



# 基于工具变量的因果推断 和因果可泛化学习

---

况琨

浙江大学计算机学院

<https://kunkuang.github.io/>

# Decision making

---

- Does predictive models guide decision making?
- System changes algorithm from A to B at some point.
- Is the new algorithm B better?
- Say algorithm that provides promotion or discount link to a different customers



Algorithm A



Algorithm B

# Decision making

---

- Measure success rate (SR)

Old Algorithm (A)	New Algorithm (B)
50/1000 (5%)	54/1000 (5.4%)



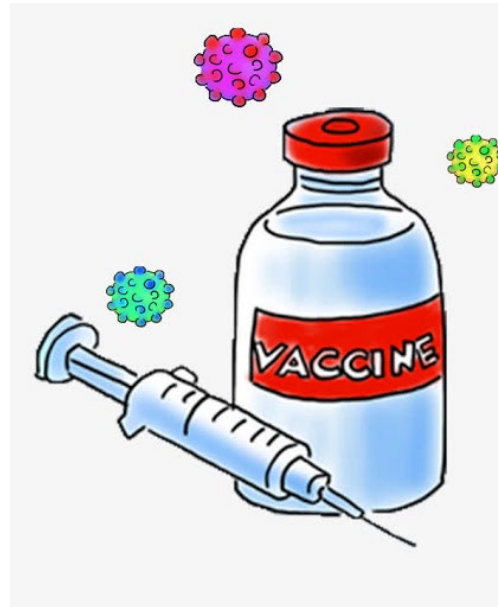
New algorithm increases overall success rate, so it is better?

	Old Algorithm (A)	New Algorithm (B)
Low-income Users	10/400 (2.5%)	4/200 (2%)
High-income Users	40/600 (6.6%)	50/800 (6.2%)
Overall	50/1000 (5%)	54/1000 (5.4%)

Which is better?

# Decision making with Causality

- **Causal Effect Estimation** is necessary for decision making!



**Causal effect estimation** plays an important role on decision making!

## A practical definition

Definition: T causes Y if and only if  
changing T leads to a change in Y,  
keep everything else constant.

**Causal effect** is defined as the magnitude by which Y is changed by a unit change in T.

**Two key points:** changing T, keeping everything else constant

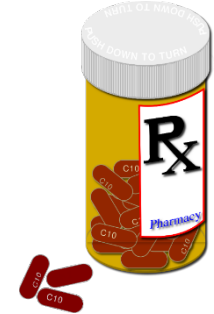
# Treatment Effect Estimation

- Treatment Variable:  $T = 1$  or  $T = 0$
- Potential Outcome:  $Y(T = 1)$  and  $Y(T = 0)$
- Individual Treatment Effect (ITE)

$$ITE(i) = Y_i(T_i = 1) - Y_i(T_i = 0)$$

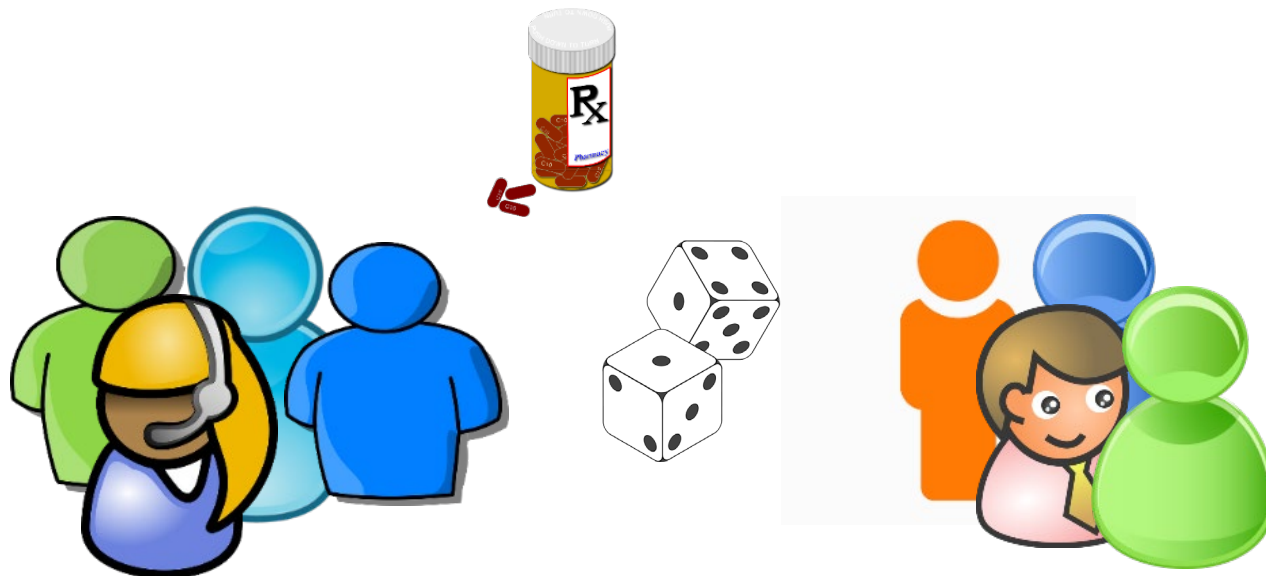
- Average Treatment Effect (ATE):

$$ATE = E[Y(T = 1) - Y(T = 0)]$$



**Two key points:** changing T, keeping everything else constant

# Randomized Experiments are the “Gold Standard”



- Drawbacks of randomized experiments:
  - Cost
  - Unethical

# Causal Inference with Observational Data

- Counterfactual problem:  $ATE = E[Y(T = 1) - Y(T = 0)]$
- In observational data, we have units with different T:

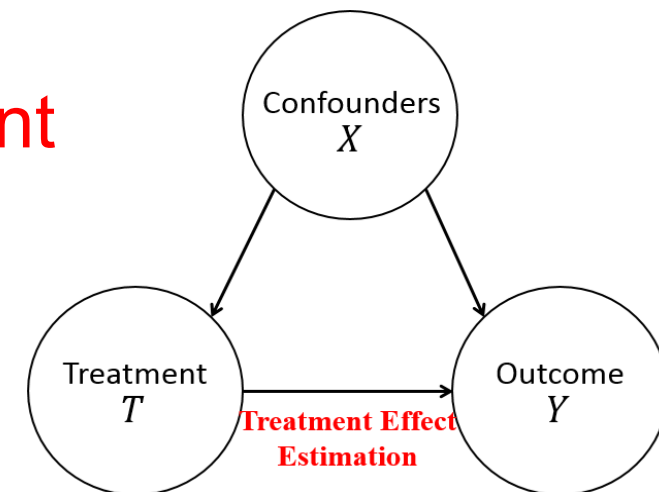
$$E[Y(T = 1)] \text{ or } E[Y(T = 0)]$$

- Can we estimate ATE by directly comparing the average outcome between groups with T=1 and T=0?

- **No, because confounders X might not be constant**

- Two key points:

- Changing T (T=1 and T=0)
- Keeping everything else (Confounder X) constant



# Causal Inference with Observational Data

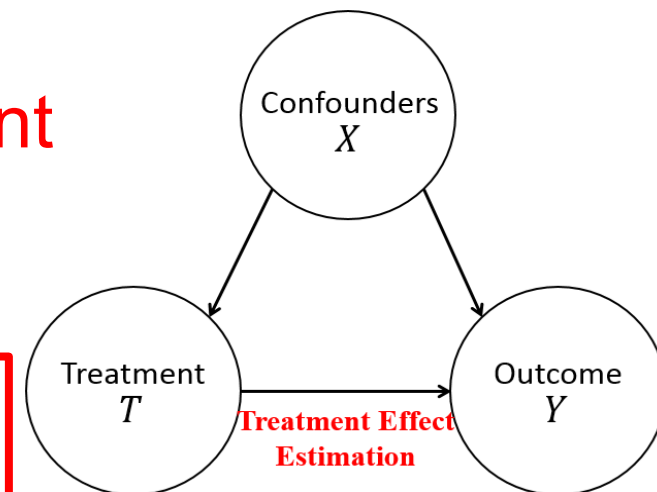
- Counterfactual problem:  $ATE = E[Y(T = 1) - Y(T = 0)]$
- In observational data, we have units with different T:

$$E[Y(T = 1)] \text{ or } E[Y(T = 0)]$$

- Can we estimate ATE by directly comparing the average outcome between groups with T=1 and T=0?
  - **No, because confounders X might not be constant**

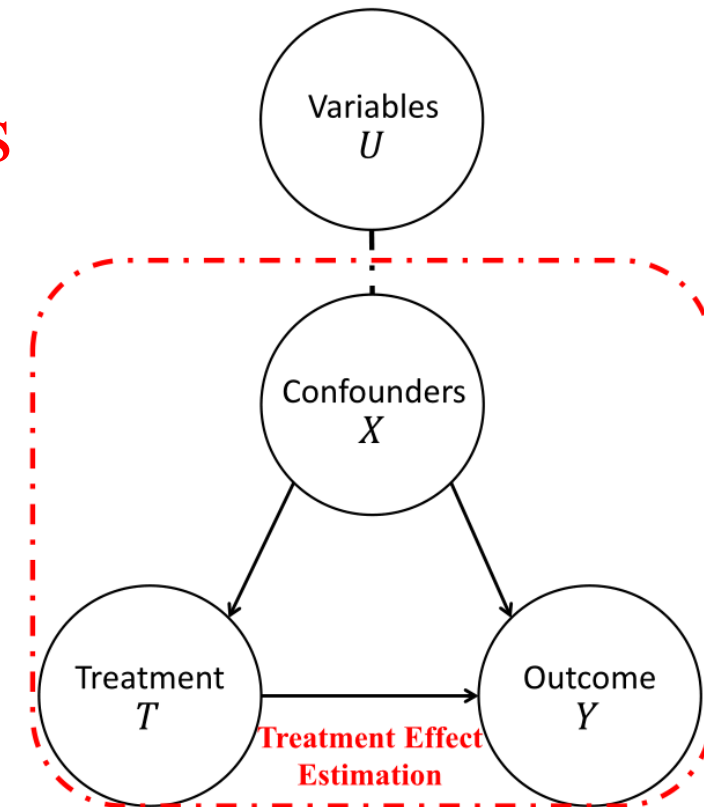
- Two key points:

