Spatiotemporal causal inference with arbitrary spillover and carryover effects

Mitsuru Mukaigawara, MD, MPP

Harvard University

August 6, 2025

Joint work with Kosuke Imai, Jason Lyall,
and Georgia Papadogeorgou

Motivation

- Increasing availability of unstructured data in social sciences
  - don't come in a nice matrix form $\leftrightsquigarrow$ survey, official statistics
  - text, images, audio, video, etc.
- How should we draw causal inference from these new types of data?
- Causal inference with spatio-temporal data
  - a time series of maps as data
  - treatment and outcome event locations in a continuous space
  - applications: crime, disease, disasters, pollution, etc.
- Methodological challenges
  1. spillover effects over space
  2. carryover effects over time
  3. infinitely many possible treatment and outcome locations
- Current practice
  1. arbitrary discretization of space
  2. strong assumptions about spillover and carryover effects

Contributions

- Causal inference with spatio-temporal data
  - impossible to estimate causal effects of each treatment event
    - unrestricted spillover and carryover effects
    - probability of each treatment realization is zero $\rightsquigarrow$ lack of overlap
  - stochastic intervention based on the distribution of treatments
- Causal estimands under stochastic intervention
  - expected number of outcome events within a region of interest
  - various stochastic interventions
    1. change the dosage while keeping the distribution identical
    2. change the distribution while keeping the dosage constant
    3. intervention over multiple time periods
- The proposed methodology can estimate:
  - average treatment effects
  - heterogeneous treatment effects
  - causal mediation effects
- Empirical application: airstrikes and insurgent violence in Iraq

Impacts of Airstrikes on Insurgent Violence in Iraq

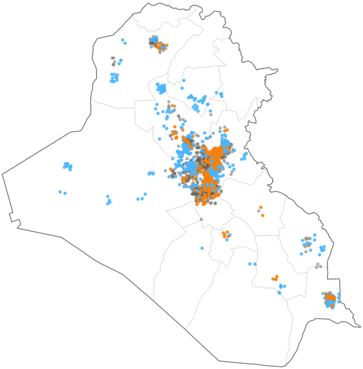
- Airstrikes as a principal tool for combating insurgency in civil wars

- Three ongoing debates:
  1. overall effectiveness: do airstrikes reduce subsequent insurgent attacks?
  2. heterogeneous effects: what factors moderate effects of airstrikes?
  3. causal mechanisms: does civilian casualty mediate effects of airstrikes?

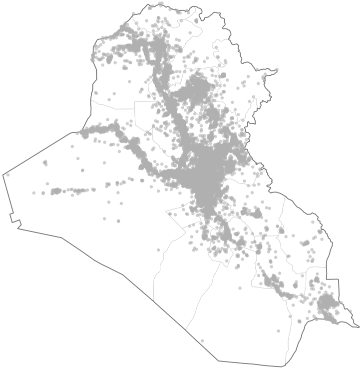
- American air campaign in Iraq:
  - declassified USAF data from Feb. 2007 to July 2008 ("surge" period)
  - date and precise geolocation for
    - airstrikes: aircraft type, number and type of bombs
    - insurgent attacks: small arms fire, improvised explosive devices
  - location of US and UK military units
    - weekly, district-level
    - troop density: soldiers per 1,000 residents
    - troop type: US Marines, US Army, and UK Army

Data: Airstrikes, Insurgent Attacks, Civilian Casualties

Airstrikes

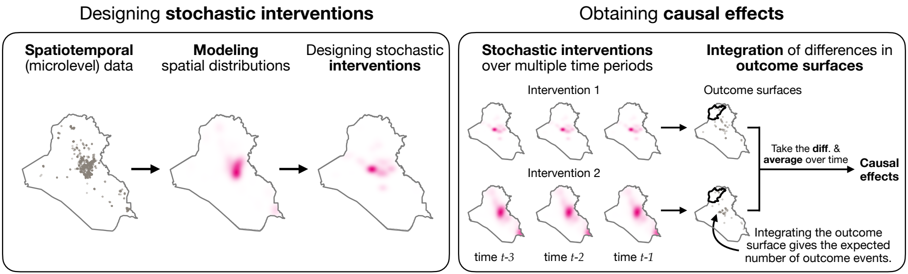


Insurgent attacks



- civilian casualty data
  - Total: 151,000 (Iraq family health survey), 200,000 (Iraq Body Count)
  - Use of satellite imagery to classify targets of airstrikes
    - civilian: residential compounds and settlements
    - non-civilian: other buildings, farms, roads, unpopulated areas, others

