Learning Decision Policies with Instrumental Variables through Double Machine Learning

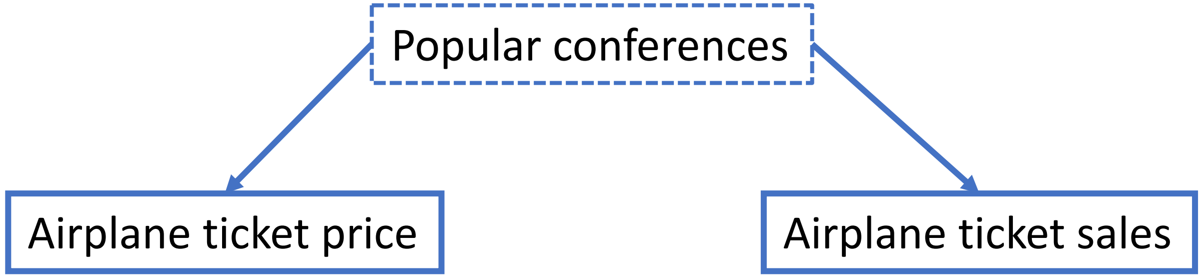
ICML 2024, https://arxiv.org/abs/2405.08498

Bill Daqian Shao

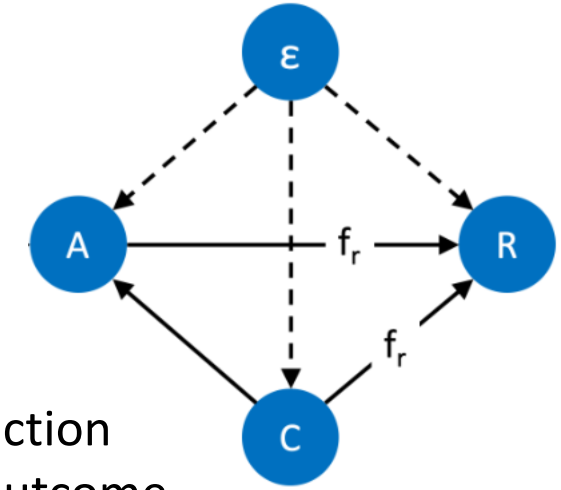
Data-driven Decision Making

- A common issue for learning from offline observational data is the existence of spurious correlations: which are relationships between variables that appear to be causal, but in fact are not.

- For example:



Data-driven Decision Making

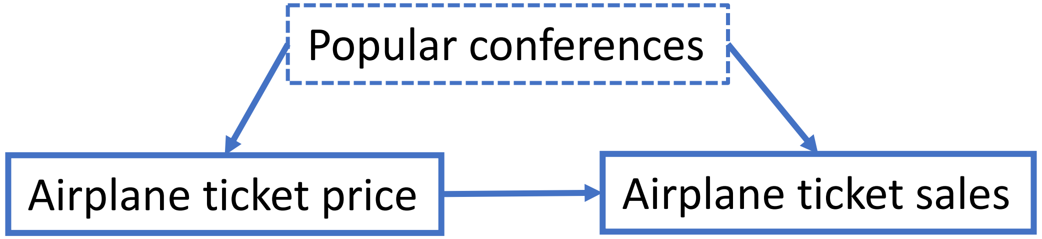


A: action
R: outcome
C: context

The causal structural model for the outcome R is specified as follows:

$$R := f_r(C, A) + \epsilon, \quad \mathbb{E}[\epsilon] = 0, \quad \mathbb{E}[\epsilon|A, C] \neq 0,$$

It has been shown (Bareinboim & Pearl, 2012) that we cannot learn the causal effect of actions in the presence of hidden confounders without structural assumptions


Popular conferences

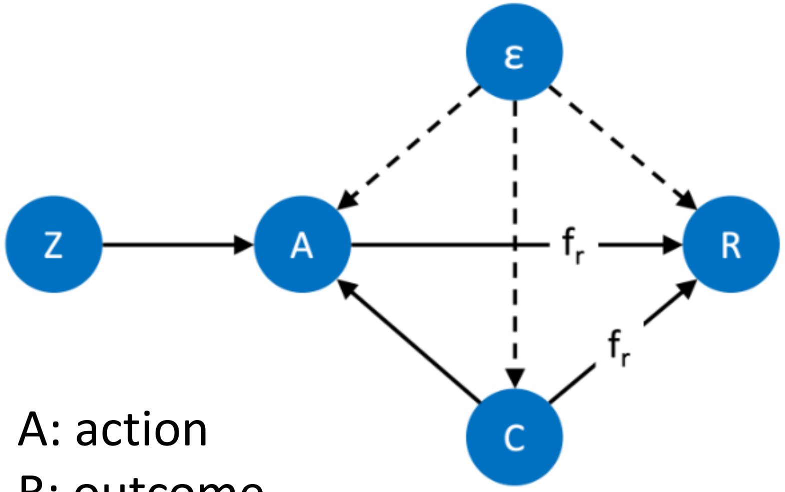
Airplane ticket price

Airplane ticket sales

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF OXFORD

Contextual Instrumental Variable Setting



A: action

R: outcome

C: context

Z: Instrument

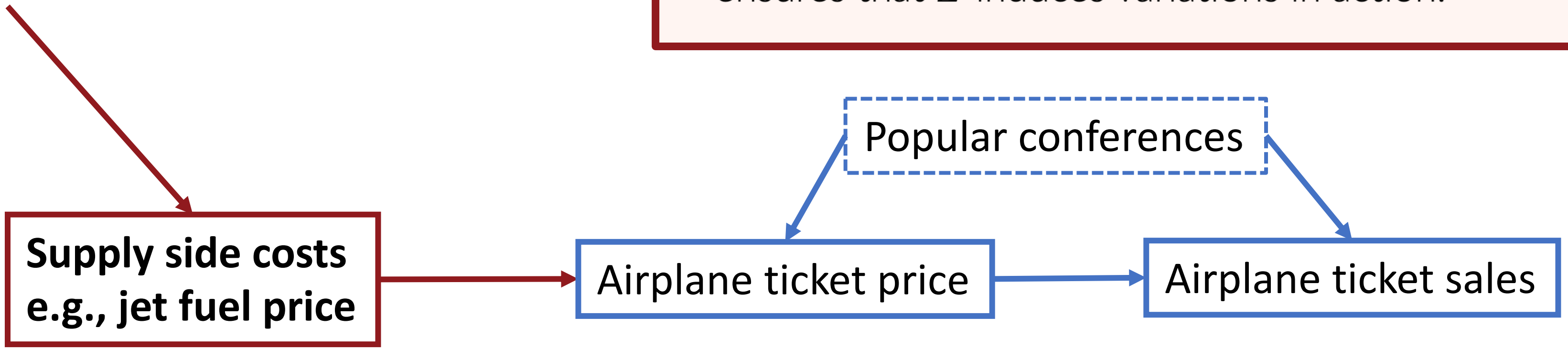
IVs are random variables independent to the hidden confounder and only directly affect the action.

How to choose a good instrument Z?

Minimal conditions to identify the causal effect:

Axiom (A) The unconfounded instrument restriction, $Z \perp\!\!\!\perp \epsilon \mid C$, i.e., the instrument Z is independent of the hidden confounder ϵ conditional on C.

Axiom (B) The relevance condition, $\mathbb{P}(A|C, Z)$ is not constant in Z, i.e., it ensures that Z induces variations in action.



