UPPSALA UNIVERSITET

Inferring Heterogeneous Causal Effects in Presence of Spatial Confounding

ICML, 2019

Muhammad Osama, Dave Zachariah, Thomas B. Schön

Division of System and Control,
Department of Information Technology,
Uppsala University

muhammad.osama@it.uu.se

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable

muhammad.osama@it.uu.se

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable
- $\mathcal{D}_n = \{y_i, z_i, \boldsymbol{s}_i\}_{i=1}^n$, where $\boldsymbol{s} \in \mathbb{R}^d$ is spatial location

muhammad.osama@it.uu.se

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable
- $\mathcal{D}_n = \{y_i, z_i, \boldsymbol{s}_i\}_{i=1}^n$, where $\boldsymbol{s} \in \mathbb{R}^d$ is spatial location
- Target quantity: Average effect of assigning $z = \widetilde{z}$ on y at location $\boldsymbol{s}$

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable
- $\mathcal{D}_n = \{y_i, z_i, \boldsymbol{s}_i\}_{i=1}^n$, where $\boldsymbol{s} \in \mathbb{R}^d$ is spatial location
- Target quantity: Average effect of assigning $z = \widetilde{z}$ on y at location $\boldsymbol{s}$

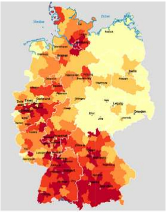
$$\tau = \frac{d}{d\widetilde{z}} \mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] \quad (1)$$

muhammad.osama@it.uu.se

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable
- $\mathcal{D}_n = \{y_i, z_i, \boldsymbol{s}_i\}_{i=1}^{n}$, where $\boldsymbol{s} \in \mathbb{R}^d$ is spatial location
- Target quantity: Average effect of assigning $z = \tilde{z}$ on y at location $\boldsymbol{s}$

$$\tau = \frac{d}{d\tilde{z}} \mathbb{E}\left[y(\tilde{z}) \mid \boldsymbol{s}\right] \tag{1}$$



$y = $ income

$z = $ age

muhammad.osama@it.uu.se

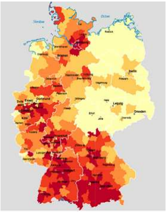
UPPSALA UNIVERSITET

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable
- $\mathcal{D}_n = \{y_i, z_i, \boldsymbol{s}_i\}_{i=1}^n$, where $\boldsymbol{s} \in \mathbb{R}^d$ is spatial location
- Target quantity: Average effect of assigning $z = \tilde{z}$ on y at location $\boldsymbol{s}$

$$\tau = \frac{d}{d\tilde{z}} \mathbb{E}\left[y(\tilde{z}) \mid \boldsymbol{s}\right] \tag{1}$$

- c: Unobserved confounding variables



$y = $ income

$z = $ age

muhammad.osama@it.uu.se

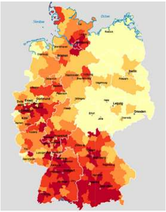
UPPSALA UNIVERSITET

Causal inference problem

- $y \in \mathbb{R}$: Outcome of interest
- $z \in \mathbb{R}$: Exposure variable
- $\mathcal{D}_n = \{y_i, z_i, \boldsymbol{s}_i\}_{i=1}^n$, where $\boldsymbol{s} \in \mathbb{R}^d$ is spatial location
- Target quantity: Average effect of assigning $z = \tilde{z}$ on y at location $\boldsymbol{s}$

$$\tau = \frac{d}{d\tilde{z}} \mathbb{E}\left[y(\tilde{z}) \mid \boldsymbol{s} \right] \tag{1}$$

- c: Unobserved confounding variables



y = income

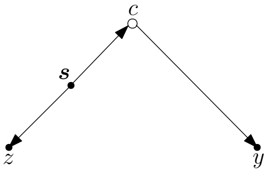
z = age

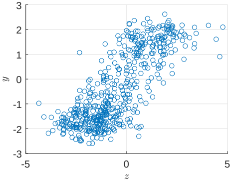
c = unemployment

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Causal Inference Problem





Example: Here $\tau = 0$ yet $\mathrm{Cov}(z, y) \neq 0$.