**Balancing Confounders' Distribution**



# Related Work

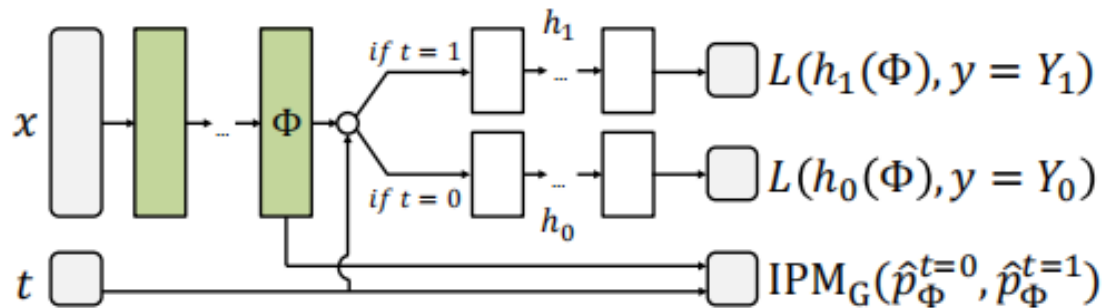
- Matching Methods
  - *Exactly Matching, Coarse Matching*
  - **Poor performance in high dimensional settings**
- Propensity Score based Methods
  - Propensity score  $e(\mathbf{X}) = p(T = 1|\mathbf{X})$
  - *Matching, Weighting, Doubly Robust*
  - **Treat all observed variables as confounders, and ignore the non-confounders**
  - **Mainly designed for binary treatment**



(a) Previous Causal Framework.

## Related Work

- Representation Learning based Methods
  - Similar representation between treatment groups.
  - Accurate prediction on factual/counterfactual outcome



$$\min_{\substack{h, \Phi \\ \|\Phi\|=1}} \frac{1}{n} \sum_{i=1}^n w_i \cdot L(h(\Phi(x_i), t_i), y_i) + \lambda \cdot \mathfrak{R}(h) \\ + \alpha \cdot \text{IPM}_G(\{\Phi(x_i)\}_{i:t_i=0}, \{\Phi(x_i)\}_{i:t_i=1}),$$

$$\text{with } w_i = \frac{t_i}{2u} + \frac{1-t_i}{2(1-u)}, \text{ where } u = \frac{1}{n} \sum_{i=1}^n t_i$$

and  $\mathfrak{R}$  is a model complexity term.

- **Confounder differentiation, binary treatment, might ignore confounders**

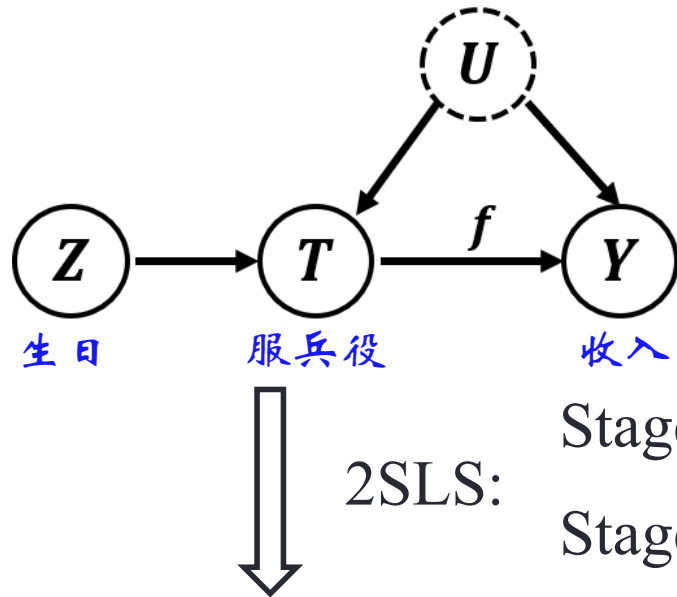
# New challenges in Big Data era

- **Automatically separate confounders**
  - Not all observed variables are confounders
  - Data-Driven Variables Decomposition ( $D^2VD$ , DeR-CFR)
- **Remove unobserved confounding bias**
  - Not all confounders are observed
  - Automatic Instrumental Variable Decomposition (AutoIV, GIV)
- **Continuous/Complex treatment effect estimation**
  - Treatment variables are not always binary
  - Generative Adversarial De-confounding (GAD, CRNet)

# New challenges in Big Data era

- **Automatically separate confounders**
  - Not all observed variables are confounders
  - Data-Driven Variables Decomposition (D<sup>2</sup>VD, DeR-CFR)
- **Remove unobserved confounding bias**
  - Not all confounders are observed
  - Automatic Instrumental Variable Decomposition (AutoIV, GIV)
- **Continuous/Complex treatment effect estimation**
  - Treatment variables are not always binary
  - Generative Adversarial De-confounding (GAD, CRNet)

# Instrumental Variable Regression



Conditions of IV (instrumental variable)

- Relevance:  $P(T|Z) \neq P(T)$
- Exclusion:  $P(Y|Z, T, U) \neq P(Y|T, U)$
- Unconfounded:  $Z \perp U$

Stage 1: regressing  $T$  on  $Z$

$$\hat{T} = \hat{g}(Z)$$

Stage 2: regressing  $Y$  on  $\hat{T}$

$$\hat{Y} = \hat{f}(\hat{T})$$

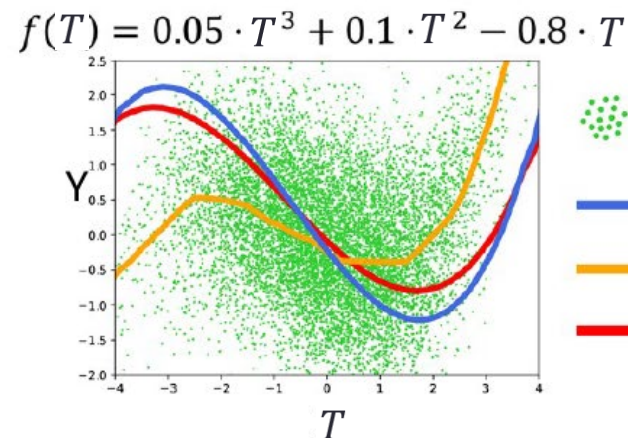
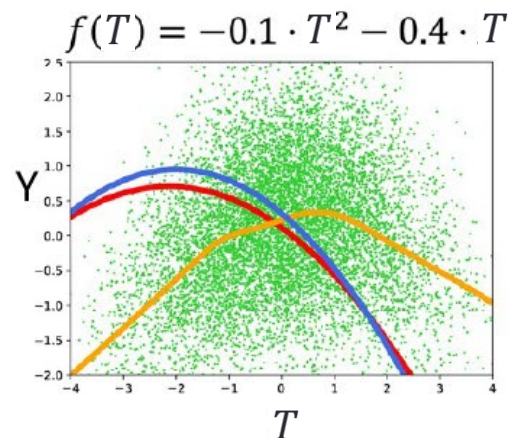


$$Z \sim \mathcal{N}(0,1)$$

$$U \sim \mathcal{N}(0,1)$$

$$T = Z + U$$

$$Y = f(T) + U$$



• Data  $P(T, Y)$

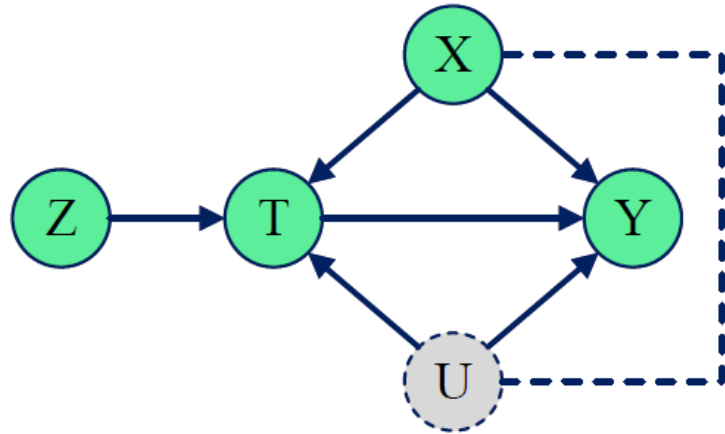
—  $f$

—  $\hat{f}^{NN}$

—  $\hat{f}^{IV}$

Requiring pre-defined IVs,  
Limited to linear setting

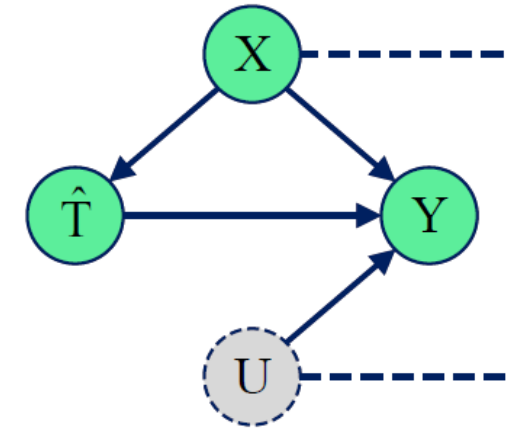
# Non-linear Instrumental Variable Regression



Non-linear IV regression (DeepIV, KernelIV et.al)