Popular conferences

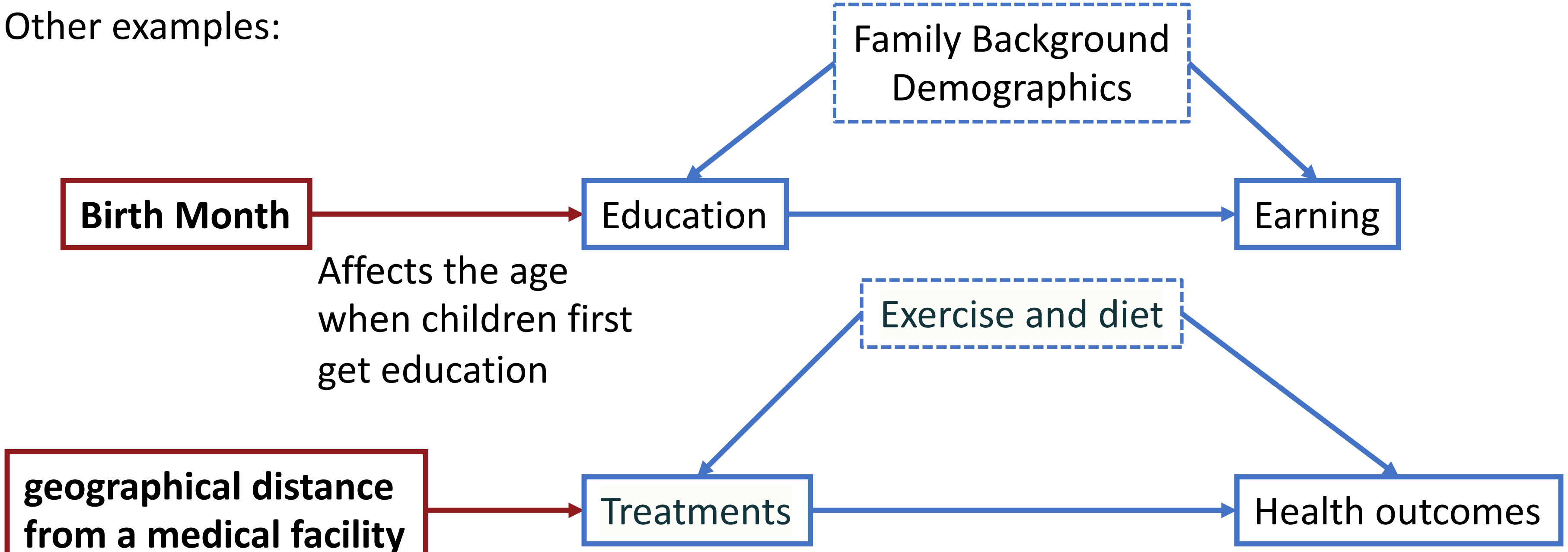
Supply side costs e.g., jet fuel price

Airplane ticket price

Airplane ticket sales

Contextual Instrumental Variable Setting

Other examples:


Family Background Demographics
Birth Month
Education
Earning
Exercise and diet
geographical distance from a medical facility
Treatments
Health outcomes

Affects the age when children first get education

People less willing to receive treatment due to the distance

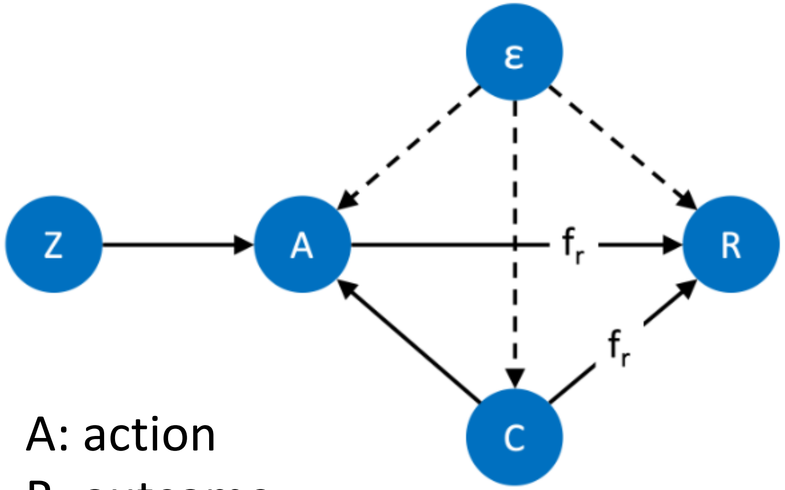
How to choose a good instrument Z?

Minimal conditions to identify the causal effect:

Axiom (A) The unconfounded instrument restriction, $Z \perp\!\!\!\perp \epsilon \mid C$, i.e., the instrument Z is independent of the hidden confounder ϵ conditional on C.

Axiom (B) The relevance condition, $\mathbb{P}(A|C, Z)$ is not constant in Z, i.e., it ensures that Z induces variations in action.

Contextual Instrumental Variable (IV)



A: action
R: outcome
C: context
Z: Instrument

$$R := f_r(C, A) + \epsilon, \quad \mathbb{E}[\epsilon] = 0, \quad \mathbb{E}[\epsilon | A, C] \neq 0,$$

Our goal is to learn the counterfactual prediction function:

$$h_0(C, A) := f_r(C, A) + \mathbb{E}[\epsilon | C] = \mathbb{E}[R | do(A), C],$$

Why learn the counterfactual prediction function h_0?

1. Learning h_0 allows us to compare between different actions when given a context C as $h_0(C, a_1) - h_0(C, a_2) = f_r(C, a_1) - f_r(C, a_2)$ for all $a_1, a_2 \in \mathcal{A}$, and in particular, $\arg\max_{a \in \mathcal{A}} h_0(C, a) = \arg\max_{a \in \mathcal{A}} f_r(C, a)$.
2. For any policy $\pi : \mathcal{C} \to \Delta(\mathcal{A})$, let $V(\pi) := \mathbb{E}_{c \sim \mathcal{P}_{\text{test}}}[h_0(c, \pi(c))]$ denote the value function of π, where $\mathcal{P}_{\text{test}}$ may differ from $\mathcal{P}_{\text{train}}$. The optimal policy can be retrieved by $\widehat{\pi}(a) = \arg\max_{a \in \mathcal{A}} \widehat{h}_0(c, a)$.

IV Regression Methods

$$R := f_r(C, A) + \epsilon$$

We can identify h$_0$ using:

$$\begin{aligned}\boxed{\mathbb{E}[R|C,Z]} &= \mathbb{E}\Big[f_r(C,A) + \mathbb{E}[\epsilon|C]\Big|C,Z\Big] \\ &= \mathbb{E}[h_0(C,A)|C,Z] \\ &= \int h_0(C,A)\boxed{\mathbb{P}(A|C,Z)}dA,\end{aligned}$$

Both observable

This is an inverse problem for definite integrals that requires the derivation of a function inside the definite integral based on numerical integral values, thus can't be solve analytically.

Existing two-stage IV regression methods:

First stage learns $\mathbb{E}[h(C,A)|c,z]$

Second stage learns $\min_{h\in\mathcal{H}} \mathbb{E}[(R - \mathbb{E}[h(C,A)|C,Z])^2].$

Recent non-linear IV regression methods use ML estimators for both stages.

Double Machine Learning

However, both regularisation and overfitting cause heavy bias (Chernozhukov et al., 2018) in estimating h_0 when the first stage estimator is naively plugged in, which causes slow convergence of the causal function estimator.

Double Machine Learning (DML)[3] is a statistical technique that debiases two-stage estimators and provides fast convergence rate guarantees for general two-stage regressions.