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Approach

- Assumptions:

muhammad.osama@it.uu.se

Approach

- Assumptions:
- $\mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] = \mathbb{E}\left[y|z = \widetilde{z},\ \boldsymbol{s}\right]$

muhammad.osama@it.uu.se

Approach

- Assumptions:
- $\mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] \;=\; \mathbb{E}\left[y|z=\widetilde{z},\; \boldsymbol{s}\right]$
- $\mathbb{E}\left[y|z=\widetilde{z},\; \boldsymbol{s}\right]$ is affine in z

muhammad.osama@it.uu.se

Approach

- Assumptions:
- $\mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] \;=\; \mathbb{E}\left[y|z = \widetilde{z},\; \boldsymbol{s}\right]$
- $\mathbb{E}\left[y|z = \widetilde{z},\; \boldsymbol{s}\right]$ is affine in z

$$y = \tau(\boldsymbol{s})z + \beta(\boldsymbol{s}) + \epsilon \tag{2}$$

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Approach

- Assumptions:
- $\mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] \;=\; \mathbb{E}\left[y|z=\widetilde{z},\;\boldsymbol{s}\right]$
- $\mathbb{E}\left[y|z=\widetilde{z},\;\boldsymbol{s}\right]$ is affine in z

$$y = \tau(\boldsymbol{s})z + \beta(\boldsymbol{s}) + \epsilon \tag{2}$$

- $\beta(\boldsymbol{s})$ is a nuisance function correlated with spatially varying exposure z.

muhammad.osama@it.uu.se

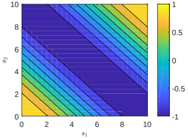
UPPSALA
UNIVERSITET

Approach

- Assumptions:
- $\mathbb{E}\left[y(\tilde{z}) \mid \boldsymbol{s}\right] \;=\; \mathbb{E}\left[y|z=\tilde{z},\; \boldsymbol{s}\right]$
- $\mathbb{E}\left[y|z=\tilde{z},\; \boldsymbol{s}\right]$ is affine in z

$$y = \tau(\boldsymbol{s})z + \beta(\boldsymbol{s}) + \epsilon \tag{2}$$

- $\beta(\boldsymbol{s})$ is a nuisance function correlated with spatially varying exposure z.



$\tau(\boldsymbol{s})$

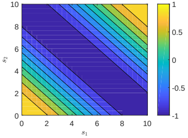
muhammad.osama@it.uu.se

Approach

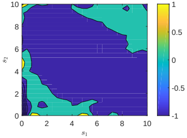
- Assumptions:
- $\mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] \;=\; \mathbb{E}\left[y | z = \widetilde{z},\; \boldsymbol{s}\right]$
- $\mathbb{E}\left[y | z = \widetilde{z},\; \boldsymbol{s}\right]$ is affine in z

$$y = \tau(\boldsymbol{s})z + \beta(\boldsymbol{s}) + \epsilon \tag{2}$$

- $\beta(\boldsymbol{s})$ is a nuisance function correlated with spatially varying exposure z.



$\tau(\boldsymbol{s})$



$\widehat{\tau}(\boldsymbol{s})$ via (2)

muhammad.osama@it.uu.se

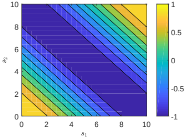
UPPSALA
UNIVERSITET

Approach

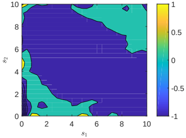
- Assumptions:
- $\mathbb{E}\left[y(\widetilde{z}) \mid \boldsymbol{s}\right] \;=\; \mathbb{E}\left[y|z=\widetilde{z},\ \boldsymbol{s}\right]$
- $\mathbb{E}\left[y|z=\widetilde{z},\ \boldsymbol{s}\right]$ is affine in z

$$y = \tau(\boldsymbol{s})z + \beta(\boldsymbol{s}) + \epsilon \tag{2}$$

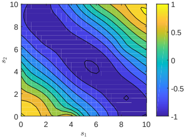
- $\beta(\boldsymbol{s})$ is a nuisance function correlated with spatially varying exposure z.



