Learning with Instrumental and Proxy Variables

Krikamol Muandet

Max Planck Institute for Intelligent Systems

Tübingen, Germany

Joint work with

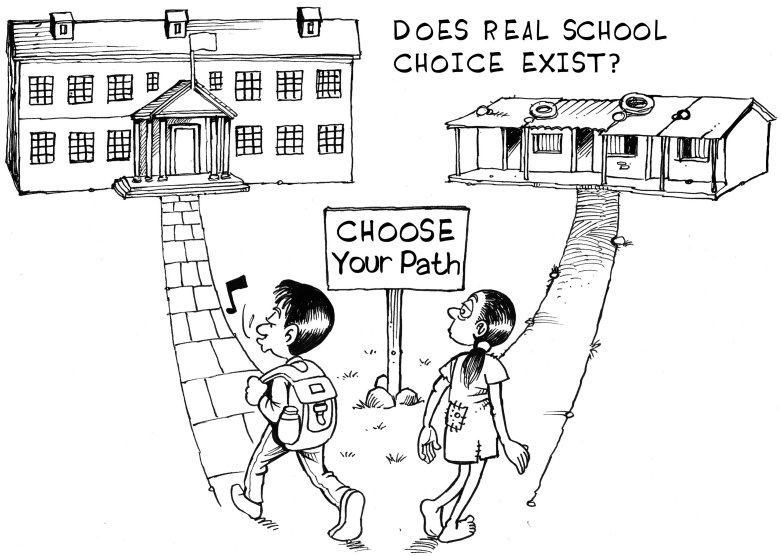
Jonas Kübler, Wittawat Jitkrittum, Arash Mehrjou, Si Kai Lee, Anant Raj, Rui Zhang, Masaaki Imaizumi, Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Bernhard Schölkopf, Ricardo Silva, Arthur Gretton, and Matt J. Kusner

Hi! PARIS Summer School 2022

July 7, 2022

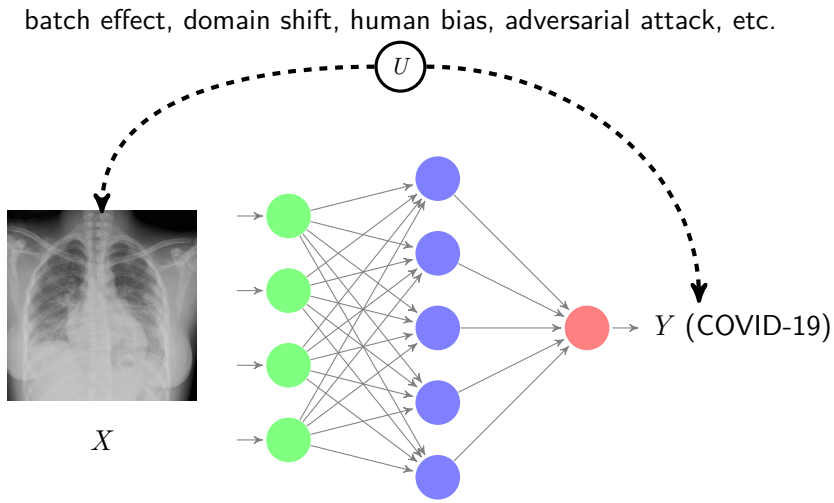
Unobserved Confounder / Spurious Correlation

Causal Inference



$\texttt{Education}\ (X) \rightarrow \texttt{Income}\ (Y)$

Machine Learning



$Y = f(X) + \varepsilon_u$

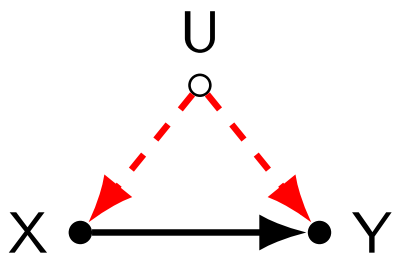
[**UAI** 2020; **NeurIPS** 2020; **ICML** 2021; Zhang et al. **Under Review**]

Spurious Correlation

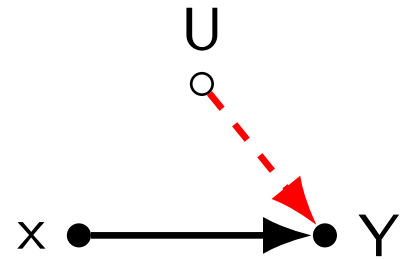
- An **interventional distribution** $P(\,Y \,|\, \mathsf{do}(X = x))$.

Spurious Correlation

- An **interventional distribution** $P(Y \mid \text{do}(X = x))$.
- $P(Y \mid \text{do}(X = x)) \neq P(Y \mid X = x)$ **observational distribution**



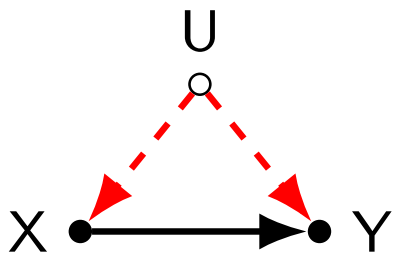
$P(Y \mid X = x)$



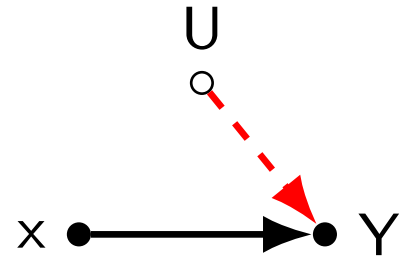
$P(Y \mid \text{do}(X = x))$

Spurious Correlation

- An **interventional distribution** $P(Y \mid \text{do}(X = x))$.
- $P(Y \mid \text{do}(X = x)) \neq P(Y \mid X = x)$ **observational distribution**

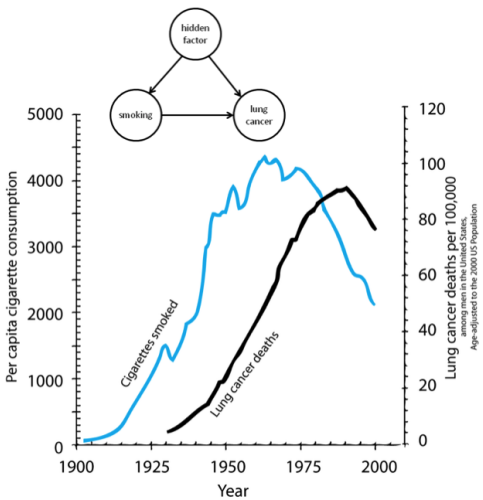


$P(Y \mid X = x)$



$P(Y \mid \text{do}(X = x))$

- The confounder U creates a **spurious correlation** between X and Y.