Methodological Overview



1. Model treatment assignment mechanism
2. Design stochastic interventions of interest
3. Estimate the counterfactual outcomes and average them over time

The Setup

- T time periods: $t = 1, 2, \ldots, T$
- Treatment variable
  - Ω: set of all possibly infinite locations that can receive the treatment
  - $W_t(s) \in \{0, 1\}$: binary treatment indicator for location s at time t
  - $W_t = \{W_t(s) : s \in \Omega\} \in \mathcal{W}$: treatment location map at time t
  - $S_{W_t} = \{s \in \Omega : W_t(s) = 1\}$: set of treatment-active locations at time t
  - $\overline{\boldsymbol{W}}_t = (W_1, W_2, \ldots, W_t)$: observed treatment history up to time t
- Outcome variable
  - $Y_t(s)$, Y_t, and $\overline{\boldsymbol{Y}}_t$ can be similarly defined
  - Potential outcome: $Y_t(\overline{\boldsymbol{w}}_t)$ where $w_t \in \mathcal{W}$ is a realized treatment and $\overline{\boldsymbol{w}}_t = (w_1, w_2, \ldots, w_t) \in \mathcal{W}^t$ is a treatment history realization at time t
  - Observed outcome: $Y_t = Y_t(\overline{\boldsymbol{W}}_t)$
  - $S_{Y_t(\overline{\boldsymbol{w}}_t)}$: set of outcome-active locations under treatment history $\overline{\boldsymbol{w}}_t$
  - History of all potential outcomes up to time t: $\overline{\mathcal{Y}}_t = \{Y_{t'}(\overline{\boldsymbol{w}}_{t'}) : \overline{\boldsymbol{w}}_{t'} \in \mathcal{W}^{t'}, t' \leq t\}$
- Time-varying confounders: X_t, $\overline{\boldsymbol{X}}_t$, $X_t(\overline{\boldsymbol{w}}_{t-1})$, and $\overline{\mathcal{X}}_t$

Causal Estimands

- Stochastic intervention: any distribution of treatment can be used
- We consider Poisson point process F_h with intensity function h
- Expected number of outcome-active locations in region B at time t under stochastic intervention F_h conducted at time t

$$\overline{N}_{Bt}(F_h) \;=\; \int_{\mathcal{W}} N_B(Y_t(\overline{\boldsymbol{W}}_{t-1}, w_t))dF_h(w_t)$$

- Further average this quantity over time:

$$\overline{N}_B(F_h) \;=\; \frac{1}{T}\sum_{t=1}^{T}\overline{N}_{Bt}(F_h)$$

- We can compare the different interventions:

$$\tau_B(F_{h'}, F_h) \;=\; \overline{N}_B(F_{h'}) - \overline{N}_B(F_h)$$

Stochastic Intervention over Multiple Time Periods

- Consider a non-dynamic stochastic intervention over L time periods

$$F_{\boldsymbol{h}} \;=\; F_{h_1} \times \cdots \times F_{h_L} \quad \text{where } \boldsymbol{h} = (h_1, h_2, \ldots, h_L)$$

- Expected number of outcome-active locations in region B at time t under stochastic intervention $F_{\boldsymbol{h}}$ conducted from time $t - L + 1$ to t

$$\overline{N}_{Bt}(F_{\boldsymbol{h}}) \;=\; \int_{\mathcal{W}} \cdots \int_{\mathcal{W}} N_B(Y_t(\overline{\boldsymbol{W}}_{t-L}, w_{t-L+1}, \ldots, w_t)) \\ dF_{h_L}(w_{t-L+1}) \cdots dF_{h_1}(w_t)$$