Stage 1: regressing  $T$  on  $Z$  and  $X$       $\hat{T} = \hat{g}(Z, X)$

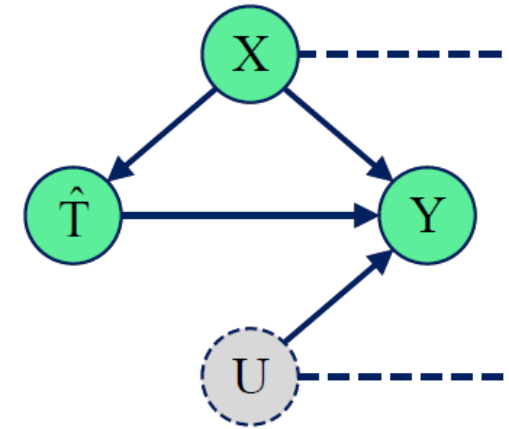
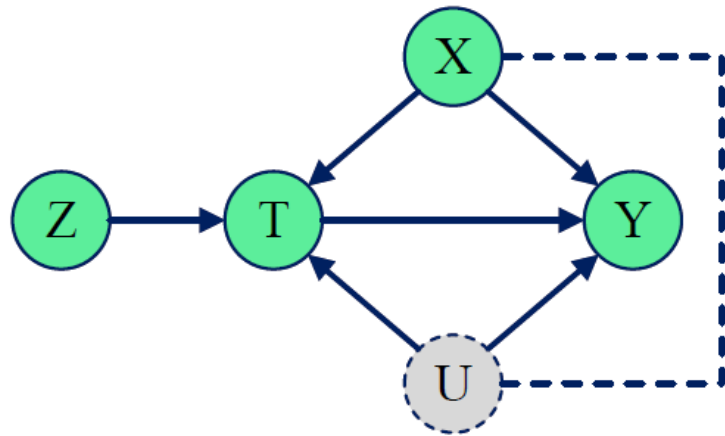
Stage 2: regressing  $Y$  on  $\hat{T}$  and  $X$       $\hat{Y} = \hat{f}(\hat{T}, X)$



Stage 1 regression brings  
confounding bias in stage 2

Confounder Balancing + IV Regression

# Confounder Balanced Instrumental Variable Regression



CB-IV (Confounder Balanced IV regression):

Stage 1 (Treatment regression): regressing  $T$  on  $Z$  and  $X$        $\hat{T} = \hat{g}(Z, X)$

**Confounder balancing:** learning a balanced confounder representation  $\phi(X)$  such that  $\hat{T} \perp \phi(X)$

Stage 2 (Outcome regression): regressing  $Y$  on  $\hat{T}$  and  $\phi(X)$        $\hat{Y} = \hat{f}(\hat{T}, \phi(X))$

# Confounder Balanced Instrumental Variable Regression

Table 2: The bias (mean  $\pm$  std) of ATE estimation on real-world data (Data- $m_Z$ - $m_X$ - $m_U$ )

IV based methods

Confounder balancing  
based methods

		Within-Sample			
Method	IHDP-2-6-0	IHDP-2-4-2	Twins-5-8-0	Twins-5-5-3	
DeepIV-LOG	2.8736 $\pm$ 0.0577	2.6227 $\pm$ 0.0651	0.0135 $\pm$ 0.0215	0.0237 $\pm$ 0.0111	
DeepIV-GMM	3.7760 $\pm$ 0.0316	3.7396 $\pm$ 0.0402	0.0194 $\pm$ 0.0047	0.0221 $\pm$ 0.0041	
OneSIV	1.7249 $\pm$ 0.3752	1.7411 $\pm$ 0.3422	0.0083 $\pm$ 0.0191	0.0080 $\pm$ 0.0167	
DFIV	3.5543 $\pm$ 0.0891	3.6218 $\pm$ 0.1038	0.0268 $\pm$ 0.0005	0.0265 $\pm$ 0.0003	
DFL	3.2018 $\pm$ 0.0496	3.1991 $\pm$ 0.0374	0.0624 $\pm$ 0.0586	0.0847 $\pm$ 0.0049	
DirectRep	0.0675 $\pm$ 0.0562	0.4600 $\pm$ 0.0711	0.0167 $\pm$ 0.0171	0.0193 $\pm$ 0.0251	
CFR	0.0854 $\pm$ 0.0579	0.4826 $\pm$ 0.0642	0.0115 $\pm$ 0.0167	0.0223 $\pm$ 0.0176	
DRCFR	0.0553 $\pm$ 0.0644	0.4336 $\pm$ 0.0692	0.0114 $\pm$ 0.0221	0.0118 $\pm$ 0.0174	
<b>CB-IV</b>	<b>0.0117 <math>\pm</math> 0.3882</b>	<b>0.1601 <math>\pm</math> 0.2499</b>	<b>0.0067 <math>\pm</math> 0.0271</b>	<b>0.0014 <math>\pm</math> 0.0249</b>	
		Out-of-Sample			
Method	IHDP-2-6-0	IHDP-2-4-2	Twins-5-8-0	Twins-5-5-3	
DeepIV-LOG	2.8760 $\pm$ 0.0553	2.6226 $\pm$ 0.0692	0.0140 $\pm$ 0.0208	0.0238 $\pm$ 0.0111	
DeepIV-GMM	3.7768 $\pm$ 0.0350	3.7388 $\pm$ 0.0416	0.0193 $\pm$ 0.0047	0.0221 $\pm$ 0.0040	
OneSIV	1.7287 $\pm$ 0.3725	1.7351 $\pm$ 0.3430	0.0082 $\pm$ 0.0191	0.0081 $\pm$ 0.0168	
DFIV	3.5538 $\pm$ 0.0904	3.6225 $\pm$ 0.1061	0.0268 $\pm$ 0.0005	0.0265 $\pm$ 0.0003	
DFL	3.2038 $\pm$ 0.0496	3.1994 $\pm$ 0.0376	0.0624 $\pm$ 0.0584	0.0846 $\pm$ 0.0046	
DirectRep	0.0608 $\pm$ 0.0817	0.4571 $\pm$ 0.0759	0.0162 $\pm$ 0.0175	0.0194 $\pm$ 0.0253	
CFR	0.0785 $\pm$ 0.0810	0.4804 $\pm$ 0.0687	0.0110 $\pm$ 0.0163	0.0225 $\pm$ 0.0180	
DRCFR	0.0450 $\pm$ 0.0953	0.4321 $\pm$ 0.0673	0.0113 $\pm$ 0.0219	0.0118 $\pm$ 0.0174	
<b>CB-IV</b>	<b>0.0150 <math>\pm</math> 0.3927</b>	<b>0.1578 <math>\pm</math> 0.2540</b>	<b>0.0065 <math>\pm</math> 0.0270</b>	<b>0.0015 <math>\pm</math> 0.0247</b>	

Requiring  
pre-defined IVs

# Confounded IVs: Estimating Individualized Causal Effect with Confounded Instruments

## Confounded IVs:

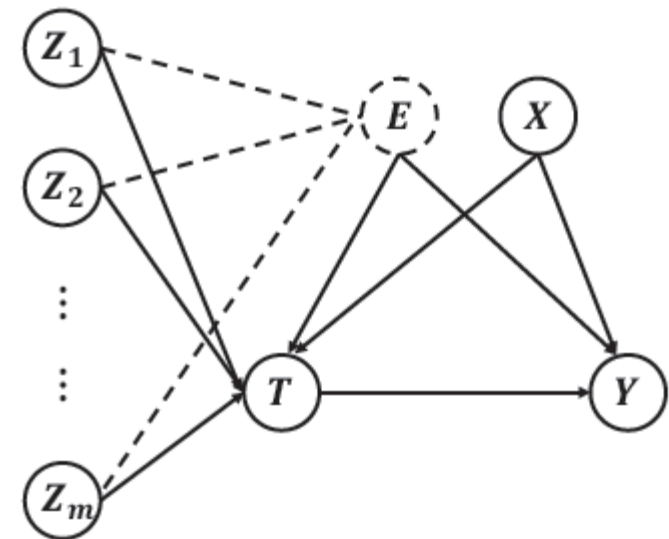
- **Violation on Unconfounded Instrument:**  $Z_i$  correlates to  $E$  conditioning on  $X$
- Often happens in real cases and leads to failure of all IV methods.

## Our setting:

- $\{Z_i\}_{i=1}^m$  represents candidates for IV
- $\{Z_i\}_{i=1}^m$  are invalid IV, called as confounded IVs

## Our Goal:

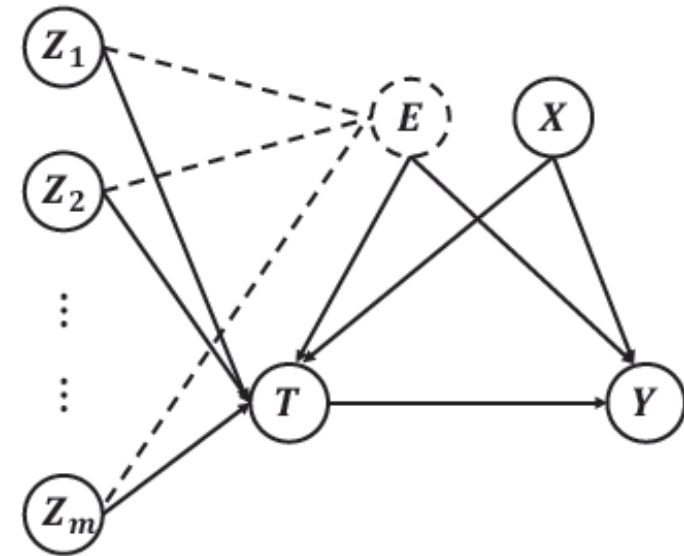
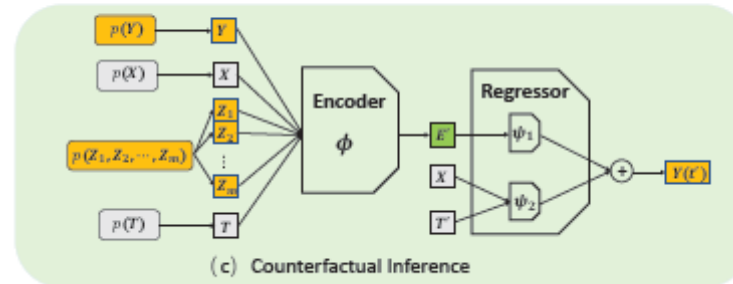
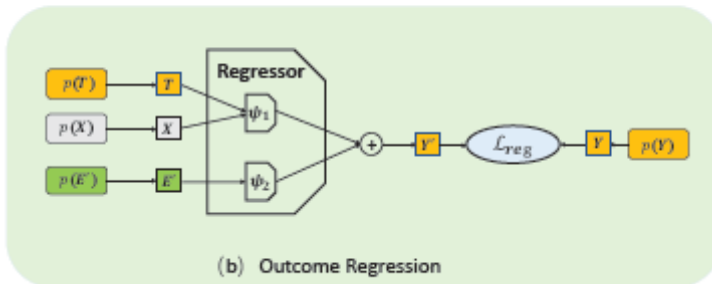
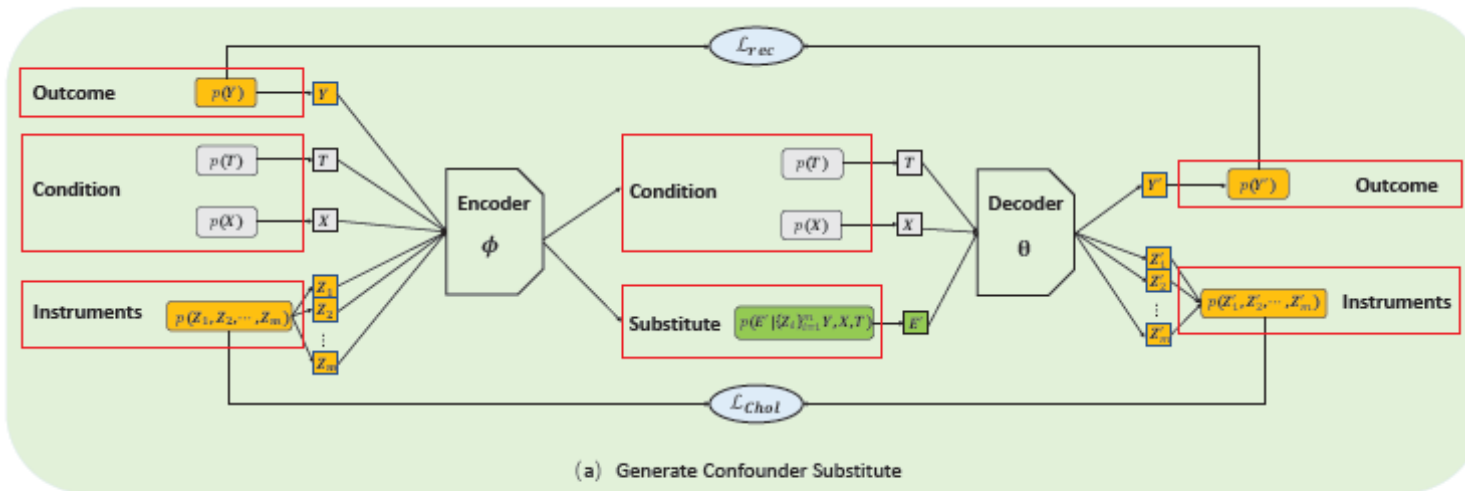
- Estimating Individual Causal Effect