$X =$ `smoking (cigarettes smoked)`

$Y =$ `lung cancer deaths`

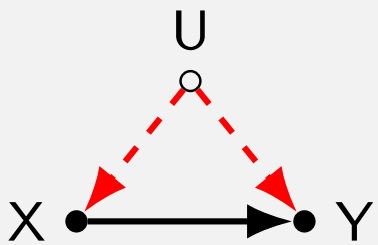
$P(Y \mid X = x) =$ `observed dealths from data`

$\text{do}(X = 0) =$ `smoking banned`

$P(Y \mid \text{do}(X = 0)) =$ `dealths after the ban`

Additive Noise Model

Newey and Powell (2003); Hoyer et al. (2008); Peters et al. (2014)

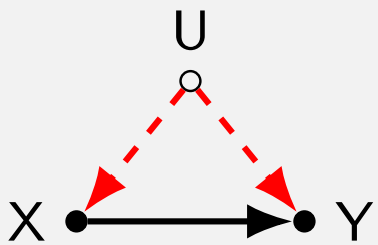


Structural Equations

$$Y \leftarrow f(X) + \varepsilon(U), \qquad \mathbb{E}[\varepsilon] = 0$$
$$X \leftarrow g(U) + \nu, \qquad \mathbb{E}[\nu] = 0$$

Additive Noise Model

Newey and Powell (2003); Hoyer et al. (2008); Peters et al. (2014)



Structural Equations

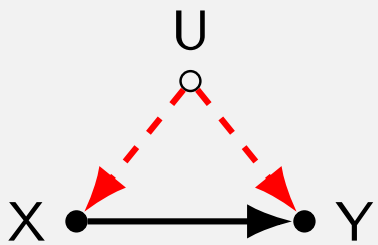
$$Y \leftarrow f(X) + \varepsilon(U), \qquad \mathbb{E}[\varepsilon] = 0$$
$$X \leftarrow g(U) + \nu, \qquad \mathbb{E}[\nu] = 0$$

- **CI**: $\mathbb{E}[\,Y \,|\, \mathsf{do}(X = x)] = f(x)$ **ML**: $\mathbb{E}[\,Y \,|\, X = x] = f(x)$

$$\mathbb{E}[\,Y \,|\, X = x] = f(x) + \mathbb{E}[\varepsilon \,|\, x]$$

Additive Noise Model

Newey and Powell (2003); Hoyer et al. (2008); Peters et al. (2014)



Structural Equations

$$Y \leftarrow f(X) + \varepsilon(U), \qquad \mathbb{E}[\varepsilon] = 0$$
$$X \leftarrow g(U) + \nu, \qquad \mathbb{E}[\nu] = 0$$

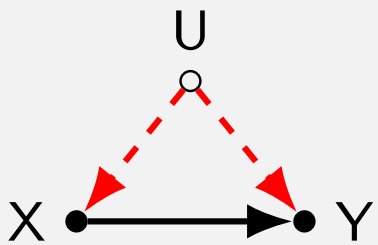
- **CI**: $\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)] = f(x)$ **ML**: $\mathbb{E}[\,Y \,|\, X = x] = f(x)$

$$\mathbb{E}[\,Y \,|\, X = x] = f(x) + \mathbb{E}[\varepsilon \,|\, x]$$

- Estimate $\widehat{\mathbb{E}}[\,Y \,|\, X = x]$ from the sample $(x_1, y_1), \ldots, (x_n, y_n) \sim P(X, Y)$
  - A biased estimate of $\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)]$.
  - The estimate $\widehat{\mathbb{E}}[\,Y \,|\, X = x]$ can be **unstable** because of $\mathbb{E}[\varepsilon \,|\, x]$.

Additive Noise Model

Newey and Powell (2003); Hoyer et al. (2008); Peters et al. (2014)



Structural Equations

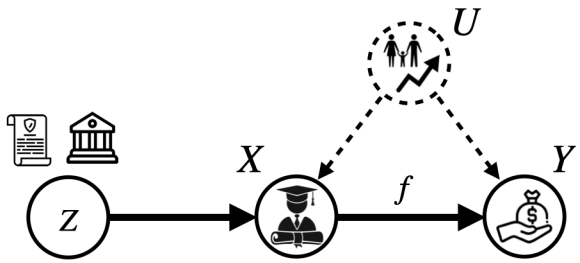
$$Y \leftarrow f(X) + \varepsilon(U), \qquad \mathbb{E}[\varepsilon] = 0$$
$$X \leftarrow g(U) + \nu, \qquad \mathbb{E}[\nu] = 0$$

- **CI**: $\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)] = f(x)$ **ML**: $\mathbb{E}[\,Y \,|\, X = x] = f(x)$

$$\mathbb{E}[\,Y \,|\, X = x] = f(x) + \mathbb{E}[\varepsilon \,|\, x]$$

- Estimate $\widehat{\mathbb{E}}[\,Y \,|\, X = x]$ from the sample $(x_1, y_1), \ldots, (x_n, y_n) \sim P(X, Y)$
  - A biased estimate of $\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)]$.
  - The estimate $\widehat{\mathbb{E}}[\,Y \,|\, X = x]$ can be **unstable** because of $\mathbb{E}[\varepsilon \,|\, x]$.
- Impossible to recover $f(x)$ without further assumptions.

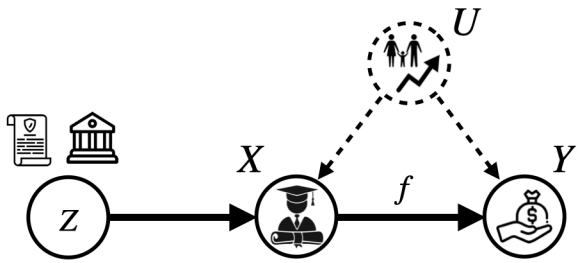
Instrumental Variables



Assumptions

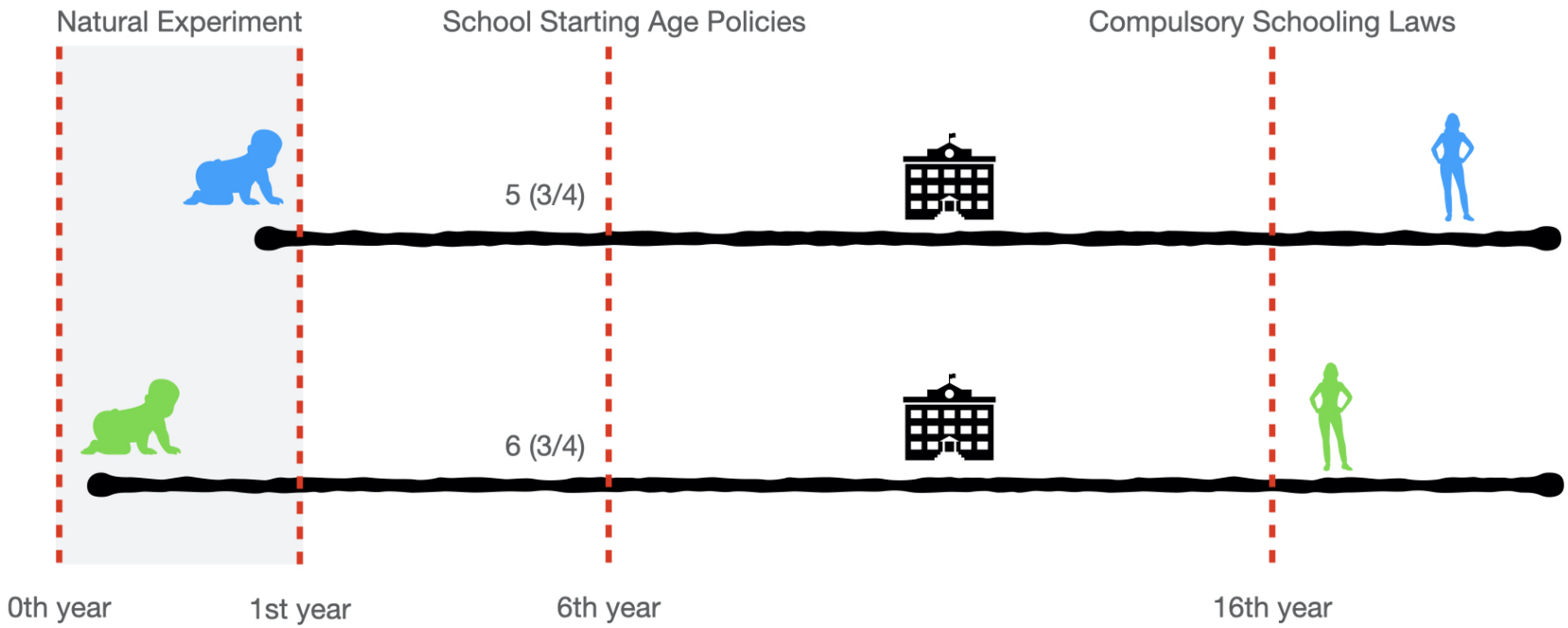
1. **Relevance**: $P(X|Z)$ is not constant in Z.
2. **Exclusion restriction**: Z affects Y only through X.
3. **Unconfoundedness**: Z is independent from U.