DML considers the problem of estimating a function of interest h as a solution to an equation (or score) of the form $\mathbb{E}[\psi(\mathcal{D}; h, \eta)] = 0,$ where η are nuisances parameters. Crucially, DML requires the score ψ to be Neyman orthogonal[4], which requires the Gateaux derivative to be zero:

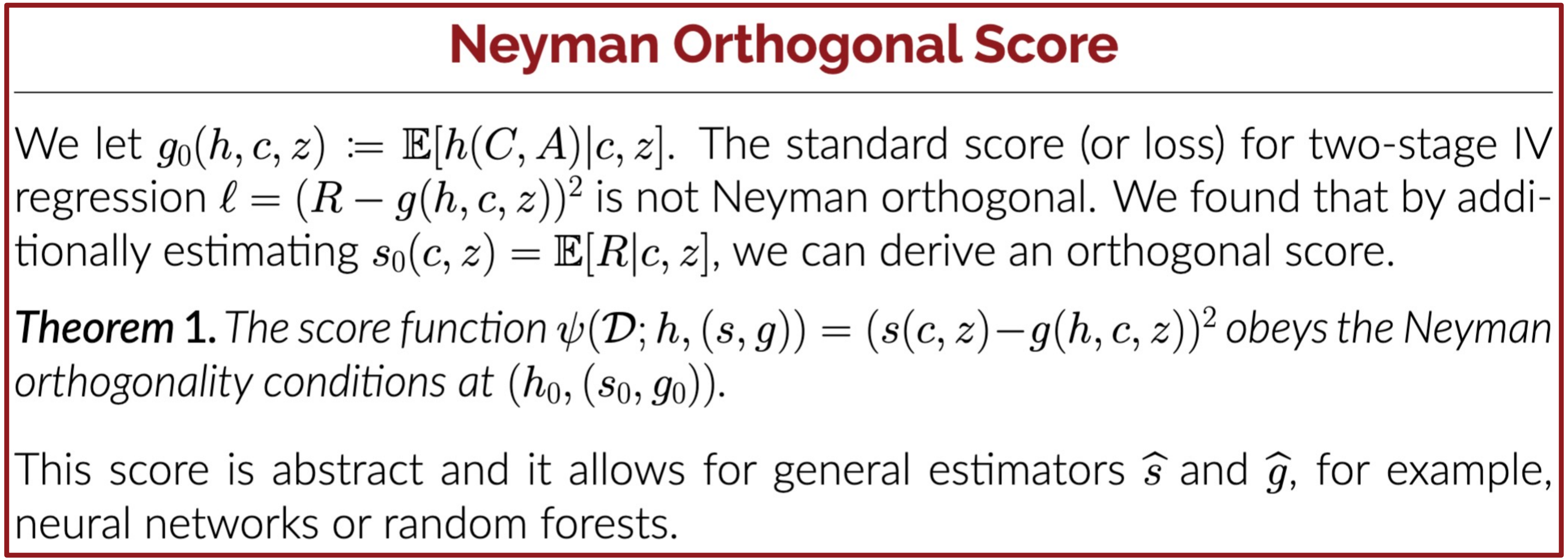
$$\frac{\partial}{\partial r}\bigg|_{r=0} \mathbb{E}[\psi(\mathcal{D}; h_0, \eta_0 + r\eta)] = 0,$$

Intuition: small changes of the nuisance parameter do not significantly affect the score function around the true parameter h_0

DML-IV

We first need to find a Neyman orthogonal score function for the IV regression problem

$$\mathbb{E}[\psi(\mathcal{D}; h, \eta)] = 0$$



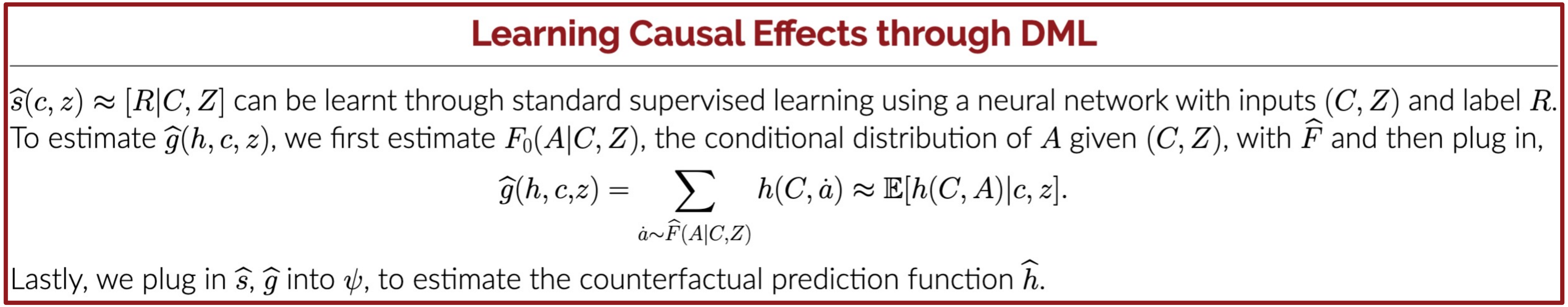
Neyman Orthogonal Score

We let $g_0(h, c, z) := \mathbb{E}[h(C, A)|c, z]$. The standard score (or loss) for two-stage IV regression $\ell = (R - g(h, c, z))^2$ is not Neyman orthogonal. We found that by additionally estimating $s_0(c, z) = \mathbb{E}[R|c, z]$, we can derive an orthogonal score.

Theorem 1. *The score function $\psi(\mathcal{D}; h, (s, g)) = (s(c, z) - g(h, c, z))^2$ obeys the Neyman orthogonality conditions at $(h_0, (s_0, g_0))$.*

This score is abstract and it allows for general estimators $\widehat{s}$ and $\widehat{g}$, for example, neural networks or random forests.

DML-IV



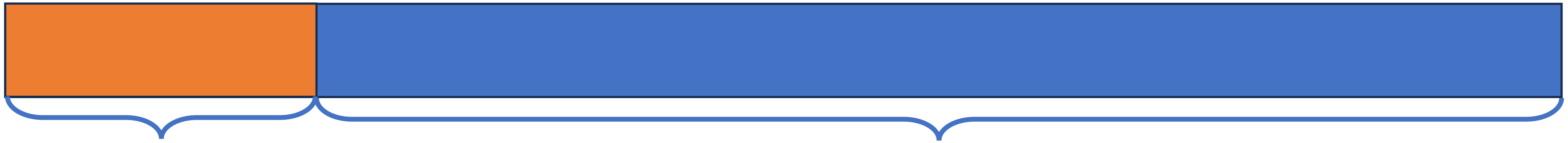
Learning Causal Effects through DML

$\widehat{s}(c,z) \approx [R|C,Z]$ can be learnt through standard supervised learning using a neural network with inputs (C,Z) and label R. To estimate $\widehat{g}(h,c,z)$, we first estimate $F_0(A|C,Z)$, the conditional distribution of A given (C,Z), with $\widehat{F}$ and then plug in,

$$\widehat{g}(h,c,z) = \sum_{\dot{a}\sim\widehat{F}(A|C,Z)} h(C,\dot{a}) \approx \mathbb{E}[h(C,A)|c,z].$$

Lastly, we plug in $\widehat{s}$, $\widehat{g}$ into ψ, to estimate the counterfactual prediction function $\widehat{h}$.