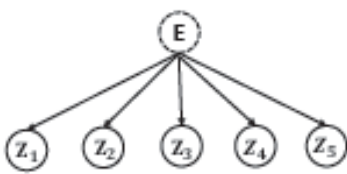
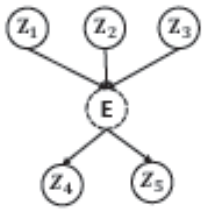
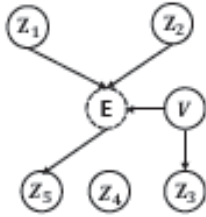
# CVAE-IV: Constructing substitute for confounders

## Conditional Independence Criteria:

- Generating  $E'$  such that  $Y \perp (Z_1, Z_2, \dots, Z_m) \mid E', T, X$
- $E'$  captures  $E$  rather than recovers  $E$

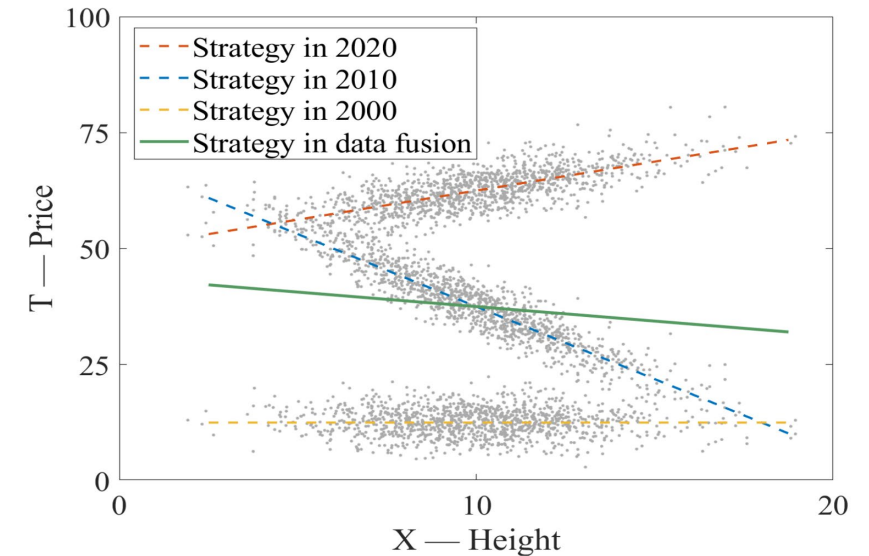
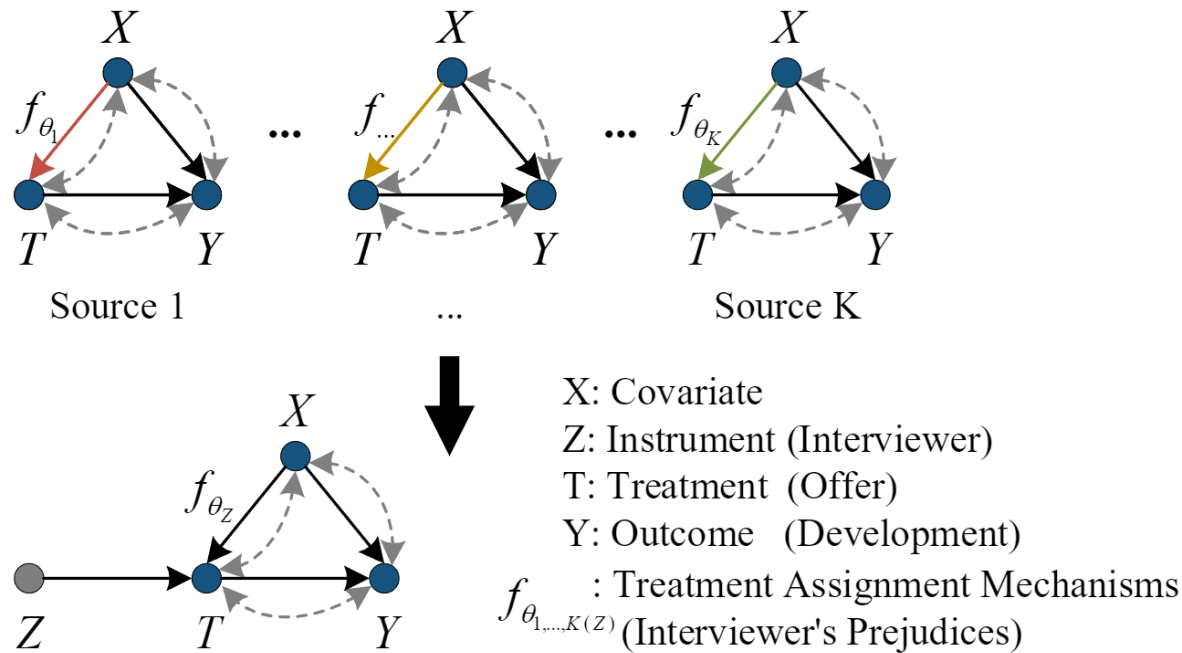


# Confounded IVs: Estimating Individualized Causal Effect with Confounded Instruments

Scenario	Function	Dim	DirectNN	2SLS-Ploy	KernelIV	DeepIV	CEVAE	ModeIV	Ours
S <sub>1</sub> : 	Linear	Low	2.091	55.437	8.565	1.977	9.959	3.120	<b>0.928</b>
		High	2.069	56.476	9.220	2.261	10.283	4.426	<b>0.858</b>
	Abs	Low	1.874	44.144	6.893	2.426	8.481	1.988	<b>0.697</b>
		High	1.671	41.731	7.998	1.675	9.921	2.088	<b>0.918</b>
	Square	Low	1.414	52.872	8.521	1.163	6.569	2.423	<b>0.490</b>
		High	1.602	41.731	9.988	1.545	7.039	2.018	<b>0.791</b>
S <sub>2</sub> : 	Linear	Low	1.788	45.287	8.922	1.910	7.621	3.314	<b>0.706</b>
		High	2.152	56.484	9.215	1.788	8.109	2.040	<b>0.601</b>
	Abs	Low	1.595	34.214	5.864	1.224	7.178	2.064	<b>0.562</b>
		High	0.836	42.181	4.030	0.835	8.015	2.134	<b>0.301</b>
	Square	Low	1.168	41.108	5.749	1.064	10.385	1.565	<b>0.125</b>
		High	1.650	51.129	6.030	1.568	9.015	2.031	<b>0.257</b>
S <sub>3</sub> : 	Linear	Low	1.650	41.009	6.617	2.023	7.234	3.638	<b>0.485</b>
		High	1.821	41.374	7.656	1.729	7.203	4.134	<b>0.608</b>
	Abs	Low	2.095	44.471	4.867	1.148	9.789	1.916	<b>0.689</b>
		High	1.590	41.132	6.330	1.484	7.293	1.972	<b>0.504</b>
	Square	Low	1.111	41.199	7.826	1.019	8.362	1.580	<b>0.516</b>
		High	2.179	41.915	6.407	1.442	8.825	1.709	<b>0.941</b>

# GIV: Learning Latent Group-IV from Data Fusion

## Multiple Treatment Assignment Mechanisms



- Data fusion with heterogenous treatment
- Unobserved confounders

□ Idea: Learning latent group instrument variable with Meta-EM algorithm

Anpeng Wu, Kun Kuang\*, Ruoxuan Xiong, Minqin Zhu, Yuxuan Liu, Bo Li, Furui Liu, Zhihua Wang, Fei Wu. Learning Instrumental Variable from Data Fusion for Treatment Effect Estimation, AAI, 2023.