Instrumental Variables



Assumptions

1. **Relevance**: $P(X|Z)$ is not constant in Z.
2. **Exclusion restriction**: Z affects Y only through X.
3. **Unconfoundedness**: Z is independent from U.

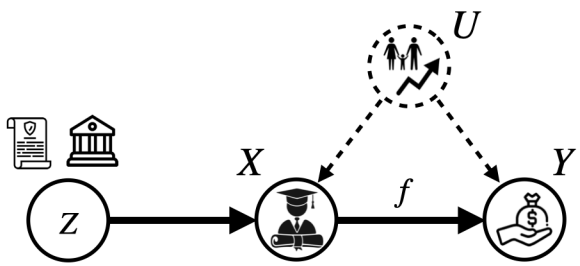


[Angrist and Krueger (1991); Angrist and Krueger (2001)]

Nobel Prize 2021

Nonparametric Instrumental Variable Regression

Newey and Powell (2003, **Econometrica**)



Assumptions

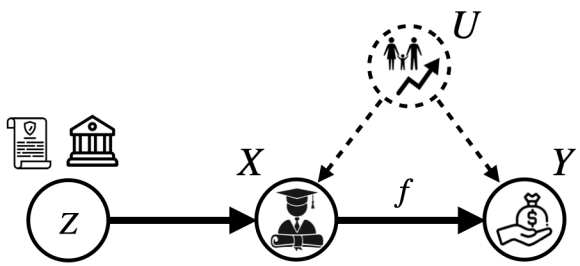
1. **Relevance**: $P(X|Z)$ is not constant in Z.
2. **Exclusion restriction**: Z affects Y only through X.
3. **Unconfoundedness**: Z is independent from U.

- Taking the conditional expectation wrt Z on both sides

$$\begin{aligned} Y &= f(X) + \varepsilon \\ \mathbb{E}[Y \mid Z] &= \mathbb{E}[f(X) \mid Z] + \underbrace{\mathbb{E}[\varepsilon \mid Z]}_{=\mathbb{E}[\varepsilon]=0} \end{aligned}$$

Nonparametric Instrumental Variable Regression

Newey and Powell (2003, **Econometrica**)



Assumptions

1. **Relevance**: $P(X|Z)$ is not constant in Z.
2. **Exclusion restriction**: Z affects Y only through X.
3. **Unconfoundedness**: Z is independent from U.

- Taking the conditional expectation wrt Z on both sides

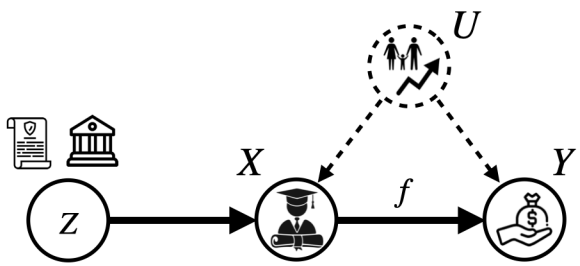
$$\begin{aligned} Y &= f(X) + \varepsilon \\ \mathbb{E}[Y \mid Z] &= \mathbb{E}[f(X) \mid Z] + \underbrace{\mathbb{E}[\varepsilon \mid Z]}_{=\mathbb{E}[\varepsilon]=0} \end{aligned}$$

- A **Fredholm integral equation** of the first kind

$$\mathbb{E}[Y \mid Z] = \int f(x)\, dP(x \mid Z)$$

Nonparametric Instrumental Variable Regression

Newey and Powell (2003, **Econometrica**)



Assumptions

1. **Relevance**: $P(X|Z)$ is not constant in Z.
2. **Exclusion restriction**: Z affects Y only through X.
3. **Unconfoundedness**: Z is independent from U.

- Taking the conditional expectation wrt Z on both sides

$$\begin{aligned} Y &= f(X) + \varepsilon \\ \mathbb{E}[Y \mid Z] &= \mathbb{E}[f(X) \mid Z] + \underbrace{\mathbb{E}[\varepsilon \mid Z]}_{=\mathbb{E}[\varepsilon]=0} \end{aligned}$$

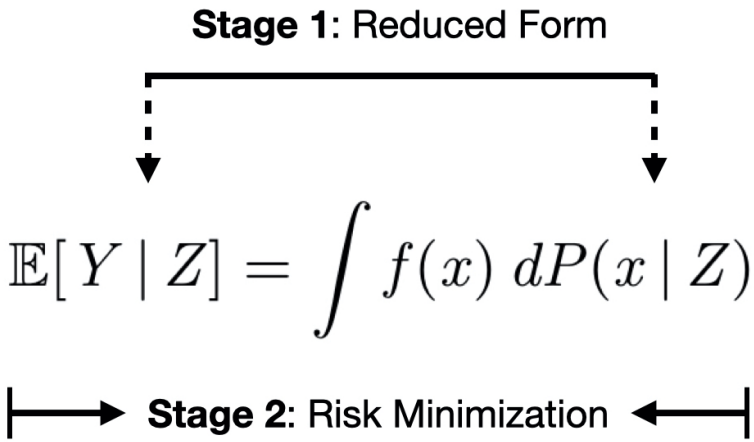
- A **Fredholm integral equation** of the first kind

$$\mathbb{E}[Y \mid Z] = \int f(x)\, dP(x \mid Z)$$

- **Completeness condition** [D'Haultfoeuille (2011, **Econ Theory**)]

For all measurable functions g, $\mathbb{E}[g(X) \mid Z] = 0$ a.s. $\Rightarrow g(X) = 0$ a.s.

