened M days ago, all estimated lagged effects for $M = 3$ are negative, whereas the estimated lagged effects for $M = 7$ are positive.
- It suggests that short interventions may reduce insurgence, while prolonged airstrikes increase them.

Empirical Results 3





- After prolonged attacks in other locations(instead of Baghdad), insurgents may shift their attacks to other areas especially for SAFs.

Conclusion

- The paper provides a framework for causal inference with treatments and outcomes following a spatio-temporal point process.
- Modelling treatment as a stochastic process provides flexibility.
- Methodology will be successful in cases where:
  - Undetected spillover frustrates traditional causal inference.
  - Treatments follow the overlap assumption.
  - Propensity model is correctly specified.
- However, this study does not address the effects of complex airstrike strategies.
- Additionally, in scenarios with high uncertainty, there were cases where the variance bound of the estimators was not properly estimated.
- The sensitivity analysis of the empirical study shows $\Gamma = 1.12$; the result is highly vulnerable to unmeasured confounders.

References

[1] **Papadogeorgou, G., Imai, K., Lyall, J., & Li, F.**
Causal inference with spatio-temporal data: estimating the effects of airstrikes on insurgent violence in Iraq.
Journal of the Royal Statistical Society Series B: Statistical Methodology 84.5 (2022): 1969-1999.