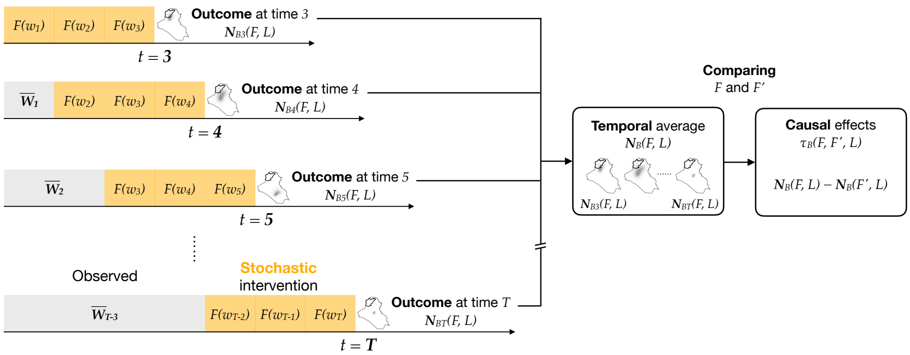
- Average this quantity over time:

$$\overline{N}_B(F_{\boldsymbol{h}}) \;=\; \frac{1}{T - L + 1} \sum_{t=L}^{T} \overline{N}_{Bt}(F_{\boldsymbol{h}})$$

- Comparison of different interventions:

$$\tau_B(F_{\boldsymbol{h}'}, F_{\boldsymbol{h}}) \;=\; \overline{N}_B(F_{\boldsymbol{h}'}) - \overline{N}_B(F_{\boldsymbol{h}})$$

Recap



- Each counterfactual outcome is conditional on the past
- Averaging is done over time
- Inference is done by letting T go infinity
- Example of causal inference with time series

Assumptions

1. Unconfoundedness: treatment is independent of all potential (past and future) paths for the outcome and time-varying confounders conditional on the observed history

Assumptions

1. Unconfoundedness: treatment is independent of all potential (past and future) paths for the outcome and time-varying confounders conditional on the observed history

$$f(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t, \{\overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}) = f(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)$$

Assumptions

1. Unconfoundedness: treatment is independent of all potential (past and future) paths for the outcome and time-varying confounders conditional on the observed history

$$f(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t, \{\overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}) = f(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)$$

2. Overlap: there exists a constant $\delta_W > 0$ such that

$$\underbrace{f(W_t = w \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)}_{propensity\ \ score} > \delta_W \cdot \underbrace{f_h(w)}_{density\ \ of\ \ F_h} \quad \text{for all } w \in \mathcal{W}$$

Assumptions

1. Unconfoundedness: treatment is independent of all potential (past and future) paths for the outcome and time-varying confounders conditional on the observed history

$$f(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t, \{\overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}) \;=\; f(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)$$

2. Overlap: there exists a constant $\delta_W > 0$ such that

$$\underbrace{f(W_t = w \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)}_{propensity\ score} \;>\; \delta_W \cdot \underbrace{f_h(w)}_{density\ of\ F_h} \qquad \text{for all } w \in \mathcal{W}$$

$\rightsquigarrow$ the ratio $f_h(w)/f(W_t = w \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)$ is bounded

The Proposed Estimator

- Inverse probability of treatment weighting (IPW)
- Kernel smoothing of spatial point patterns
- Estimated outcome surface at $\omega \in \Omega$ under the intervention F_h

$$\widehat{Y}_t(F_h;\omega) = \underbrace{\frac{\overbrace{f_h(W_t)}^{\textit{counterfactual distribution}}}{\hat{f}(W_t \mid \overline{\boldsymbol{W}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}, \overline{\boldsymbol{X}}_t)}}_{\textit{actual distribution}} \underbrace{\sum_{s \in S_{Y_t}} K_b(\|\omega - s\|)}_{\textit{spatially smoothed outcome}}$$

where K_b is the scaled Kernel function with bandwidth parameter b

- Estimated number of outcome-active locations in region B

$$\widehat{\widehat{N}}_{Bt}(F_h) = \int_B \widehat{Y}_t(F_h;\omega) d\omega$$

- Averaging over time

$$\widehat{\widehat{N}}_B(F_h) = \frac{1}{T}\sum_{t=1}^{T} \widehat{\widehat{N}}_{Bt}(F_h)$$