K-fold cross fitting

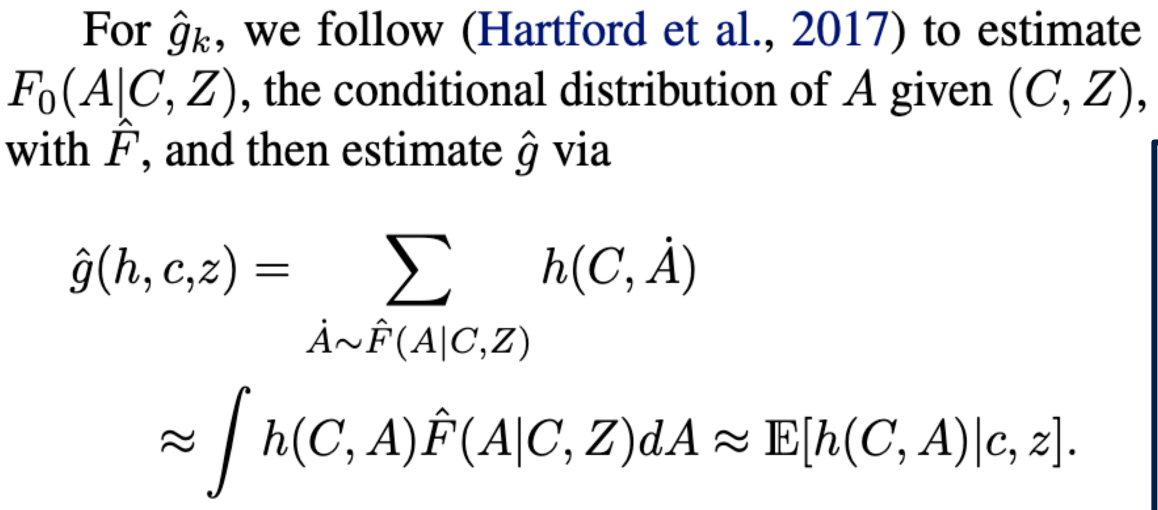


For each fold of data that is used to fit h

The rest of the data is used to train the nuisances parameters: s and g

DML-IV

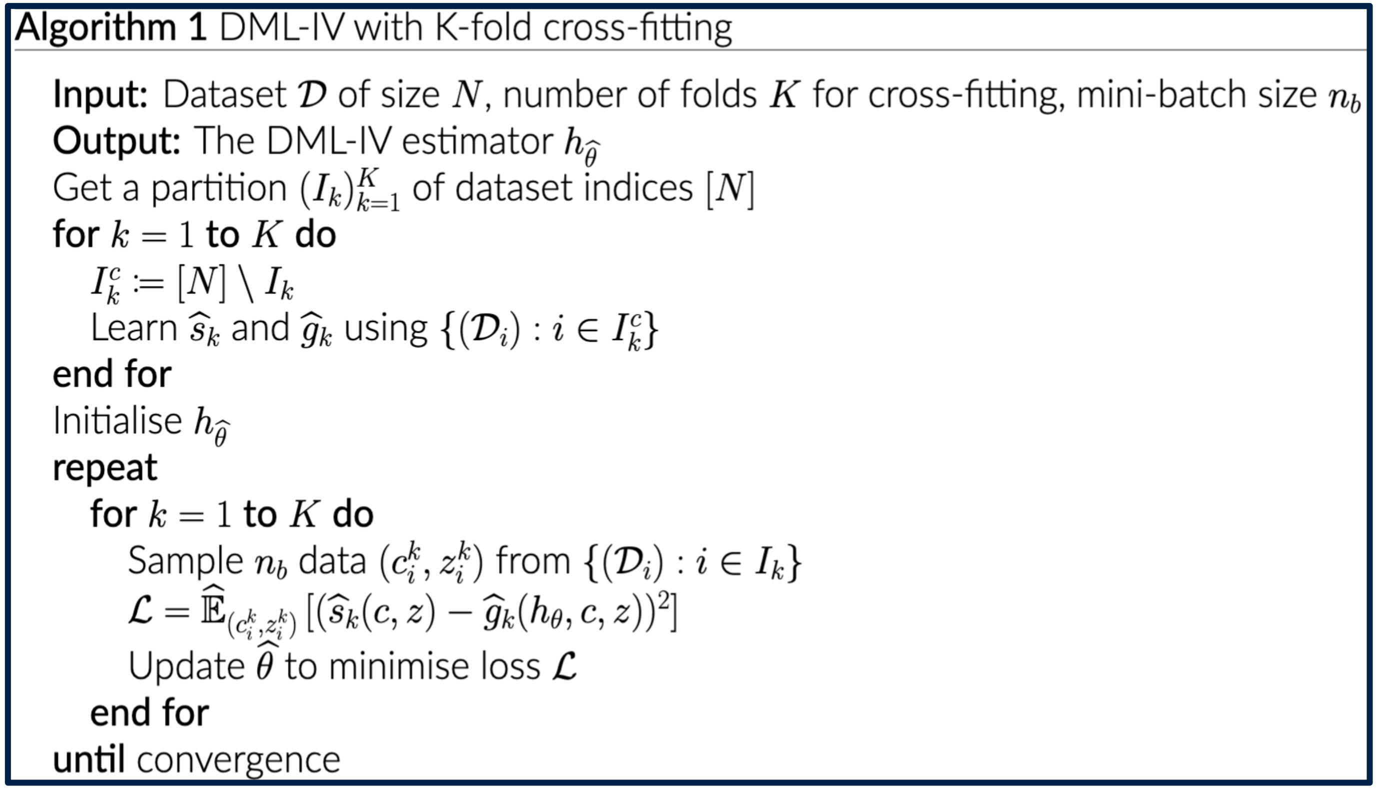
For each fold we use data from $I_k^c := [N] \setminus I_k$ to learn $\hat{s}_k \approx \mathbb{E}[R|C,Z]$ via supervised learning



For $\hat{g}_k$, we follow (Hartford et al., 2017) to estimate $F_0(A|C,Z)$, the conditional distribution of A given (C,Z), with $\hat{F}$, and then estimate $\hat{g}$ via