$\tau(\boldsymbol{s})$



$\widehat{\tau}(\boldsymbol{s})$ via (2)



$\widehat{\tau}(\boldsymbol{s})$ proposed

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Error-in-variables model

- Let $w = y - \mathbb{E}\left[y|\boldsymbol{s}\right]$ and $v = z - \mathbb{E}\left[z|\boldsymbol{s}\right]$ [1]

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Error-in-variables model

- Let $w = y - \mathbb{E}\left[y|\boldsymbol{s}\right]$ and $v = z - \mathbb{E}\left[z|\boldsymbol{s}\right]$ [1]
- (2) becomes

$$w = \tau(\boldsymbol{s})v + \epsilon \tag{3}$$

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Error-in-variables model

- Let $w = y - \mathbb{E}\left[y|\boldsymbol{s}\right]$ and $v = z - \mathbb{E}\left[z|\boldsymbol{s}\right]$ [1]
- (2) becomes

$$w = \tau(\boldsymbol{s})v + \epsilon \tag{3}$$

- The effect $\tau(\boldsymbol{s})$ is directly identifiable from (3) which we parameterize as

$$\tau_{\boldsymbol{\theta}}(\boldsymbol{s}) \in \{f(\boldsymbol{s}) : f = \boldsymbol{\phi}(\boldsymbol{s})^\top \boldsymbol{\theta}\},$$

muhammad.osama@it.uu.se

Error-in-variables model

- Let $w = y - \mathbb{E}\left[y|\boldsymbol{s}\right]$ and $v = z - \mathbb{E}\left[z|\boldsymbol{s}\right]$ [1]
- (2) becomes

$$w = \tau(\boldsymbol{s})v + \epsilon \tag{3}$$

- The effect $\tau(\boldsymbol{s})$ is directly identifiable from (3) which we parameterize as

$$\tau_{\boldsymbol{\theta}}(\boldsymbol{s}) \in \{f(\boldsymbol{s}) : f = \boldsymbol{\phi}(\boldsymbol{s})^\top \boldsymbol{\theta}\},$$

- Residuals w and v are not observed but estimated so that

$$w = \underbrace{\left(y - \widehat{\mathbb{E}[y|\boldsymbol{s}]}\right)}_{\widehat{w}} + \underbrace{\left(\widehat{\mathbb{E}[y|\boldsymbol{s}]} - \mathbb{E}[y|\boldsymbol{s}]\right)}_{\widetilde{w}},$$

$$v = \underbrace{\left(z - \widehat{\mathbb{E}[z|\boldsymbol{s}]}\right)}_{\widehat{v}} + \underbrace{\left(\widehat{\mathbb{E}[z|\boldsymbol{s}]} - \mathbb{E}[z|\boldsymbol{s}]\right)}_{\widetilde{v}},$$

where $\widetilde{w}$ and $\widetilde{v}$ denote errors

muhammad.osama@it.uu.se

Proposed robust method

- Then (3) becomes

$$\widehat{w} = \left(\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta}(\boldsymbol{s})\right)^{\top}\boldsymbol{\theta} + \widetilde{\epsilon}$$

where $\boldsymbol{\delta}(\boldsymbol{s}) = \widetilde{v}\boldsymbol{\phi}(\boldsymbol{s})$ is an unobserved random deviation

muhammad.osama@it.uu.se

Proposed robust method

- Then (3) becomes

$$\widehat{w} = \left(\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta}(\boldsymbol{s})\right)^{\top}\boldsymbol{\theta} + \widetilde{\epsilon}$$

  where $\boldsymbol{\delta}(\boldsymbol{s}) = \widetilde{v}\boldsymbol{\phi}(\boldsymbol{s})$ is an unobserved random deviation
- Robust estimator with tolerance against worst-case deviation $\boldsymbol{\delta}(\boldsymbol{s})$