Fredholm Integral Equation: Two-Stage Estimation



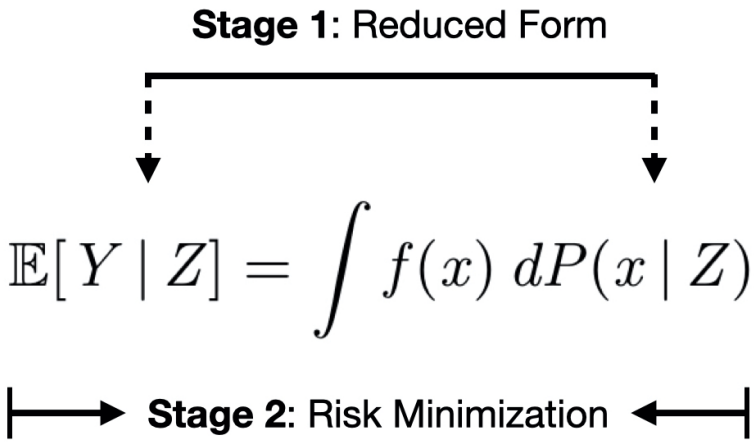
Population risk

$$\min_{f \in \mathcal{F}} \mathbb{E}_Z \left[\left(\mathbb{E}[Y \mid Z] - \mathbb{E}_{X|Z}[f(X)] \right)^2 \right]$$

Empirical risk: $(x_i, y_i, z_i) \sim P(X, Y, Z)$

$$\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \left[\left(\widehat{\mathbb{E}}[Y \mid z_i] - \widehat{\mathbb{E}}_{X|z_i}[f(X)] \right)^2 \right]$$

Fredholm Integral Equation: Two-Stage Estimation



Population risk

$$\min_{f \in \mathcal{F}} \mathbb{E}_Z \left[\left(\mathbb{E}[Y \mid Z] - \mathbb{E}_{X|Z}[f(X)] \right)^2 \right]$$

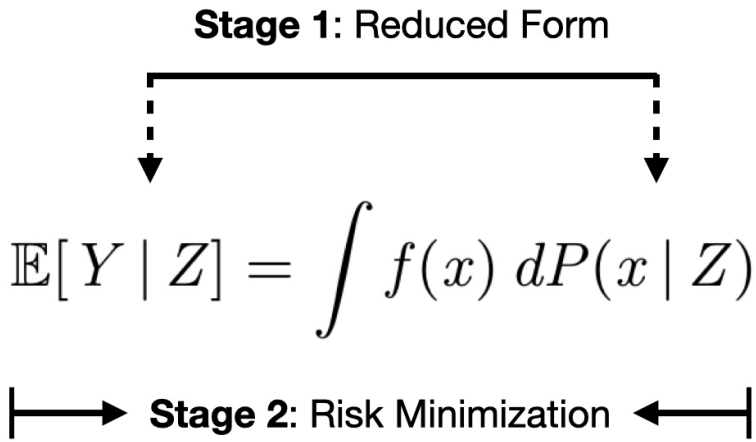
Empirical risk: $(x_i, y_i, z_i) \sim P(X, Y, Z)$

$$\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \left[\left(\widehat{\mathbb{E}}[Y \mid z_i] - \widehat{\mathbb{E}}_{X|z_i}[f(X)] \right)^2 \right]$$

Previous work

- **2SLS** [Angrist and Imbens (1996)]
- **SieveIV** [Newey and Powell (2003)]
- **DeepIV** [Hartford (2017, **ICML**)]
- **KernelIV** [Singh (2019, **NeurIPS**)]

Fredholm Integral Equation: Two-Stage Estimation



Population risk

$$\min_{f \in \mathcal{F}} \mathbb{E}_Z \left[\left(\mathbb{E}[Y \mid Z] - \mathbb{E}_{X|Z}[f(X)] \right)^2 \right]$$

Empirical risk: $(x_i, y_i, z_i) \sim P(X, Y, Z)$

$$\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \left[\left(\widehat{\mathbb{E}}[Y \mid z_i] - \widehat{\mathbb{E}}_{X|z_i}[f(X)] \right)^2 \right]$$

Previous work

- **2SLS** [Angrist and Imbens (1996)]
- **SieveIV** [Newey and Powell (2003)]
- **DeepIV** [Hartford (2017, **ICML**)]
- **KernelIV** [Singh (2019, **NeurIPS**)]

Challenges

- Require **data splitting** for two-stage estimation.
- Estimate $P(X \mid Z)$ using **one observation** from $P(X \mid Z = z)$.
- The first stage is known as "**forbidden regression**" in econometrics.
- It violates **Vapnik's principle** [Vapnik 1998].

DualIV: From Two-Stage Estimation to Two-Player Game

Muandet, Mehrjou, Lee, Raj. (2020, **NeurIPS**); Liao et al (2020, **NeurIPS**)

$$
\begin{aligned}
\min_{f\in\mathcal{F}} R(f) &= \min_{f\in\mathcal{F}} \mathbb{E}_Z[(\mathbb{E}[Y\mid Z]-\mathbb{E}_{X|Z}[f(X)])^2] \\
&\overset{(a)}{=} \min_{f\in\mathcal{F}} \mathbb{E}_Z[\max_{u\in\mathbb{R}}\{\mathbb{E}_{X|Z}[f(X)]u-\mathbb{E}[Y|Z]u-\frac{1}{2}u^2\}] \\
&\overset{(b)}{=} \min_{f\in\mathcal{F}}\max_{u\in\mathcal{U}} \mathbb{E}_Z[\mathbb{E}_{X|Z}[f(X)]u(Z)-\mathbb{E}[Y|Z]u(Z)-\frac{1}{2}u^2(Z)] \\
&= \min_{f\in\mathcal{F}}\max_{u\in\mathcal{U}} \mathbb{E}_Z[(\mathbb{E}_{X|Z}[f(X)]-\mathbb{E}[Y|Z])u(Z)-\frac{1}{2}u^2(Z)] \\
&= \boldsymbol{\min_{f\in\mathcal{F}}\max_{u\in\mathcal{U}} \mathbb{E}_{XYZ}[(f(X)-Y)u(Z)-\frac{1}{2}u^2(Z)]}
\end{aligned}
$$

(a) **Fenchel duality**: Let $\ell_y(\cdot)=\ell(y,\cdot)$ be a convex loss. For $\ell(y,y')=(y-y')^2$,