$$\hat{g}(h,c,z) = \sum_{\dot{A}\sim\hat{F}(A|C,Z)} h(C,\dot{A})$$

$$\approx \int h(C,A)\hat{F}(A|C,Z)dA \approx \mathbb{E}[h(C,A)|c,z].$$



Algorithm 1 DML-IV with K-fold cross-fitting

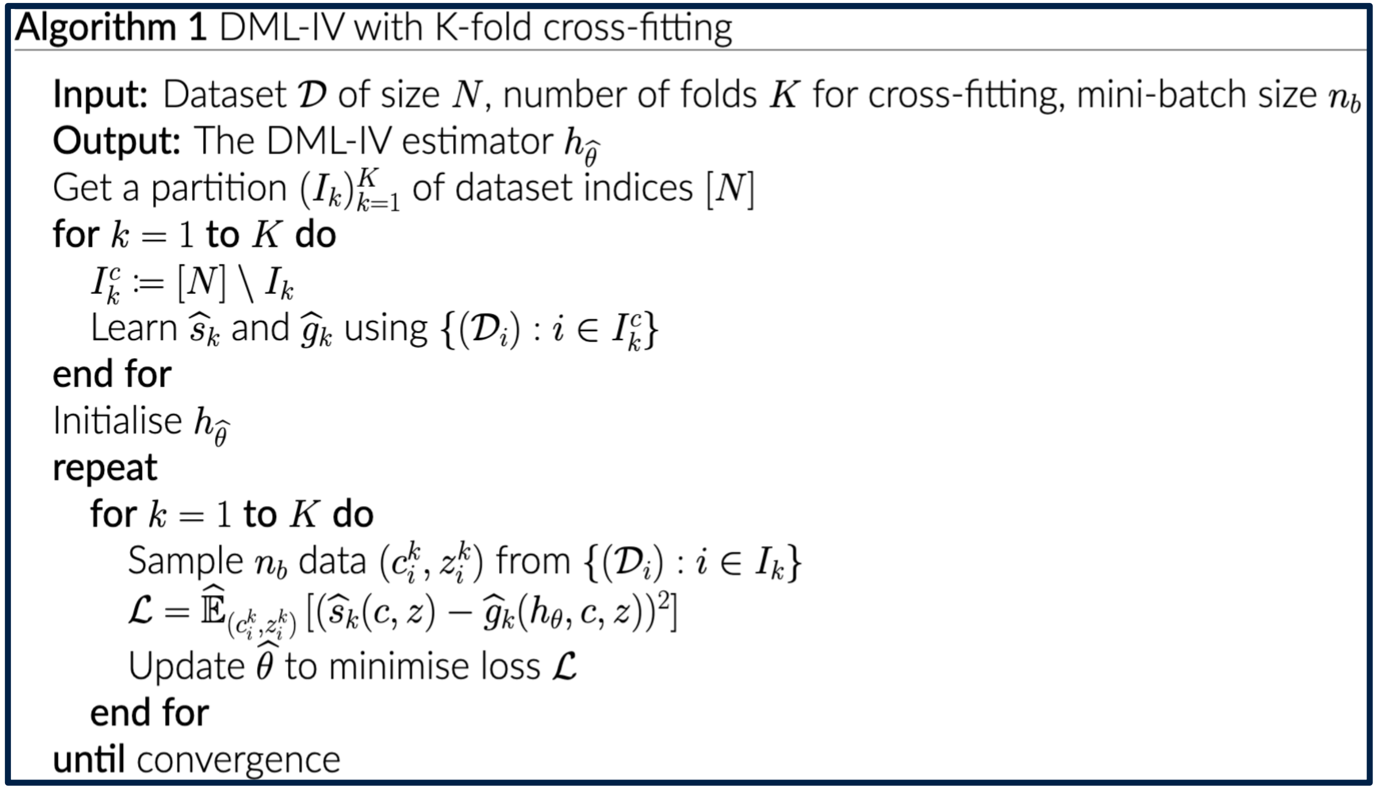
```
Input: Dataset $\mathcal{D}$ of size $N$, number of folds $K$ for cross-fitting, mini-batch size $n_b$
Output: The DML-IV estimator $h_{\widehat{\theta}}$
Get a partition $(I_k)_{k=1}^K$ of dataset indices $[N]$
for $k = 1$ to $K$ do
    $I_k^c := [N] \setminus I_k$
    Learn $\widehat{s}_k$ and $\widehat{g}_k$ using $\{(\mathcal{D}_i) : i \in I_k^c\}$
end for
Initialise $h_{\widehat{\theta}}$
repeat
    for $k = 1$ to $K$ do
        Sample $n_b$ data $(c_i^k, z_i^k)$ from $\{(\mathcal{D}_i) : i \in I_k\}$
        $\mathcal{L} = \widehat{\mathbb{E}}_{(c_i^k, z_i^k)}[(\widehat{s}_k(c,z) - \widehat{g}_k(h_\theta, c, z))^2]$
        Update $\widehat{\theta}$ to minimise loss $\mathcal{L}$
    end for
until convergence
```

DML-IV

We sample **mini-batch** (c_i^k, z_i^k) from $\mathcal{D}_{I_k}$ and optimize the following loss:

$$\hat{\mathbb{E}}_{(c_i^k, z_i^k)}\left[(\hat{s}_k(c, z) - \hat{g}_k(h_\theta, c, z))^2\right]$$

$$= \sum_{(c_i^k, z_i^k)} \frac{1}{n_b}\left((\hat{s}_k(c, z) - \hat{g}_k(h_\theta, c, z))^2\right)$$

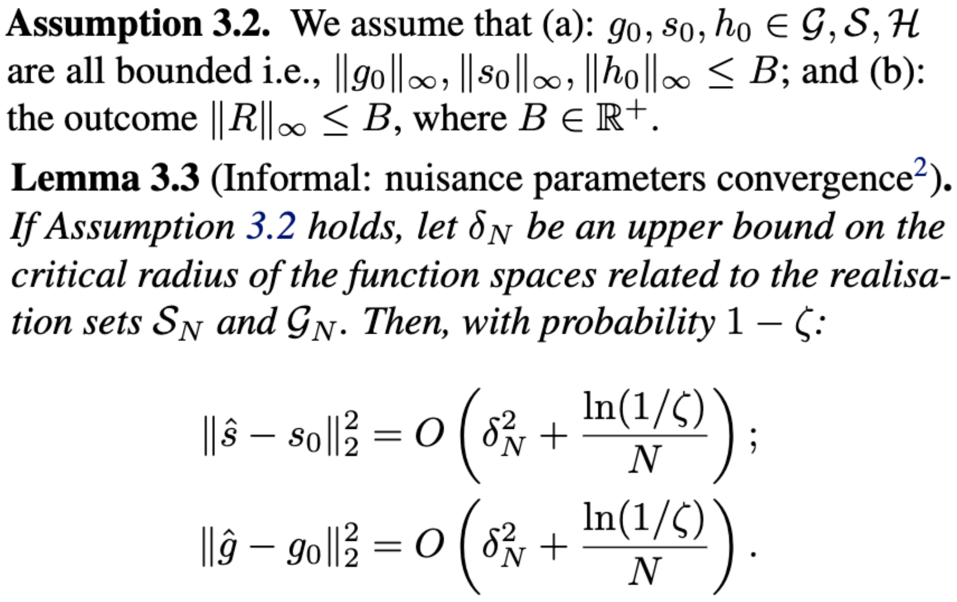


Algorithm 1 DML-IV with K-fold cross-fitting

```
Input: Dataset $\mathcal{D}$ of size $N$, number of folds $K$ for cross-fitting, mini-batch size $n_b$
Output: The DML-IV estimator $h_{\widehat{\theta}}$
Get a partition $(I_k)_{k=1}^K$ of dataset indices $[N]$
for $k = 1$ to $K$ do
   $I_k^c := [N] \setminus I_k$
   Learn $\widehat{s}_k$ and $\widehat{g}_k$ using $\{(\mathcal{D}_i) : i \in I_k^c\}$
end for
Initialise $h_{\widehat{\theta}}$
repeat
   for $k = 1$ to $K$ do
      Sample $n_b$ data $(c_i^k, z_i^k)$ from $\{(\mathcal{D}_i) : i \in I_k\}$
      $\mathcal{L} = \widehat{\mathbb{E}}_{(c_i^k, z_i^k)}[(\widehat{s}_k(c, z) - \widehat{g}_k(h_\theta, c, z))^2]$
      Update $\widehat{\theta}$ to minimise loss $\mathcal{L}$
   end for
until convergence
```

Theoretical Guarantees

In order to achieve the $O(N^{-1/2})$ convergence rate guarantee, apart from having a Neyman orthogonal score and using k-fold cross fitting, we additionally need that the nuisances parameter converges at $o(N^{-1/4})$.