Estimation for Intervention over Multiple Time Periods L

- Estimated outcome surface at $\omega \in \Omega$

$$\widehat{Y}_t(F_{\boldsymbol{h}}; \omega) = \underbrace{\prod_{j=t-L+1}^{t} \frac{f_{h_{t-j+1}}(W_j)}{\hat{f}(W_j \mid \overline{\boldsymbol{W}}_{j-1}, \overline{\boldsymbol{Y}}_{j-1}, \overline{\boldsymbol{X}}_j)}}_{product\ of\ L\ ratios} \sum_{s \in S_{Y_t}} K_b(\|\omega - s\|)$$

- Estimated number of outcome-active locations in region B

$$\widehat{\overline{N}}_{Bt}(F_{\boldsymbol{h}}) = \int_B \widehat{Y}_t(F_{\boldsymbol{h}}; \omega) d\omega$$

- Averaging over time

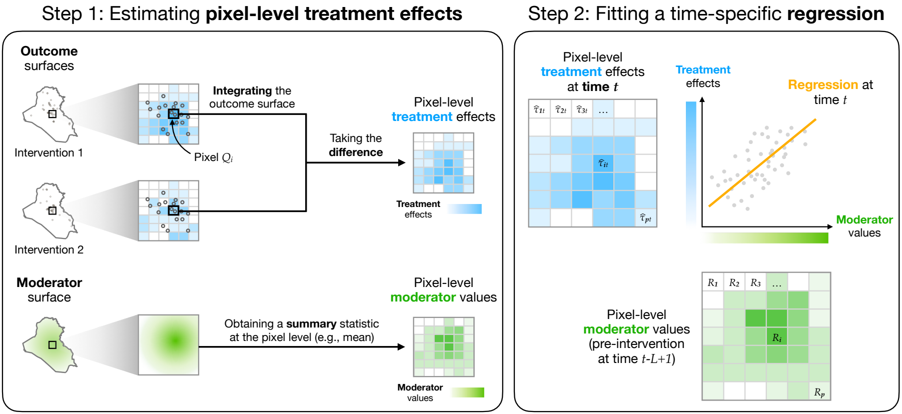
$$\widehat{\overline{N}}_B(F_{\boldsymbol{h}}) = \frac{1}{T - L + 1} \sum_{t=L}^{T} \widehat{\overline{N}}_{Bt}(F_{\boldsymbol{h}})$$

- Asymptotic normality

$$\sqrt{T}\left(\widehat{\overline{N}}_B(F_{\boldsymbol{h}}) - \overline{N}_B(F_{\boldsymbol{h}})\right) \xrightarrow{d} \mathcal{N}(0, v)$$

- Hájek estimator (normalized weights) for efficiency

Heterogeneous Treatment Effects



- Separate regression for each time period
- Inference by averaging coefficients over time

Causal Mechanisms

- Mediator
  - mediator at location $s \in \Omega$: $M_t(s) \in \mathcal{M}$
  - collection of mediator values: $\{M_t(s), s \in \Omega\}$
  - mediator history: $\boldsymbol{M}_t = (M_1, M_2, \ldots, M_t)$
  - potential values: $M_t(\overline{\boldsymbol{w}}_t, \overline{\boldsymbol{m}}_{t-1})$
- Potential outcome: $Y_t(\overline{\boldsymbol{w}}_t, \overline{m}_t)$
- Time-varying covariates: $\boldsymbol{X}_t(\overline{\boldsymbol{w}}_t, \overline{m}_t)$

- Stochastic intervention: $F = (F_W(w), F_{M|w}(m))$
  - F_W: intervention distribution for W
  - $F_{M|w}$: intervention distribution for M given $W = w$
- Causal estimands