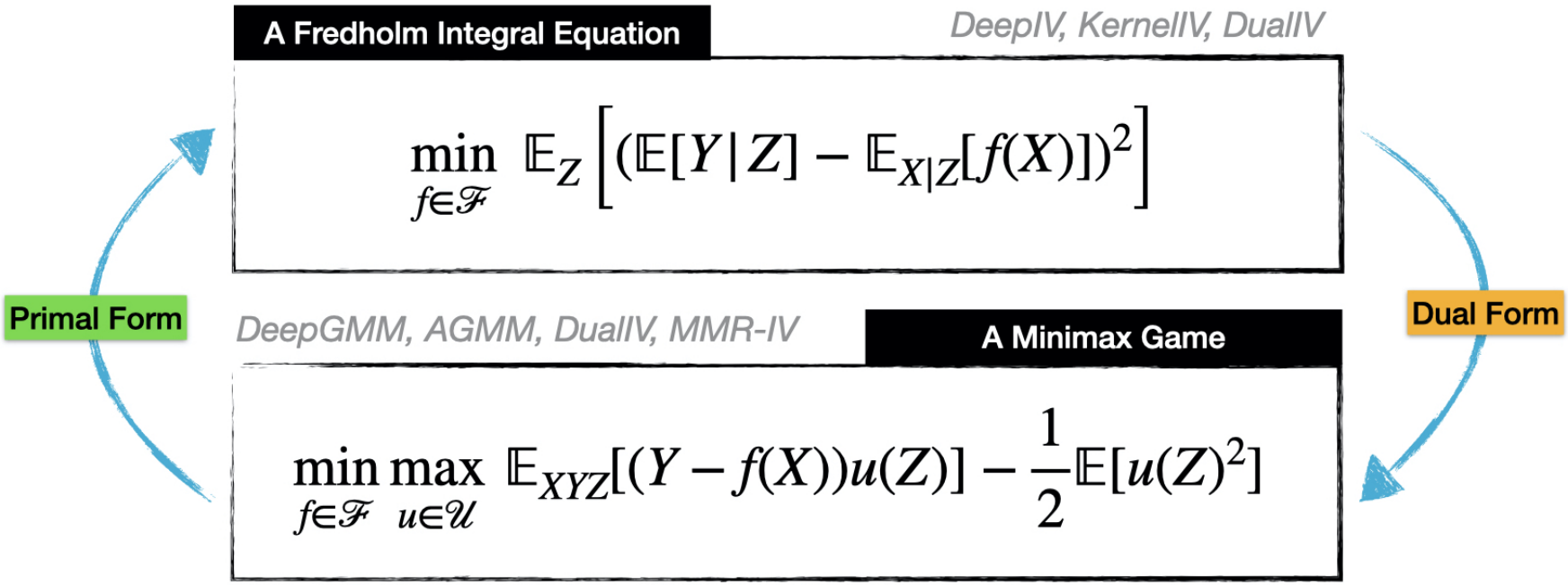
$$\ell_y(v)=\max_u\{uv-\ell_y^\star(u)\},\quad \ell_y^\star(u)=uy+\frac{1}{2}u^2.$$

(b) **Interchangeability**: $\mathbb{E}_\omega[\max_{u\in\mathbb{R}} f(u,\omega)]=\max_{u(\cdot)\in\mathcal{U}}\mathbb{E}_\omega[f(u(\omega),\omega)]$.

Dai et al. (2017; Lemma 1), Rockafellar and Wets (1998; Ch. 14), and Shapiro et al. (2014; Ch. 7)

DualIV: From Two-Stage Estimation to Two-Player Game

Muandet, Mehrjou, Lee, Raj. (2020, **NeurIPS**); Liao et al (2020, **NeurIPS**)



[Liao et al (2020, **NeurIPS**); Bennett et al. (2019, **NeurIPS**); Muandet et al. (2020, **UAI**); Dikkala et al. (2020, **NeurIPS**); Zhang et al. (2022, **Under Review**)]

DualIV: From Two-Stage Estimation to Two-Player Game

Muandet, Mehrjou, Lee, Raj. (2020, **NeurIPS**); Liao et al (2020, **NeurIPS**)

Data: $(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n) \sim P(X, Y, Z)$

- **Adversarial learning** ($\mathcal{F} = \{f_{\boldsymbol{\psi}} : \boldsymbol{\psi} \in \mathbb{R}^q\}$ and $\mathcal{U} = \{u_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbb{R}^p\}$)

$$\min_{\boldsymbol{\psi} \in \mathbb{R}^q} \max_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} [(y_i - f_{\boldsymbol{\psi}}(x_i)) u_{\boldsymbol{\theta}}(z_i)] - \frac{\gamma}{2} \|u_{\boldsymbol{\theta}}\|^2 + \frac{\lambda}{2} \|f_{\boldsymbol{\psi}}\|^2$$

[Bennett et al. (2019, **NeurIPS**); Muandet et al. (2020, **UAI**); Dikkala et al. (2020, **NeurIPS**); Zhang et al. (2022, **UR**)]

DualIV: From Two-Stage Estimation to Two-Player Game

Muandet, Mehrjou, Lee, Raj. (2020, **NeurIPS**); Liao et al (2020, **NeurIPS**)

Data: $(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n) \sim P(X, Y, Z)$

- **Adversarial learning** ($\mathcal{F} = \{f_{\boldsymbol{\psi}} : \boldsymbol{\psi} \in \mathbb{R}^q\}$ and $\mathcal{U} = \{u_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbb{R}^p\}$)

$$\min_{\boldsymbol{\psi} \in \mathbb{R}^q} \max_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} [(y_i - f_{\boldsymbol{\psi}}(x_i)) u_{\boldsymbol{\theta}}(z_i)] - \frac{\gamma}{2} \|u_{\boldsymbol{\theta}}\|^2 + \frac{\lambda}{2} \|f_{\boldsymbol{\psi}}\|^2$$

- **M-estimators** (A unit ball within an RKHS, i.e., $\mathcal{U} = \{u : \mathcal{H}_{k_{\mathcal{Z}}}, \|u\| \leq 1\}$)

$$\min_{f \in \mathcal{F}} R_V(f) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - f(x_i)) k_{\mathcal{Z}}(z_i, z_j) (y_j - f(x_j)) + \frac{\lambda}{2} \|f\|^2$$

[Bennett et al. (2019, **NeurIPS**); Muandet et al. (2020, **UAI**); Dikkala et al. (2020, **NeurIPS**); Zhang et al. (2022, **UR**)]

DualIV: From Two-Stage Estimation to Two-Player Game

Muandet, Mehrjou, Lee, Raj. (2020, **NeurIPS**); Liao et al (2020, **NeurIPS**)

Data: $(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n) \sim P(X, Y, Z)$

- **Adversarial learning** ($\mathcal{F} = \{f_{\boldsymbol{\psi}} : \boldsymbol{\psi} \in \mathbb{R}^q\}$ and $\mathcal{U} = \{u_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbb{R}^p\}$)

$$\min_{\boldsymbol{\psi}\in\mathbb{R}^q} \max_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}[(y_i - f_{\boldsymbol{\psi}}(x_i))u_{\boldsymbol{\theta}}(z_i)] - \frac{\gamma}{2}\|u_{\boldsymbol{\theta}}\|^2 + \frac{\lambda}{2}\|f_{\boldsymbol{\psi}}\|^2$$

- **M-estimators** (A unit ball within an RKHS, i.e., $\mathcal{U} = \{u : \mathcal{H}_{k_{\mathcal{Z}}}, \|u\| \leq 1\}$)

$$\min_{f\in\mathcal{F}} R_V(f) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(y_i - f(x_i))k_{\mathcal{Z}}(z_i, z_j)(y_j - f(x_j)) + \frac{\lambda}{2}\|f\|^2$$

- **Kernel machines** ($\mathcal{F} \in \mathcal{H}_{k_{\mathcal{X}}}$ and $\mathcal{U} = \{u : \mathcal{H}_{k_{\mathcal{Z}}}, \|u\| \leq 1\}$)

$$f(x) = \sum_{i=1}^{n}\alpha_i k_{\mathcal{X}}(x_i, x), \quad \boldsymbol{\alpha} = (\mathbf{KLK} + \lambda\mathbf{K})^{-1}\mathbf{KL}\boldsymbol{y}$$

where $\mathbf{K}_{ij} = k_{\mathcal{X}}(x_i, x_j)$, $\mathbf{L}_{ij} = k_{\mathcal{Z}}(z_i, z_j)/n^2$, and $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$.