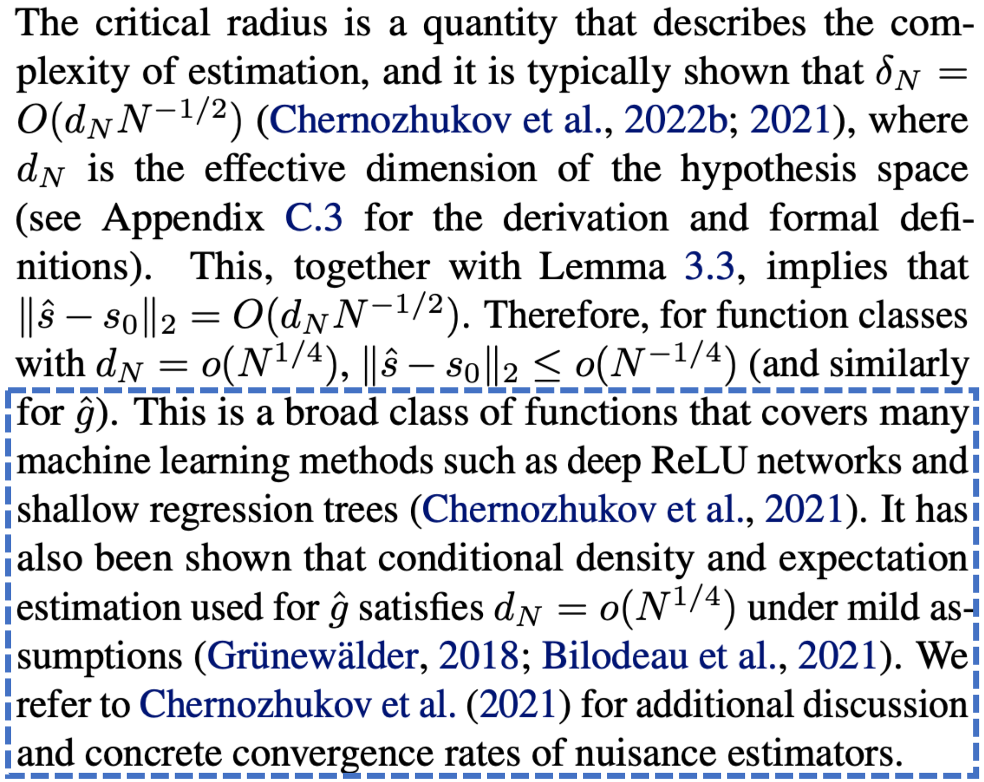


Assumption 3.2. We assume that (a): $g_0, s_0, h_0 \in \mathcal{G}, \mathcal{S}, \mathcal{H}$ are all bounded i.e., $\|g_0\|_\infty, \|s_0\|_\infty, \|h_0\|_\infty \leq B$; and (b): the outcome $\|R\|_\infty \leq B$, where $B \in \mathbb{R}^+$.

Lemma 3.3 (Informal: nuisance parameters convergence[2]). *If Assumption 3.2 holds, let δ_N be an upper bound on the critical radius of the function spaces related to the realisation sets $\mathcal{S}_N$ and $\mathcal{G}_N$. Then, with probability $1-\zeta$:*

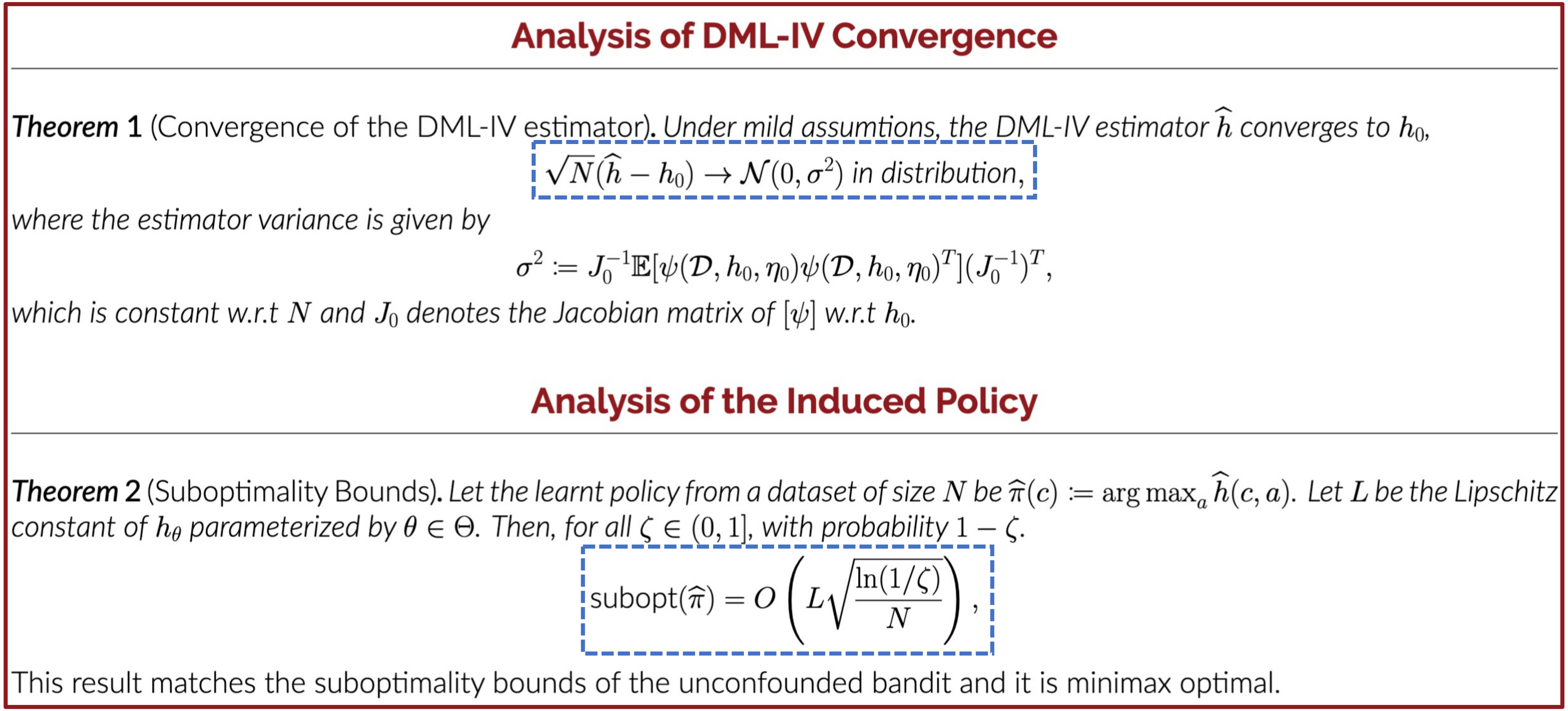
$$\|\hat{s} - s_0\|_2^2 = O\left(\delta_N^2 + \frac{\ln(1/\zeta)}{N}\right);$$

$$\|\hat{g} - g_0\|_2^2 = O\left(\delta_N^2 + \frac{\ln(1/\zeta)}{N}\right).$$



The critical radius is a quantity that describes the complexity of estimation, and it is typically shown that $\delta_N = O(d_N N^{-1/2})$ (Chernozhukov et al., 2022b; 2021), where d_N is the effective dimension of the hypothesis space (see Appendix C.3 for the derivation and formal definitions). This, together with Lemma 3.3, implies that $\|\hat{s} - s_0\|_2 = O(d_N N^{-1/2})$. Therefore, for function classes with $d_N = o(N^{1/4})$, $\|\hat{s} - s_0\|_2 \leq o(N^{-1/4})$ (and similarly for $\hat{g}$). This is a broad class of functions that covers many machine learning methods such as deep ReLU networks and shallow regression trees (Chernozhukov et al., 2021). It has also been shown that conditional density and expectation estimation used for $\hat{g}$ satisfies $d_N = o(N^{1/4})$ under mild assumptions (Grünewälder, 2018; Bilodeau et al., 2021). We refer to Chernozhukov et al. (2021) for additional discussion and concrete convergence rates of nuisance estimators.