# GIV: Learning Latent Group-IV from Data Fusion

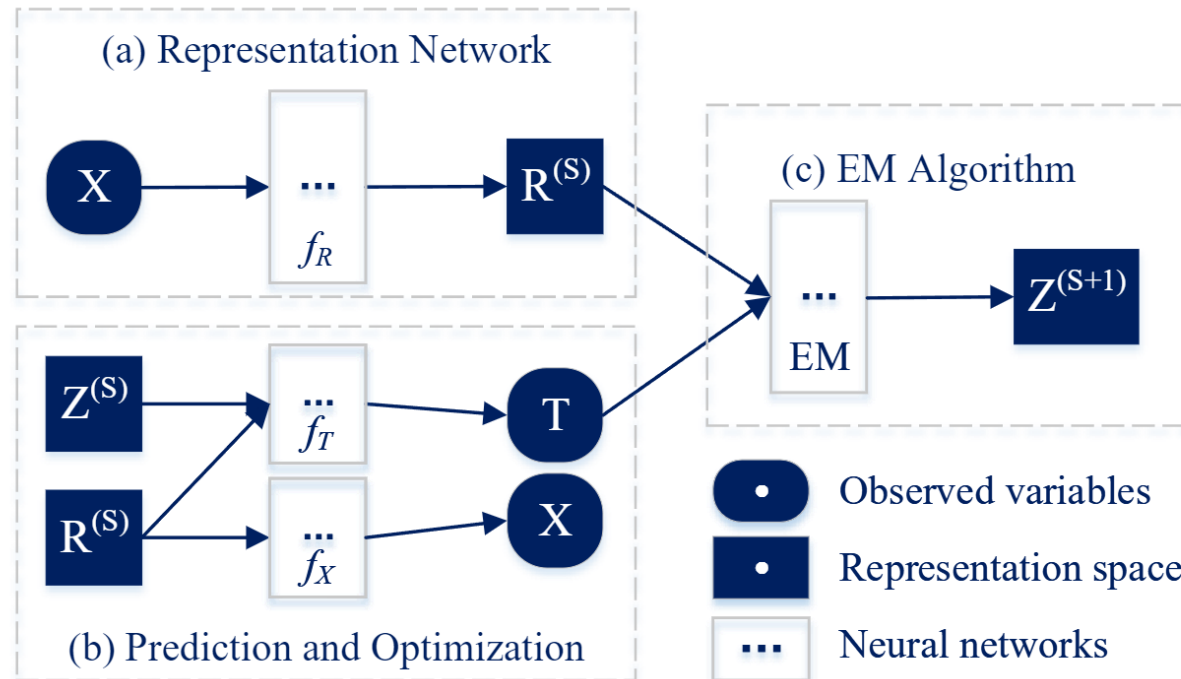


Figure 2: Overview of Meta-EM Architecture.

# GIV: Learning Latent Group-IV from Data Fusion

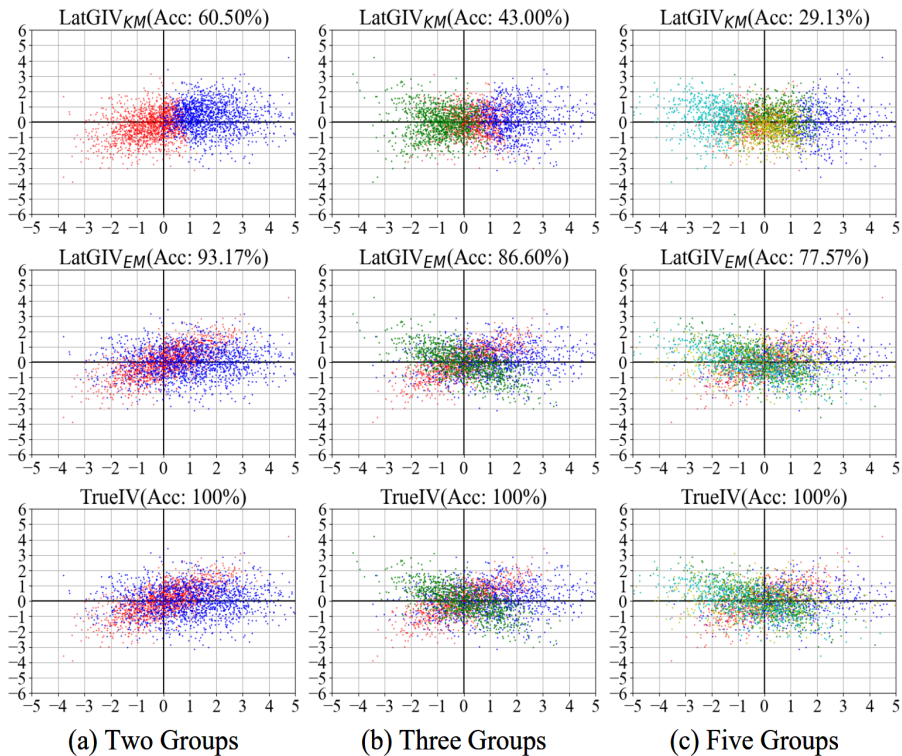
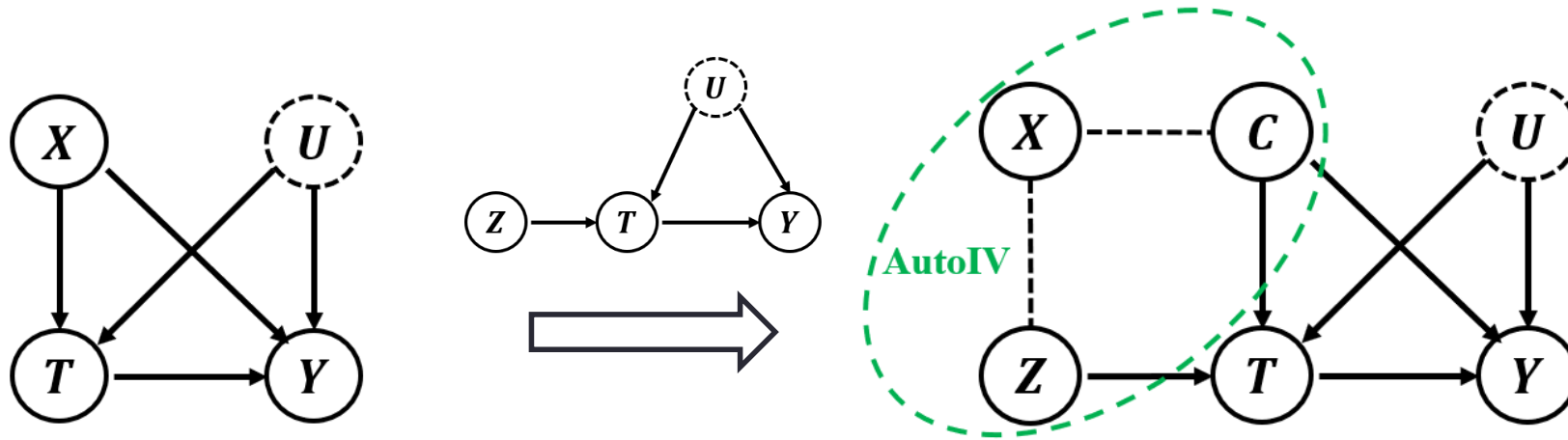


Table 6: The Full Results of MSE  $mean(std)$  of IHDP & PM-CMR Dataset with  $T=do(0)$

	IHDP Dataset								
	Poly2SLS	NN2SLS	KernelIV	DualIV	DeepIV	OneSIV	DFIV	DeepGMM	AGMM
NoneIV	0.350(0.129)	0.397(0.179)	0.322(0.121)	0.370(0.067)	0.365(0.081)	0.438(0.136)	1.130(0.367)	0.156(0.083)	0.147(0.057)
UAS	0.349(0.130)	0.443(0.152)	0.323(0.121)	0.395(0.090)	0.366(0.082)	0.398(0.117)	1.151(0.392)	0.159(0.097)	0.147(0.060)
WAS	0.349(0.130)	0.418(0.105)	0.322(0.122)	0.560(0.145)	0.367(0.086)	0.394(0.118)	1.138(0.389)	0.157(0.075)	0.154(0.062)
ModeIV	0.350(0.130)	0.453(0.186)	0.327(0.120)	0.561(0.152)	0.379(0.086)	0.423(0.123)	1.127(0.348)	0.168(0.051)	0.157(0.063)
AutoIV	>100	0.379(0.140)	0.323(0.122)	0.377(0.073)	0.365(0.088)	0.404(0.117)	1.106(0.350)	0.160(0.059)	0.147(0.058)
LatGIV <sub>KM</sub> *	0.166(0.069)	0.202(0.098)	0.203(0.058)	<b>0.411(0.078)</b>	0.309(0.075)	0.321(0.106)	1.046(0.263)	<b>0.126(0.062)</b>	0.117(0.047)
LatGIV <sub>EM</sub>	<b>0.048(0.023)</b>	<b>0.143(0.079)</b>	<b>0.094(0.049)</b>	0.423(0.083)	<b>0.294(0.072)</b>	<b>0.251(0.083)</b>	<b>1.024(0.282)</b>	<b>0.151(0.041)</b>	<b>0.090(0.037)</b>
TrueIV	<b>0.042(0.021)</b>	<b>0.120(0.037)</b>	<b>0.075(0.019)</b>	<b>0.422(0.074)</b>	<b>0.284(0.063)</b>	<b>0.237(0.062)</b>	<b>0.974(0.264)</b>	0.161(0.050)	<b>0.082(0.030)</b>
	PM-CMR Dataset								
	Poly2SLS	NN2SLS	KernelIV	DualIV	DeepIV	OneSIV	DFIV	DeepGMM	AGMM
NoneIV	0.206(0.054)	0.450(0.123)	0.198(0.072)	0.232(0.046)	0.358(0.091)	0.269(0.078)	1.099(0.220)	0.118(0.021)	0.098(0.029)
UAS	0.206(0.054)	0.477(0.151)	0.197(0.072)	0.259(0.059)	0.340(0.099)	0.264(0.083)	1.056(0.214)	0.128(0.036)	0.096(0.027)
WAS	0.206(0.054)	0.428(0.156)	0.227(0.063)	0.638(0.178)	0.360(0.072)	0.288(0.079)	1.074(0.247)	0.141(0.042)	0.147(0.058)
ModeIV	0.206(0.054)	0.455(0.139)	0.198(0.074)	0.448(0.115)	0.356(0.098)	0.278(0.090)	1.105(0.228)	0.119(0.047)	0.099(0.029)
AutoIV	0.205(0.055)	0.416(0.106)	0.196(0.072)	0.316(0.125)	0.346(0.080)	0.289(0.092)	1.063(0.223)	0.133(0.045)	0.099(0.031)
LatGIV <sub>KM</sub> *	0.265(0.159)	0.523(0.141)	0.192(0.066)	0.303(0.069)	0.312(0.073)	0.264(0.081)	<b>1.064(0.246)</b>	0.128(0.031)	0.094(0.024)
LatGIV <sub>EM</sub>	<b>0.091(0.041)</b>	<b>0.248(0.058)</b>	<b>0.170(0.079)</b>	<b>0.303(0.068)</b>	<b>0.275(0.076)</b>	<b>0.245(0.088)</b>	<b>1.032(0.229)</b>	<b>0.121(0.027)</b>	<b>0.066(0.015)</b>
TrueIV	<b>0.036(0.008)</b>	<b>0.172(0.072)</b>	<b>0.079(0.017)</b>	<b>0.303(0.062)</b>	<b>0.171(0.044)</b>	<b>0.172(0.044)</b>	1.100(0.261)	<b>0.120(0.031)</b>	<b>0.037(0.007)</b>

Figure 2: Reconstruction Accuracy of the Group IV with Different Group Number.