[Bennett et al. (2019, **NeurIPS**); Muandet et al. (2020, **UAI**); Dikkala et al. (2020, **NeurIPS**); Zhang et al. (2022, **UR**)]

DualIV: From Two-Stage Estimation to Two-Player Game

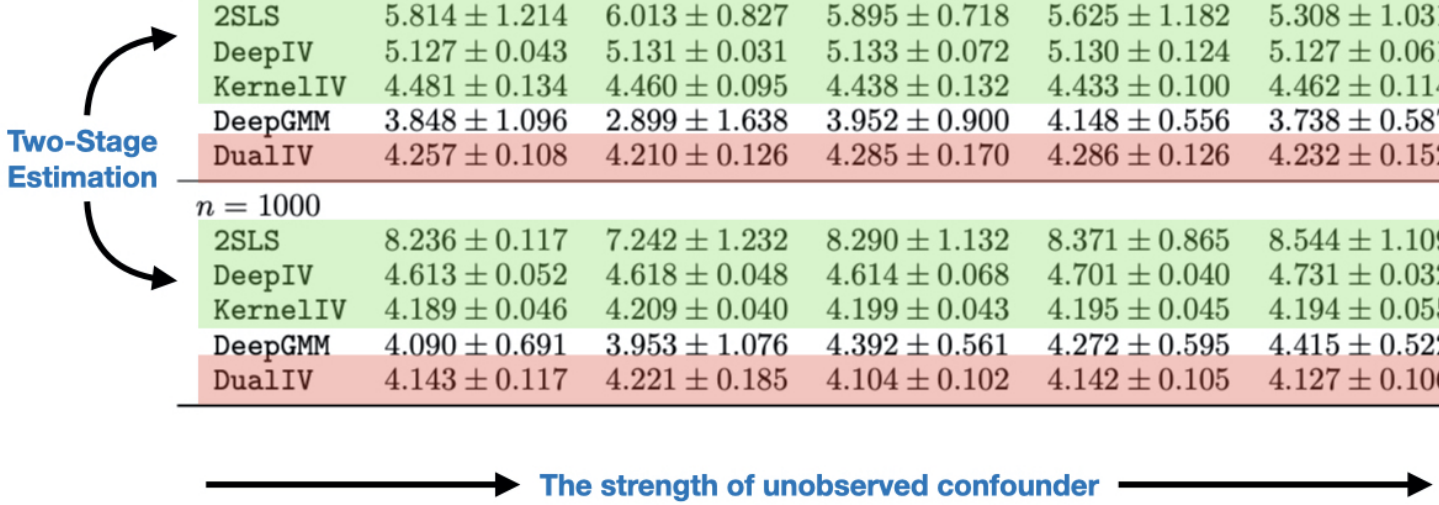
Muandet, Mehrjou, Lee, Raj. (2020, **NeurIPS**)

Demand design: $Y = f(X) + \varepsilon$ where $X = (P, T, S)$ and $Z = (C, T, S)$

- Y: sale, P: price, (T, S): time of year and customer sentiment
- Sale Y and price P are confounded by supply-side market forces
- C: supply cost shifter (instrument)

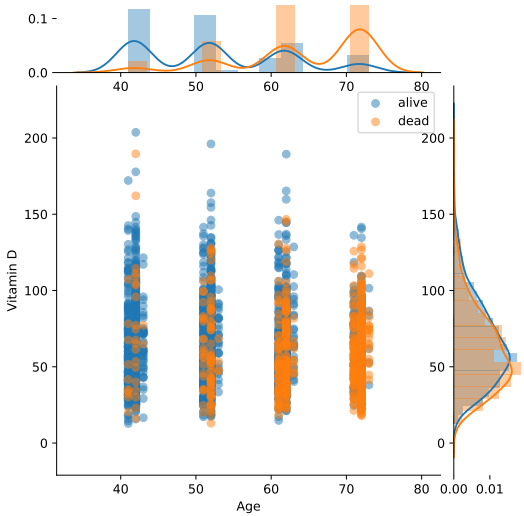
Table 1: Comparisons of IV regression methods in small (top) and medium (bottom) sample size regimes. We report the $\log_{10}$ mean squared error (MSE) and its standard deviations over 20 trials.

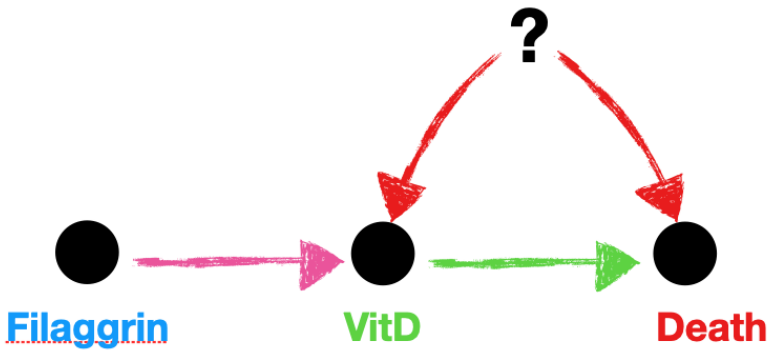
$n = 50$	$\log_{10}$ **Mean Squared Error (MSE)**				
	$\rho = 0.1$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0.9$
2SLS	5.814 ± 1.214	6.013 ± 0.827	5.895 ± 0.718	5.625 ± 1.182	5.308 ± 1.031
DeepIV	5.127 ± 0.043	5.131 ± 0.031	5.133 ± 0.072	5.130 ± 0.124	5.127 ± 0.061
KernelIV	4.481 ± 0.134	4.460 ± 0.095	4.438 ± 0.132	4.433 ± 0.100	4.462 ± 0.114
DeepGMM	3.848 ± 1.096	2.899 ± 1.638	3.952 ± 0.900	4.148 ± 0.556	3.738 ± 0.587
DualIV	4.257 ± 0.108	4.210 ± 0.126	4.285 ± 0.170	4.286 ± 0.126	4.232 ± 0.152
$n = 1000$					
2SLS	8.236 ± 0.117	7.242 ± 1.232	8.290 ± 1.132	8.371 ± 0.865	8.544 ± 1.109
DeepIV	4.613 ± 0.052	4.618 ± 0.048	4.614 ± 0.068	4.701 ± 0.040	4.731 ± 0.032
KernelIV	4.189 ± 0.046	4.209 ± 0.040	4.199 ± 0.043	4.195 ± 0.045	4.194 ± 0.055
DeepGMM	4.090 ± 0.691	3.953 ± 1.076	4.392 ± 0.561	4.272 ± 0.595	4.415 ± 0.522
DualIV	4.143 ± 0.117	4.221 ± 0.185	4.104 ± 0.102	4.142 ± 0.105	4.127 ± 0.106