Theoretical Guarantees



Analysis of DML-IV Convergence

Theorem 1 (Convergence of the DML-IV estimator). *Under mild assumtions, the DML-IV estimator $\widehat{h}$ converges to h_0,*

$$\boxed{\sqrt{N}(\widehat{h} - h_0) \to \mathcal{N}(0, \sigma^2) \textit{ in distribution},}$$

where the estimator variance is given by

$$\sigma^2 := J_0^{-1}\mathbb{E}[\psi(\mathcal{D}, h_0, \eta_0)\psi(\mathcal{D}, h_0, \eta_0)^T](J_0^{-1})^T,$$

which is constant w.r.t N and J_0 denotes the Jacobian matrix of $[\psi]$ w.r.t h_0.

Analysis of the Induced Policy

Theorem 2 (Suboptimality Bounds). *Let the learnt policy from a dataset of size N be $\widehat{\pi}(c) := \arg\max_a \widehat{h}(c, a)$. Let L be the Lipschitz constant of h_θ parameterized by $\theta \in \Theta$. Then, for all $\zeta \in (0, 1]$, with probability $1 - \zeta$.*

$$\boxed{\text{subopt}(\widehat{\pi}) = O\left(L\sqrt{\frac{\ln(1/\zeta)}{N}}\right),}$$

This result matches the suboptimality bounds of the unconfounded bandit and it is minimax optimal.

Experiments

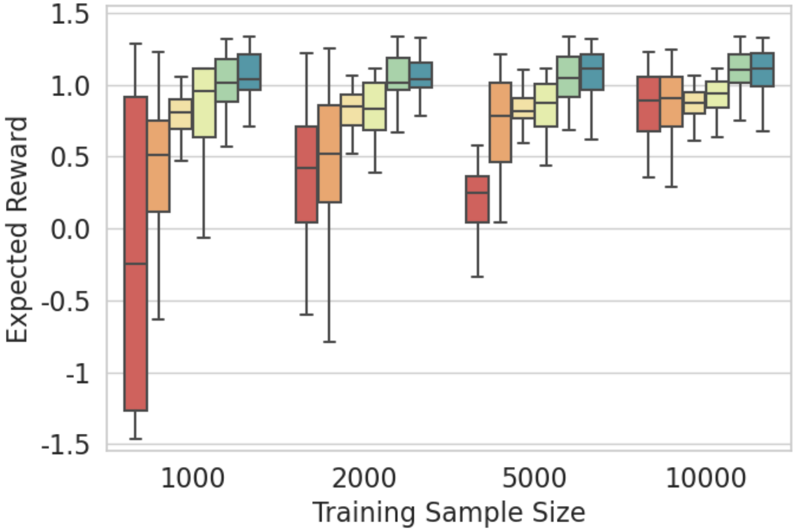
- We consider datasets with both low- and high-dimensional contexts, as well as semi-synthetic real-world datasets.

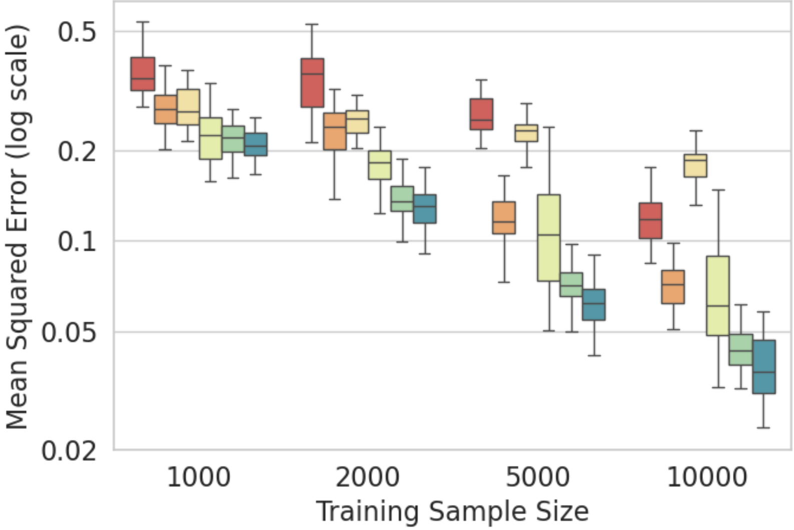
- We evaluate DML-IV, and a computationally efficient version of DML-IV, referred to as CE-DML-IV, which does not apply K-fold cross-fitting. In CE-DML-IV, nuisances parameter estimators are trained only once (instead of K times) using the entire dataset. It can also be considered as an ablation study for K-fold cross-fitting.

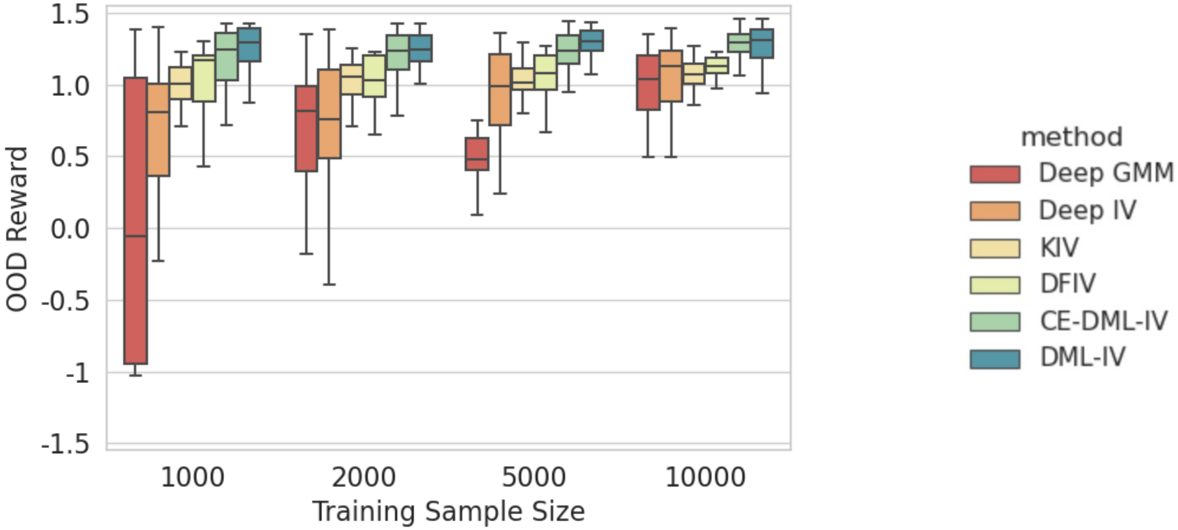
- Note that CE-DML-IV lacks the theoretical convergence rate guarantees but it still enjoys the partial debiasing effect from the Neyman orthogonal score and trades off computational complexity with bias. We found that CE-DML-IV empirically performs as well as standard DML-IV on low-dimensional datasets.

Experiments

Synthetic airplane ticket sales dataset, where the causal function is a complex non-linear function.