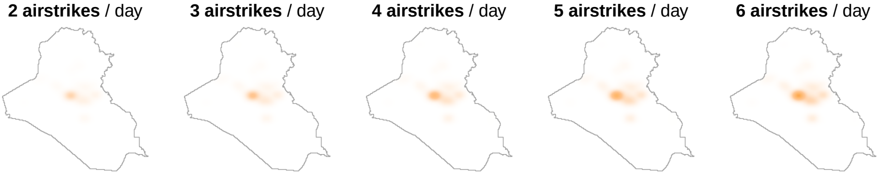
$$\underbrace{\tau_B(F', F'')}_{\text{total effect}} = \underbrace{\tau_B^{\text{IE}}(F'_{M|w}, F''_{M|w}; F'_W)}_{\text{indirect effect}} + \underbrace{\tau_B^{\text{DE}}(F'_W, F''_W; F'_{M|w})}_{\text{direct effect}}$$

Empirical Analysis: Setup

- Estimate the baseline treatment distribution f_0
  - inhomogeneous Poisson process regression
  - 2006 data, separate from the 2007 evaluation data
  - covariates: aid, histories of air strikes, show of force, and insurgent attacks (1, 7, and 30 days), log population, time splines, distances from rivers, major roads, cities, and settlements

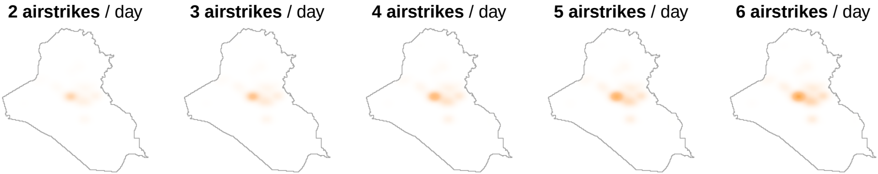
- Questions:
  1. How does increasing airstrikes affect insurgent violence?
     $\rightsquigarrow$ vary $c > 0$ for $h(\omega) = c \cdot f_0(\omega)$
  2. How does the shift in the prioritization of certain locations for airstrikes change the spatial pattern of insurgent attacks?
     $\rightsquigarrow$ vary $\alpha > 0$ for $h_\alpha(\omega) \propto f_0(\omega) d_\alpha(\omega)$ with $\int_\Omega h_\alpha(\omega) d\omega = c$
     - power density $d_\alpha(\omega) \propto d(\omega)^\alpha$
     - $d(\omega) =$ the normal density centered at s_f with precision α

Intervention by Picture

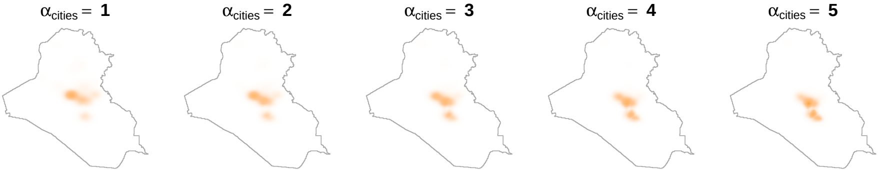


(a) Counterfactual interventions with intensified airstrikes

Intervention by Picture

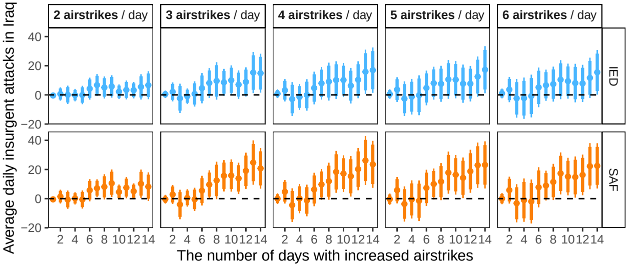


(a) Counterfactual interventions with intensified airstrikes

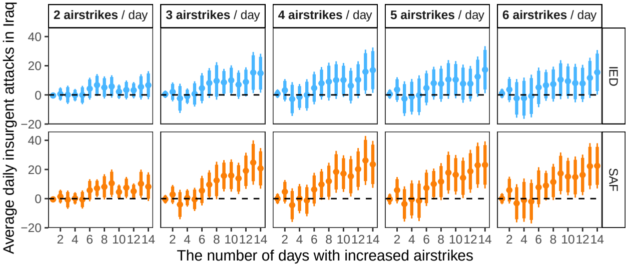


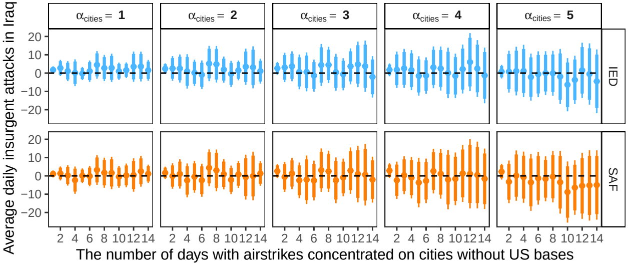
(b) Counterfactual interventions with location shifts