Improvement over two-stage methods: $\sim$20% ($n = 50$) and $\sim$9% ($n = 1000$)

Vitamin D Data



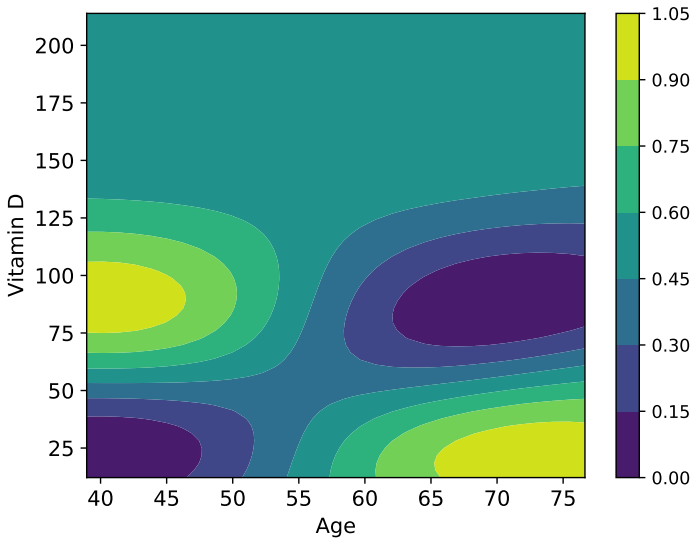


- A 10-year study on 2571 individuals aged 40–71.
- There are 4 variables:
  1. **Age** (at baseline)
  2. **Filaggrin** (binary indicator of filaggrin mutations)
  3. **VitD** (vitamin D level at baseline)
  4. **Death** (binary indicator of death during study)
- The goal is to evaluate the potential effect of VitD on Death.

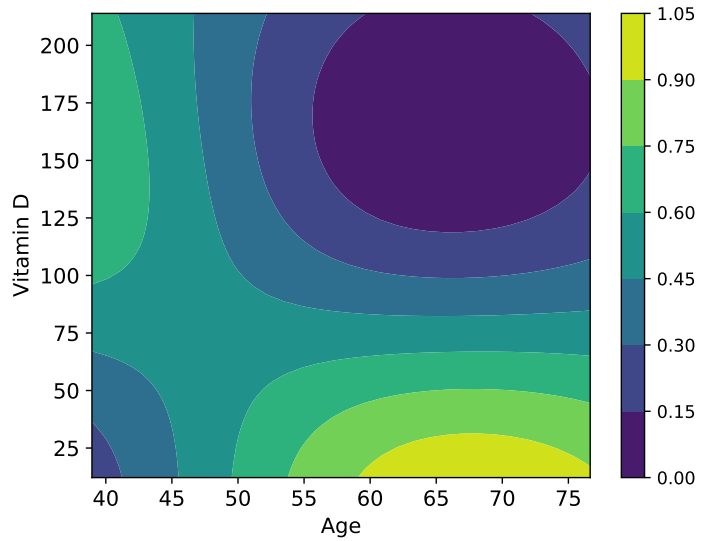
$$X = \text{VitD}, \quad Y = \text{Death}, \quad Z = \text{Filaggrin}$$

Vitamin D Data

$$\text{Death} = f(\text{VitD}, \text{Age})$$



(a) Kernel Ridge Regression (IV: None)

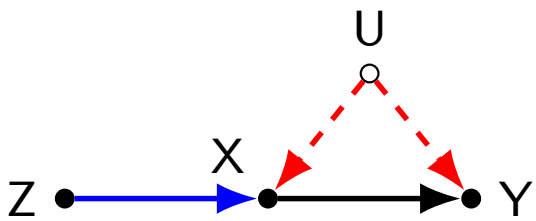


(b) Our Method (IV: Filaggrin Mutation)

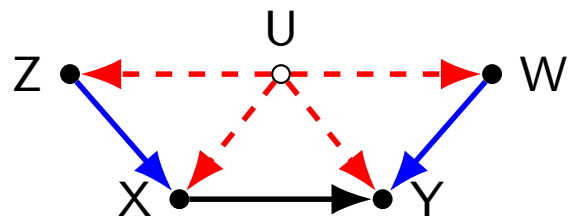
Proximal Causal Learning with Kernels

Mastouri, Zhu, Gultchin, Korba, Silva, Kusner, Gretton, Muandet (2021, **ICML**)

Instrumental Variable



Proxy Variables



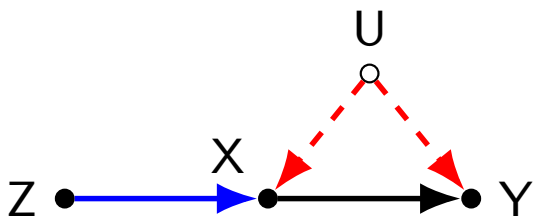
- Treatment-inducing proxy Z and outcome-inducing proxy W

[Miao et al. (2018, **Biometrika**)]

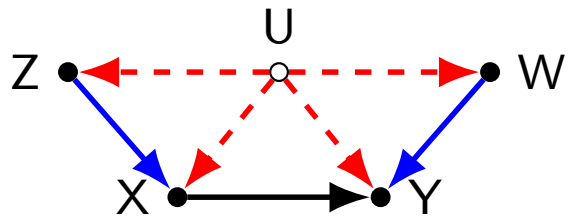
Proximal Causal Learning with Kernels

Mastouri, Zhu, Gultchin, Korba, Silva, Kusner, Gretton, Muandet (2021, **ICML**)

Instrumental Variable



Proxy Variables



- Treatment-inducing proxy Z and outcome-inducing proxy W
- The **causal effect** estimation

$$\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)] = \mathbb{E}_W[h(x, W)]$$

where h is a **confounding bridge** function satisfying an integral equation

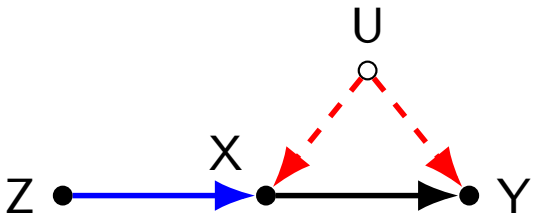
$$\mathbb{E}[\,Y \,|\, x, z] = \int h(x, w)\, dP(w \,|\, x, z)$$

[Miao et al. (2018, **Biometrika**)]

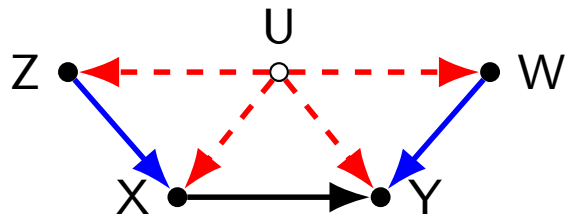
Proximal Causal Learning with Kernels

Mastouri, Zhu, Gultchin, Korba, Silva, Kusner, Gretton, Muandet (2021, **ICML**)