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \max_{\boldsymbol{\delta} \in \Delta} \sqrt{\mathbb{E}_{\mathrm{n}}\left[|\widehat{w} - (\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta})^{\top}\boldsymbol{\theta}|^2 \right]} \right\} \quad (4)$$

muhammad.osama@it.uu.se

Proposed robust method

- Then (3) becomes

$$\widehat{w} = \big(\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta}(\boldsymbol{s})\big)^\top \boldsymbol{\theta} + \widetilde{\epsilon}$$

  where $\boldsymbol{\delta}(\boldsymbol{s}) = \widetilde{v}\boldsymbol{\phi}(\boldsymbol{s})$ is an unobserved random deviation

- Robust estimator with tolerance against worst-case deviation $\boldsymbol{\delta}(\boldsymbol{s})$

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \max_{\boldsymbol{\delta} \in \Delta} \sqrt{\mathbb{E}_{\mathrm{n}}\left[|\widehat{w} - (\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta})^\top \boldsymbol{\theta}|^2 \right]} \right\} \tag{4}$$

  where

$$\Delta = \left\{ \boldsymbol{\delta} : \mathbb{E}_{\mathrm{n}}\left[|\delta_k|^2\right] \leq n^{-1}\, \mathbb{E}_{\mathrm{n}}\left[|\widehat{v}\phi_k(\boldsymbol{s})|^2\right], \forall k \right\}$$

muhammad.osama@it.uu.se

Proposed robust method

- Then (3) becomes

$$\widehat{w} = \left(\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta}(\boldsymbol{s})\right)^{\top}\boldsymbol{\theta} + \widetilde{\epsilon}$$

  where $\boldsymbol{\delta}(\boldsymbol{s}) = \widetilde{v}\boldsymbol{\phi}(\boldsymbol{s})$ is an unobserved random deviation

- Robust estimator with tolerance against worst-case deviation $\boldsymbol{\delta}(\boldsymbol{s})$

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \max_{\boldsymbol{\delta} \in \Delta} \sqrt{\mathbb{E}_{\mathrm{n}}\left[|\widehat{w} - (\widehat{v}\boldsymbol{\phi}(\boldsymbol{s}) + \boldsymbol{\delta})^{\top}\boldsymbol{\theta}|^2\right]} \right\} \qquad (4)$$

  where

$$\Delta = \left\{\boldsymbol{\delta} : \mathbb{E}_{\mathrm{n}}\left[|\delta_k|^2\right] \leq n^{-1}\,\mathbb{E}_{\mathrm{n}}\left[|\widehat{v}\phi_k(\boldsymbol{s})|^2\right], \forall k\right\}$$

- (4) is a convex problem and can be solved using coordinate descent.

muhammad.osama@it.uu.se

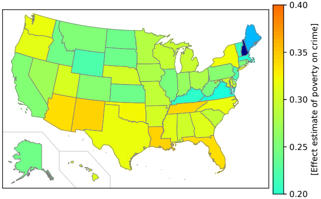
Real data

- y : Number of crimes, z : number of poor families across states $\boldsymbol{s} = \{1, \ldots, 50\}$

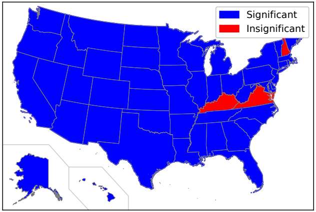
muhammad.osama@it.uu.se

Real data

- y : Number of crimes, z : number of poor families across states $\boldsymbol{s} = \{1, \ldots, 50\}$



(a) Estimate $\widehat{\tau}(\boldsymbol{s})$



(b) Significance at 5% level

- Results consistent with previous findings [2]

muhammad.osama@it.uu.se

UPPSALA UNIVERSITET

Conclusion

- We propose an orthogonalization-based strategy for estimating heterogeneous effects from spatial data in presence of spatially varying confounding variables
- Our proposed method is robust to errors-in-variables
- Visit poster $\# \ 80$ at Pacific Ballroom $6.30pm - 9pm$

muhammad.osama@it.uu.se

References

Chernozukhov et al., *Double machine learning for treatment and causal parameters*, cemmap working paper, Centre for Microdata Methods and Practice, 2016.

Ellis et al., *Crime, delinquency, and social status: A reconsideration*, Journal of Offender Rehabilitation, 2001.

muhammad.osama@it.uu.se