Increasing the Expected Number of Airstrikes from 1 to 6 per Day Leads to More Insurgent Attacks with Large *L*



Increasing the Expected Number of Airstrikes from 1 to 6 per Day Leads to More Insurgent Attacks with Large *L*





Heterogeneous Treatment Effects

- Hypothesis: airstrikes may further increase insurgent violence where US and UK armed forces are present
- Survey evidence: strong resentment against foreign forces
- Armored vehicles are clear targets of insurgents

Heterogeneous Treatment Effects

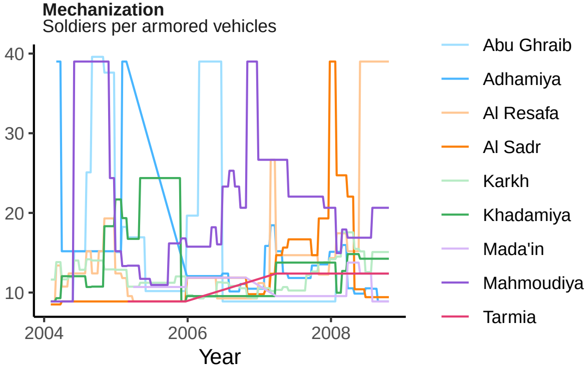
- Hypothesis: airstrikes may further increase insurgent violence where US and UK armed forces are present
- Survey evidence: strong resentment against foreign forces
- Armored vehicles are clear targets of insurgents

Heterogeneous Treatment Effects

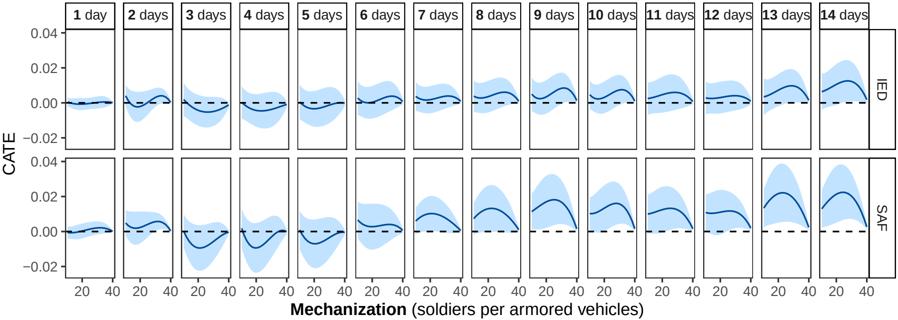
- Hypothesis: airstrikes may further increase insurgent violence where US and UK armed forces are present
- Survey evidence: strong resentment against foreign forces
- Armored vehicles are clear targets of insurgents

Heterogeneous Treatment Effects

- Hypothesis: airstrikes may further increase insurgent violence where US and UK armed forces are present
- Survey evidence: strong resentment against foreign forces
- Armored vehicles are clear targets of insurgents



CATE Is Positively Associated with Mechanization



Civilian Casualty as a Causal Mechanism

- Airstrikes $\longrightarrow$ civilian casualty $\longrightarrow$ insurgent's response
- Stochastic intervention:
  - F_W: same as before
  - $F_{M|w}$: adjust for population density, distances from roads, cities, residential buildings, and settlements as key covariates

$$F_{M|w} = \frac{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t) + 1 - \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}$$

Civilian Casualty as a Causal Mechanism

- Airstrikes $\longrightarrow$ civilian casualty $\longrightarrow$ insurgent's response
- Stochastic intervention:
  - F_W: same as before
  - $F_{M|w}$: adjust for population density, distances from roads, cities, residential buildings, and settlements as key covariates

$$F_{M|w} = \frac{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t) + 1 - \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}$$