# AutoIV: Counterfactual Learning with Unobserved Confounders via Automatically generating IVs



## Conditions of IV

- Relevance:  $P(T|Z) \neq P(T)$
- **Exclusion:**  $P(Y|Z, T, C) \neq P(Y|T, C)$
- Unconfounded:  $Z \perp C$



Mutual Information  
Representation Learning

**But exclusion might not be satisfied**

# AutoIV: Counterfactual Learning with Unobserved Confounders via Automatically generating IVs

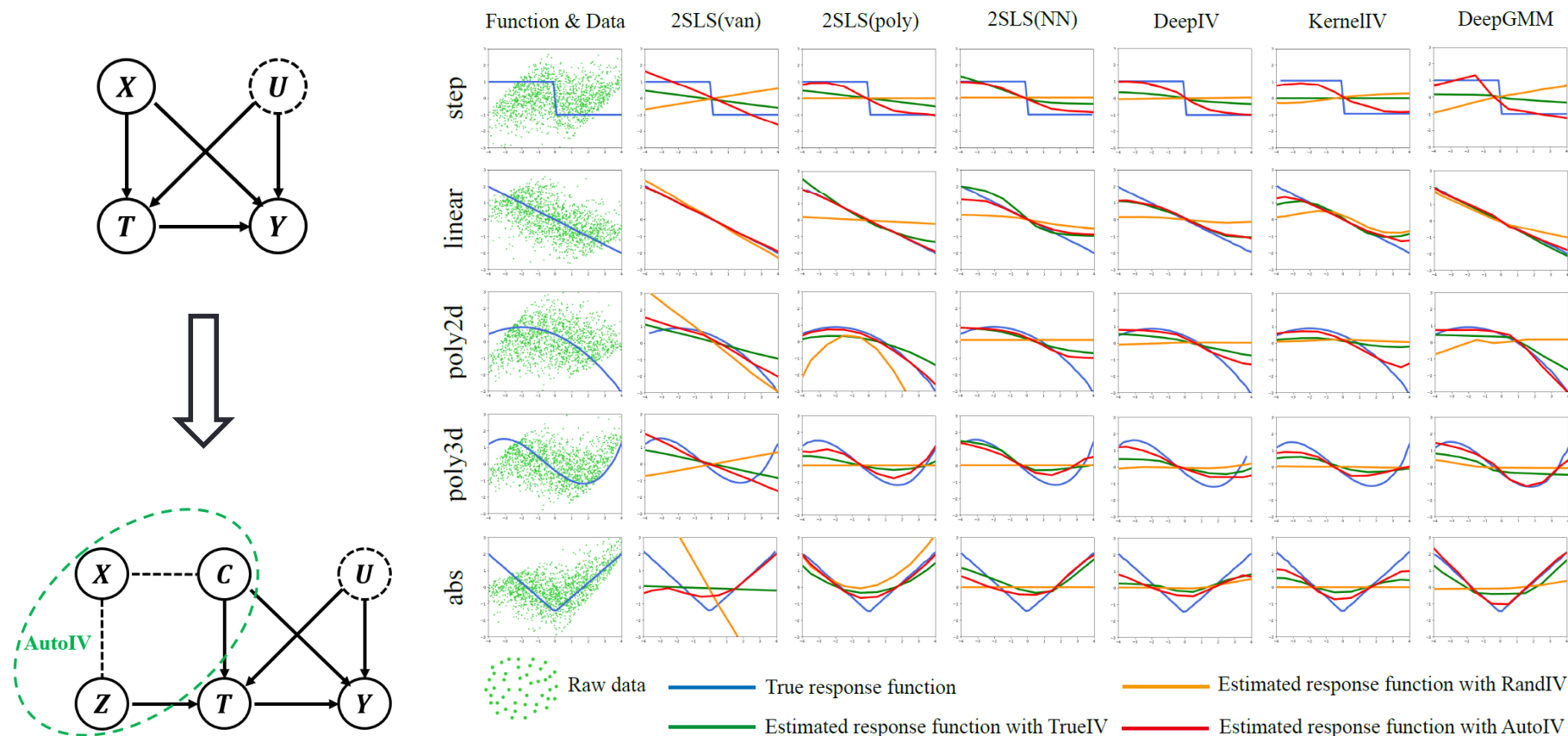


Figure 2: Response function prediction in low-dimensional scenarios.

Yuan J, Wu A, Kuang K, et al. Auto IV: Counterfactual Prediction via Automatic Instrumental Variable Decomposition[J]. TKDD, 2022.

# Summary

- Confounder Balanced IV (CB-IV):
  - **With pre-defined IV**, simultaneously address biases from both observed and unobserved confounders
- Confounded IV:
  - **With invalid IVs**, construct substitute for confounders
- Group IV (G-IV):
  - **Without IV**, learn data heterogeneity as latent group IV
- Auto IV:
  - **Without IV**, generate representations to serve the role of IVs

# IVs in Causal Inference and Machine Learning

## Instrumental Variables in Causal Inference and Machine Learning: A Survey

Anpeng Wu, Kun Kuang, Ruoxuan Xiong, Fei Wu, *Senior Member, IEEE*

**Abstract**—Causal inference is the process of using assumptions, study designs, and estimation strategies to draw conclusions about the causal relationships between variables based on data. This allows researchers to better understand the underlying mechanisms at work in complex systems and make more informed decisions. In many settings, we may not fully observe all the confounders that affect both the treatment and outcome variables, complicating the estimation of causal effects. To address this problem, a growing literature in both causal inference and machine learning proposes to use Instrumental Variables (IV). This paper serves as the first effort to systematically and comprehensively introduce and discuss the IV methods and their applications in both causal inference and machine learning. First, we provide the formal definition of IVs and discuss the identification problem of IV regression methods under different assumptions. Second, we categorize the existing work on IV methods into three streams according to the focus on the proposed methods, including two-stage least squares with IVs, control function with IVs, and evaluation of IVs. For each stream, we present both the classical causal inference methods, and recent developments in the machine learning literature. Then, we introduce a variety of applications of IV methods in real-world scenarios and provide a summary of the available datasets and algorithms. Finally, we summarize the literature, discuss the open problems and suggest promising future research directions for IV methods and their applications. We also develop a toolkit of IVs methods reviewed in this survey at <https://github.com/causal-machine-learning-lab/mliv>.

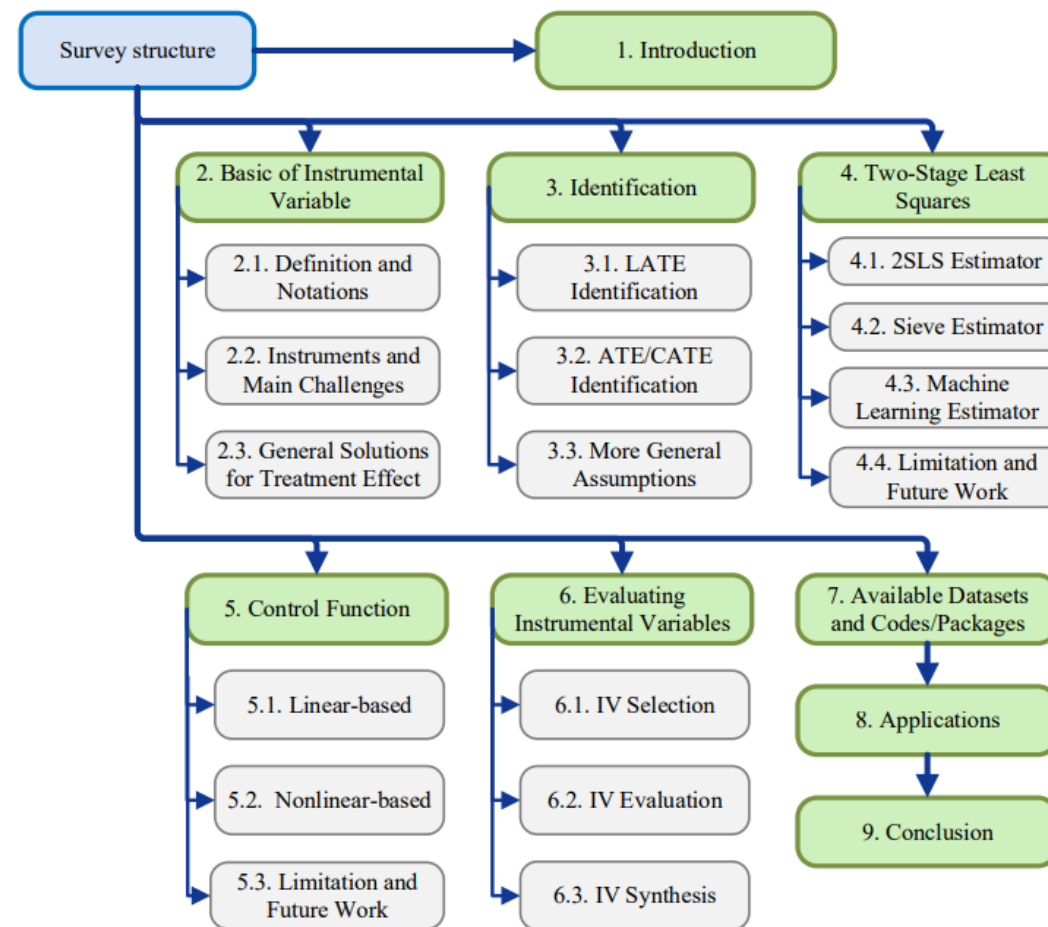


Fig. 2: Outline of the Survey.