Instrumental Variable



Proxy Variables



- Treatment-inducing proxy Z and outcome-inducing proxy W
- The **causal effect** estimation

$$\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)] = \mathbb{E}_W[h(x, W)]$$

where h is a **confounding bridge** function satisfying an integral equation

$$\mathbb{E}[\,Y \,|\, x, z] = \int h(x, w)\, dP(w \,|\, x, z)$$

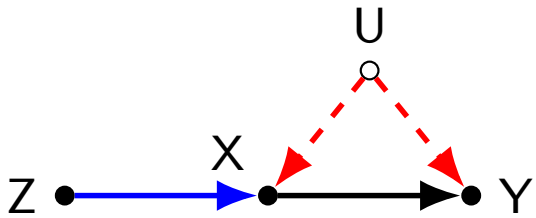
- The impact of **legalized abortion** (X) on **crime** (Y) [Donohue and Levitt (2001)]
  (Supreme Court's 1973 decision in *Roe* v. *Wade*)

[Miao et al. (2018, **Biometrika**)]

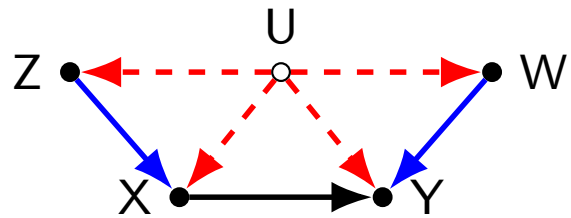
Proximal Causal Learning with Kernels

Mastouri, Zhu, Gultchin, Korba, Silva, Kusner, Gretton, Muandet (2021, **ICML**)

Instrumental Variable



Proxy Variables



- Treatment-inducing proxy Z and outcome-inducing proxy W
- The **causal effect** estimation

$$\mathbb{E}[\,Y \,|\, \mathrm{do}(X = x)] = \mathbb{E}_W[h(x, W)]$$

where h is a **confounding bridge** function satisfying an integral equation

$$\mathbb{E}[\,Y \,|\, x, z] = \int h(x, w)\, dP(w \,|\, x, z)$$

- The impact of **legalized abortion** (X) on **crime** (Y) [Donohue and Levitt (2001)]
  (Supreme Court's 1973 decision in *Roe* v. *Wade*)
- The impact of **grade retention** (X) on **cognitive outcome** (Y) [Deaner (2018)]

[Miao et al. (2018, **Biometrika**)]

Fantastic "Instruments" and Where to Find Them

Mendelian Randomization

[Adam (2019, **Nature**)]

Issue: weak instruments

Fantastic "Instruments" and Where to Find Them

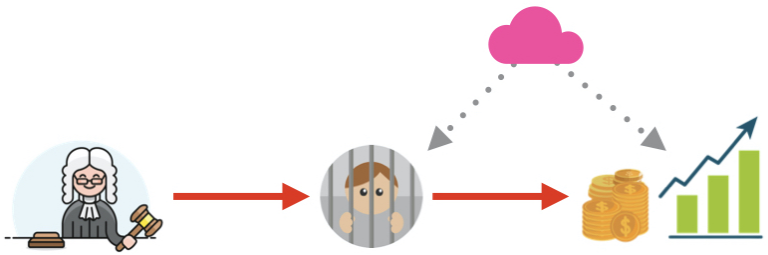
Mendelian Randomization

[Adam (2019, **Nature**)]

Issue: weak instruments

Judge Leniency Design

[Kling (2006, **AER**)]



Issue: algorithmic decision making

Fantastic "Instruments" and Where to Find Them

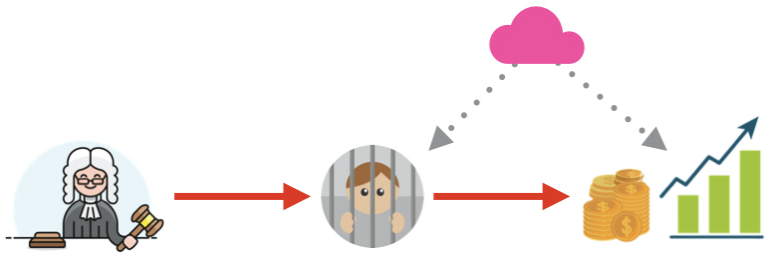
Mendelian Randomization

[Adam (2019, **Nature**)]

Issue: weak instruments

Judge Leniency Design

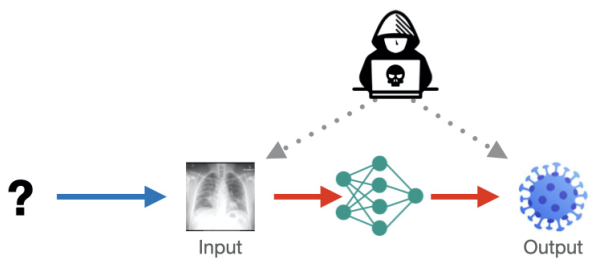
[Kling (2006, **AER**)]



Issue: algorithmic decision making

Adversarial Examples

[N.A.]



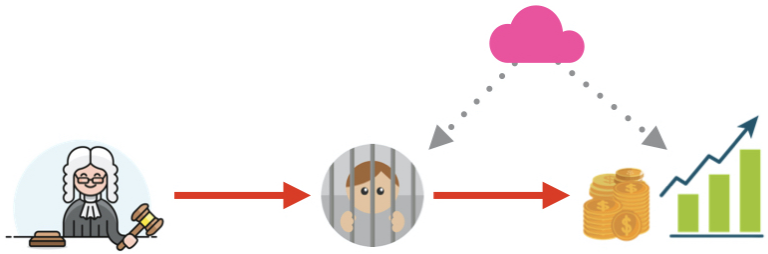
Issue: prone to adversarial attack

Fantastic "Instruments" and Where to Find Them

Mendelian Randomization
[Adam (2019, **Nature**)]

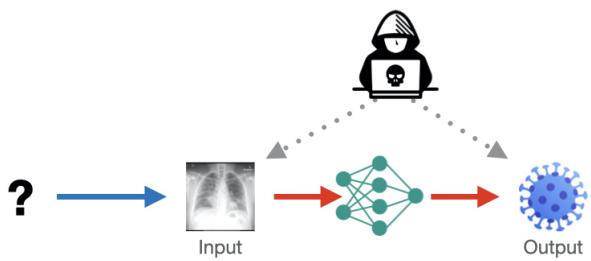
Issue: weak instruments

Judge Leniency Design
[Kling (2006, **AER**)]



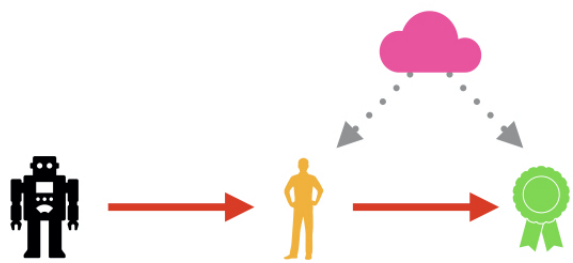
Issue: algorithmic decision making

Adversarial Examples
[N.A.]



Issue: prone to adversarial attack

Algorithmic Instruments
[Ngo et al. (2021, **ICML**)]



Issue: non-compliance