Civilian Casualty as a Causal Mechanism

- Airstrikes $\longrightarrow$ civilian casualty $\longrightarrow$ insurgent's response
- Stochastic intervention:
  - F_W: same as before
  - $F_{M|w}$: adjust for population density, distances from roads, cities, residential buildings, and settlements as key covariates

$$F_{M|w} = \frac{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t) + 1 - \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}$$

Civilian Casualty as a Causal Mechanism

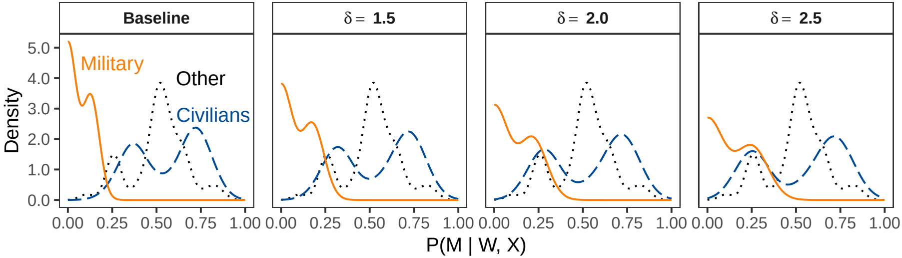
- Airstrikes $\longrightarrow$ civilian casualty $\longrightarrow$ insurgent's response
- Stochastic intervention:
  - F_W: same as before
  - $F_{M|w}$: adjust for population density, distances from roads, cities, residential buildings, and settlements as key covariates

$$F_{M|w} = \frac{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t) + 1 - \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}$$

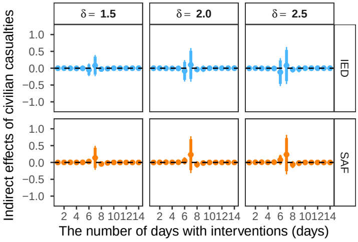
Civilian Casualty as a Causal Mechanism

- Airstrikes $\longrightarrow$ civilian casualty $\longrightarrow$ insurgent's response
- Stochastic intervention:
  - F_W: same as before
  - $F_{M|w}$: adjust for population density, distances from roads, cities, residential buildings, and settlements as key covariates

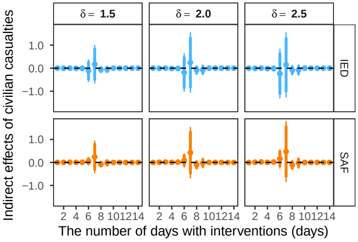
$$F_{M|w} = \frac{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}{\delta \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t) + 1 - \Pr(M_t = m_t \mid W_t, \boldsymbol{X}_t)}$$



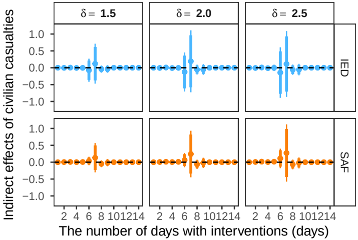
Civilian Casualty Does Not Mediate the Effects of Airstrikes



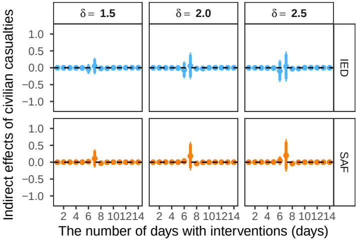
(a) Baghdad Governorate



(b) Entire Iraq



(c) Rural Iraq



(d) Urban Iraq

Concluding Remarks

- A new approach to causal inference with spatio-temporal data
  - directly model point patterns without arbitrary aggregation
  - allow for unstructured spillover and carryover effects
- Key idea: stochastic intervention
  - consider treatment distributions rather than fixed treatment values
  - can handle infinitely many possible treatment locations
- Three methods
  1. average treatment effects
  2. heterogeneous treatment effects
  3. causal mediation effects
- R package: geocausal available at CRAN
- Paper at https://imai.fas.harvard.edu/research/spatiotempo.html