Anpeng Wu, Kun Kuang, Ruoxuan Xiong, Fei Wu, Instrumental Variables in Causal Inference and Machine Learning: A Survey[J]. arXiv preprint arXiv:2212.05778, 2022.

# IVs in Causal Inference and Machine Learning

## mliv

```
from mliv.dataset.demand import gen_data
from mliv.utils import CausalDataset
gen_data()
data = CausalDataset('./Data/Demand/0.5_1.0_0.0_10000/1/')

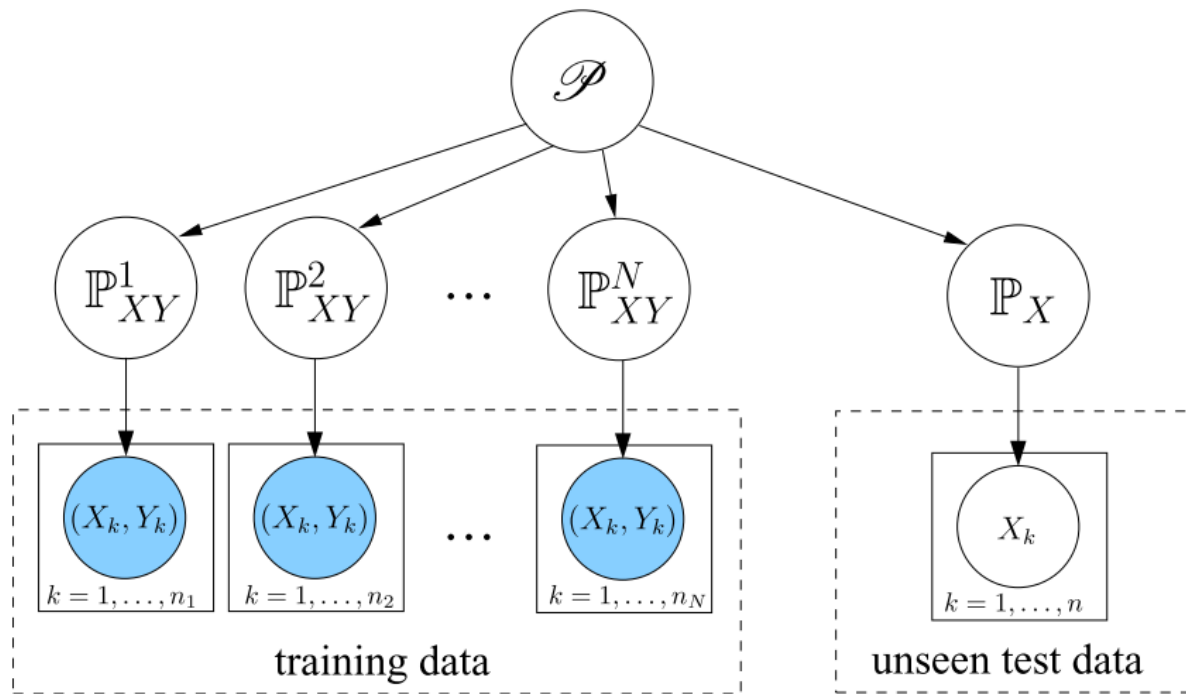
from mliv.inference import Vanilla2SLS
from mliv.inference import Poly2SLS
from mliv.inference import NN2SLS
from mliv.inference import OneSIV
from mliv.inference import KernelIV
from mliv.inference import DualIV
from mliv.inference import DFL
from mliv.inference import AGMM
from mliv.inference import DeepGMM
from mliv.inference import DFIV
from mliv.inference import DeepIV          # Tensorflow & keras

for mod in [OneSIV, KernelIV, DualIV, DFL, AGMM, DeepGMM, DFIV, Vanilla2SLS, Poly2SLS, NN2SLS]:
    model = mod()
    model.config['num'] = 100
    model.config['epochs'] = 10
    model.fit(data)

print(mod)
```

Anpeng Wu, Kun Kuang, Ruoxuan Xiong, Fei Wu, Instrumental Variables in Causal Inference and Machine Learning: A Survey[J]. arXiv preprint arXiv:2212.05778, 2022.

# 基于工具变量回归的因果可泛化学习



- Given data from different observed environments  $e \in \mathcal{E}$  :

$$(X^e, Y^e) \sim F^e, \quad e \in \mathcal{E}$$

- The task is to predict  $Y$  given  $X$  such that the prediction works well (is “robust”) for “all possible” (including unseen) environments

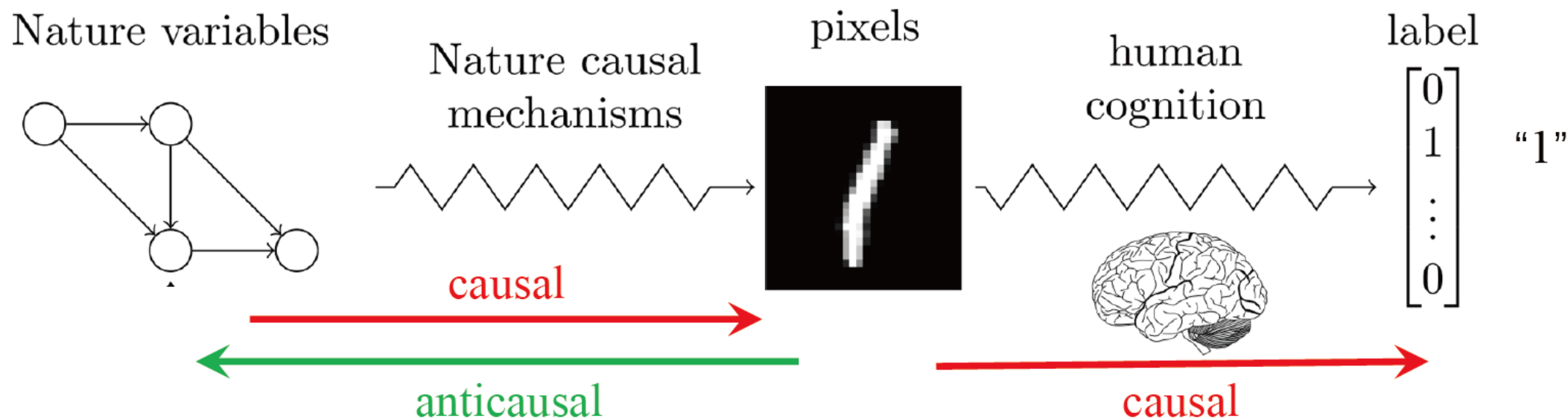
Domain Generation  
Invariant Causal Prediction  
Causal Transfer Learning

## 基于工具变量回归的因果可泛化学习

Data generating process (DGP) assumption for each domain  $m$ :

$X^m$ : input features;  $Y^m$ : labels;

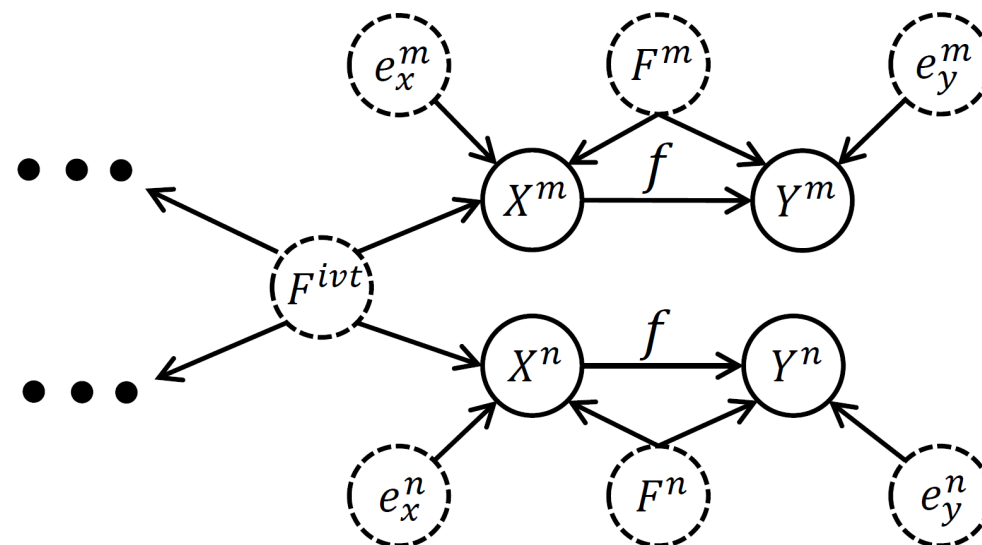
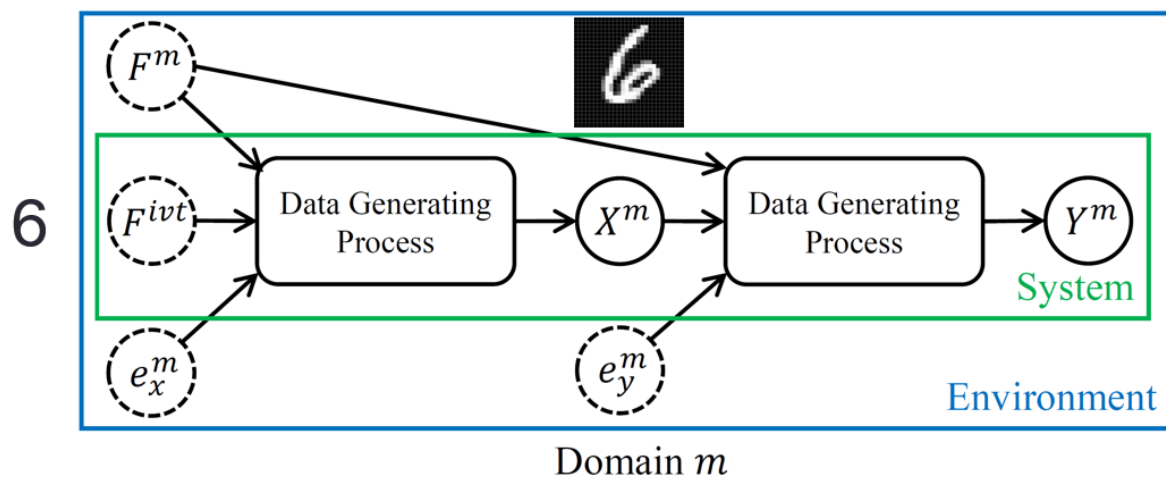
$F^{ivt}$ : domain-invariant factor;  $F^m$ : domain-specific factor;  $e_x^m, e_y^m$ : error term.



# 基于工具变量回归的因果可泛化学习

**Assumption 1.** Data distributions of different domains satisfy the data generating process and causal graph, where only **the factor  $F^{ivt}$  and relationship  $f$  are invariant** across domains.

**Theorem 1.** For any two domains  $m$  and  $n$ , if  $m \neq n$ , then  **$X^n$  is a valid instrumental variable of domain  $m$ .**



## 基于工具变量回归的因果可泛化学习

By taking expectation of  $Y^m$  conditional on  $X^n$ , we have:

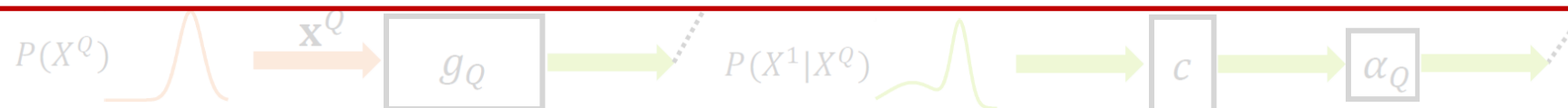
$$\begin{aligned}\mathbb{E}[Y^m|X^n] &= \mathbb{E}[f(X^m)|X^n] + \mathbb{E}[F^m|X^n] + \mathbb{E}[e_y^m|X^n] \\ &= \int f(X^m)dP(X^m|X^n)\end{aligned}$$

**First stage:** estimate  $P(X^m|X^n)$

**Second stage:** use the approximation of  $P(X^m|X^n)$  and  $Y^m$  to learn invariant function  $f(X^m)$



**Remark: With IV regression, we only require the label information from one domain**



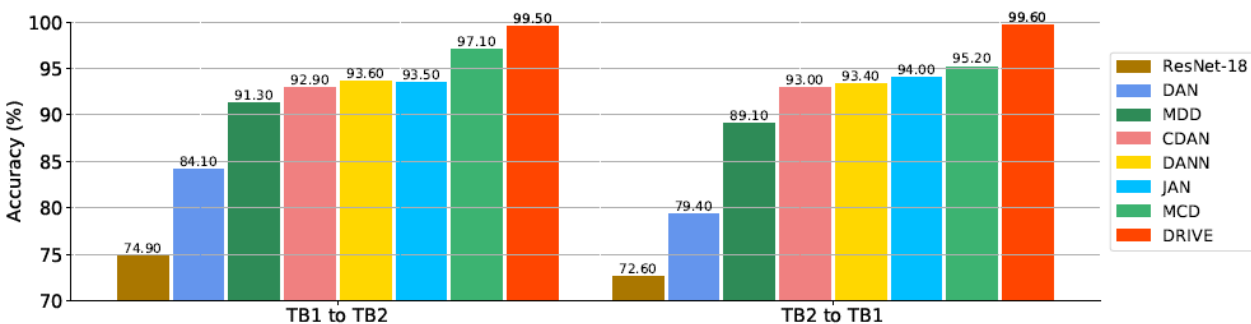
# 基于工具变量回归的因果可泛化学习

## Results on public domain generalization datasets.

RESULTS (%) FOR DOMAIN GENERALIZATION ON OFFICE-HOME DATASET.

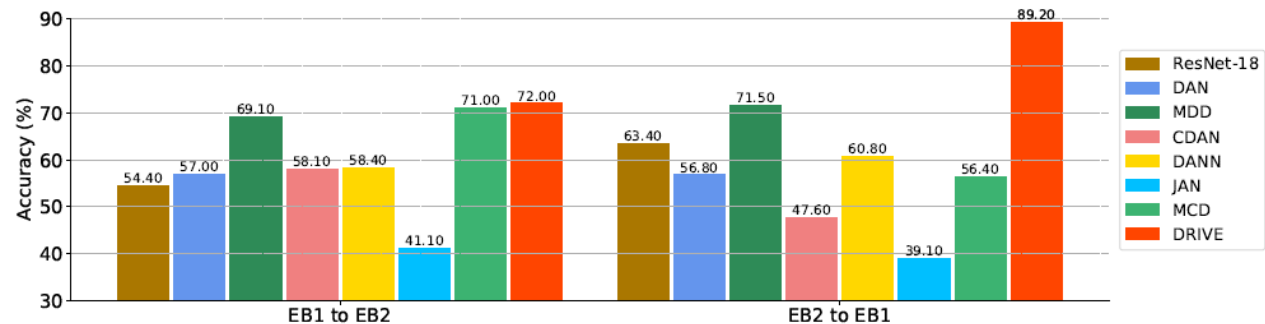
Methods	Art	Clipart	Product	Real-World	Average
DeepAll [8]	52.15	45.86	70.86	73.15	60.51
JiGen [8]	53.04	47.51	71.47	72.79	61.20
DSON [44]	59.37	44.70	71.84	74.68	62.90
RSC [16]	58.42	<b>47.90</b>	71.63	74.54	63.12
DRIVE w/o IV	55.53 ± 0.21	45.92 ± 0.50	71.64 ± 0.35	74.49 ± 0.05	61.90 ± 0.20
DRIVE w/o pre	59.30 ± 0.06	47.65 ± 0.30	72.03 ± 0.57	75.55 ± 0.24	63.63 ± 0.11
DRIVE	<b>60.40 ± 0.26</b>	47.73 ± 0.28	<b>72.63 ± 0.18</b>	<b>76.14 ± 0.10</b>	<b>64.23 ± 0.09</b>

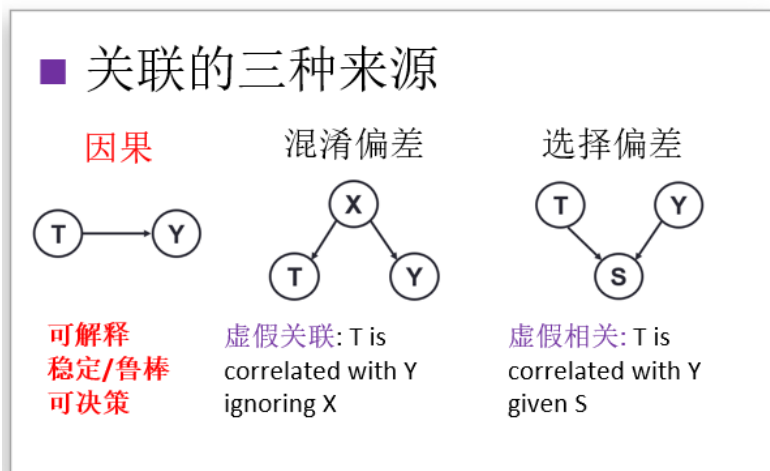
## Results on biased datasets.



RESULTS (%) FOR DOMAIN GENERALIZATION ON PACS DATASET.

Methods	Art	Cartoon	Photo	Sketch	Average
DeepAll [8]	78.96	72.93	96.28	70.59	79.94
JiGen [8]	79.42	75.25	96.03	71.35	80.51
MASF [10]	80.29	77.17	94.99	71.69	81.04
DGER [61]	80.70	76.40	96.65	71.77	81.38
Epi-FCR [21]	82.1	77.0	93.9	73.0	81.5
MMLD [36]	81.28	77.16	96.09	72.29	81.83
EISNet [49]	81.89	76.44	95.93	74.33	82.15
L2A-OT [63]	83.3	78.2	96.2	73.6	82.8
DDAIG [62]	<b>84.2</b>	78.1	95.3	74.7	83.1
DRIVE w/o IV	79.40 ± 0.10	76.93 ± 0.09	95.75 ± 0.10	74.44 ± 0.07	81.63 ± 0.03
DRIVE w/o pre	81.95 ± 0.25	77.55 ± 0.31	96.64 ± 0.34	75.65 ± 0.10	82.95 ± 0.14
DRIVE	83.36 ± 0.70	<b>78.76 ± 0.08</b>	<b>96.87 ± 0.18</b>	<b>78.68 ± 0.96</b>	<b>84.42 ± 0.11</b>



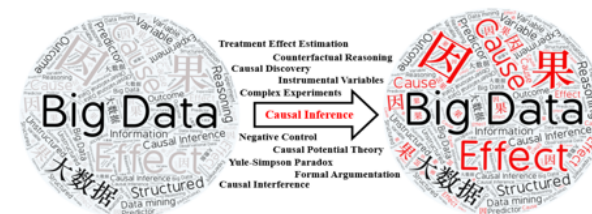


机器学习赋能  
因果推理



因果启发机器学习

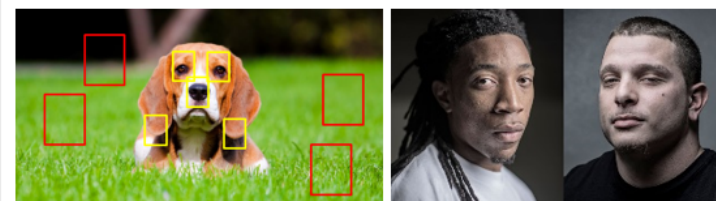
### ■ 大数据因果推理



### ■ 因果可信学习

可解释性、稳定性

公平性、可决策性



# Thank You!

Kun Kuang

[kunkuang@zju.edu.cn](mailto:kunkuang@zju.edu.cn)

Homepage: <https://kunkuang.github.io/>