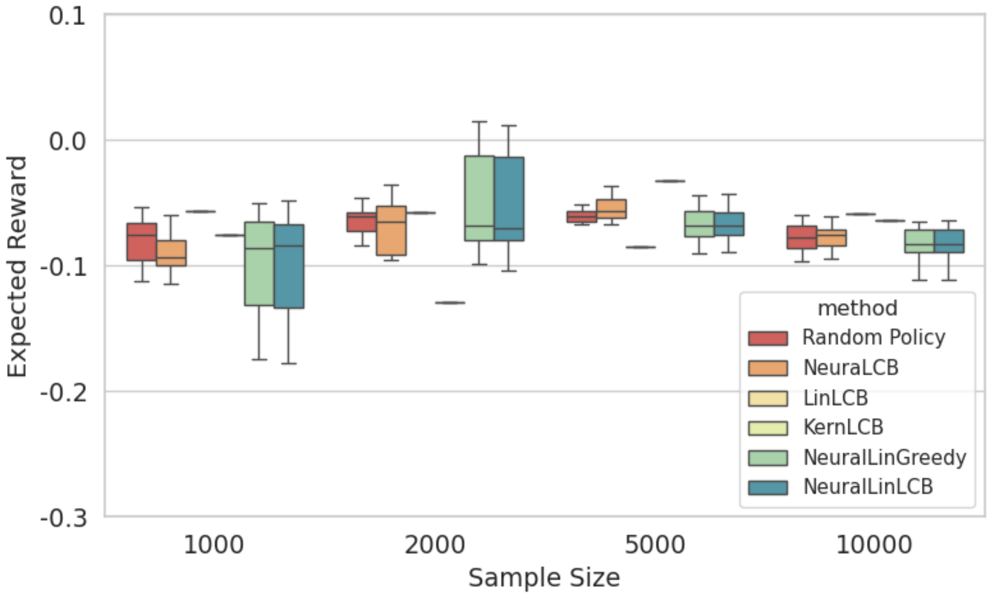






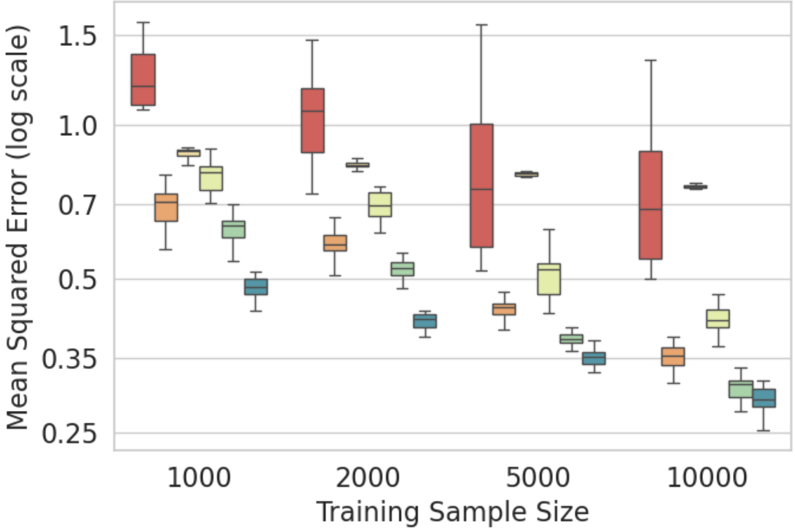
Experiments

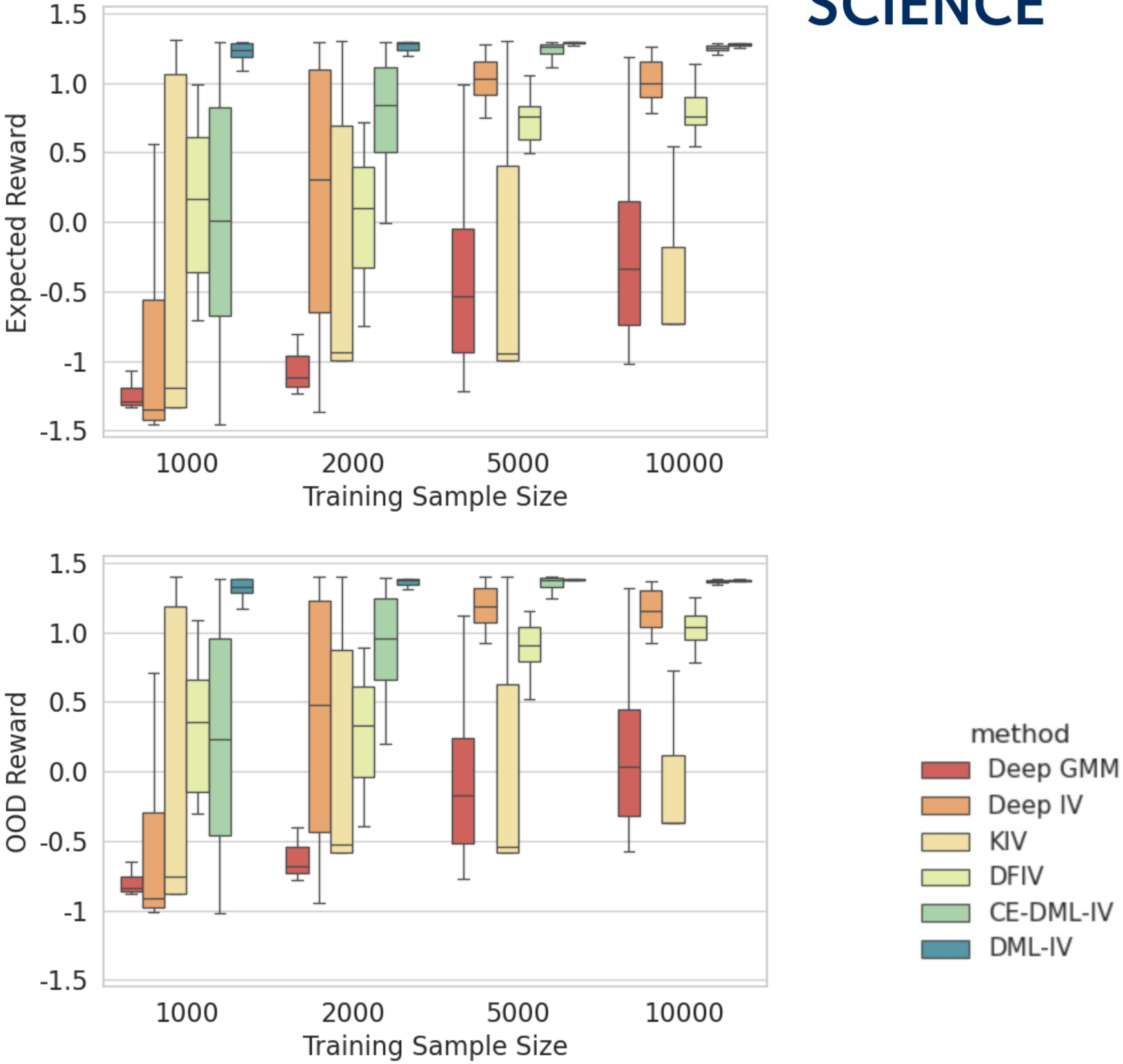
Standard offline bandit algorithms that don’t explicitly consider IVs failed to learn meaningful policies when the data is confounded.
(On airplane ticket sales dataset)



Experiments

Synthetic airplane ticket sales dataset, where we replace the customer type variable ∈ [1..7] (a context variable) with MNIST images of the same digits.

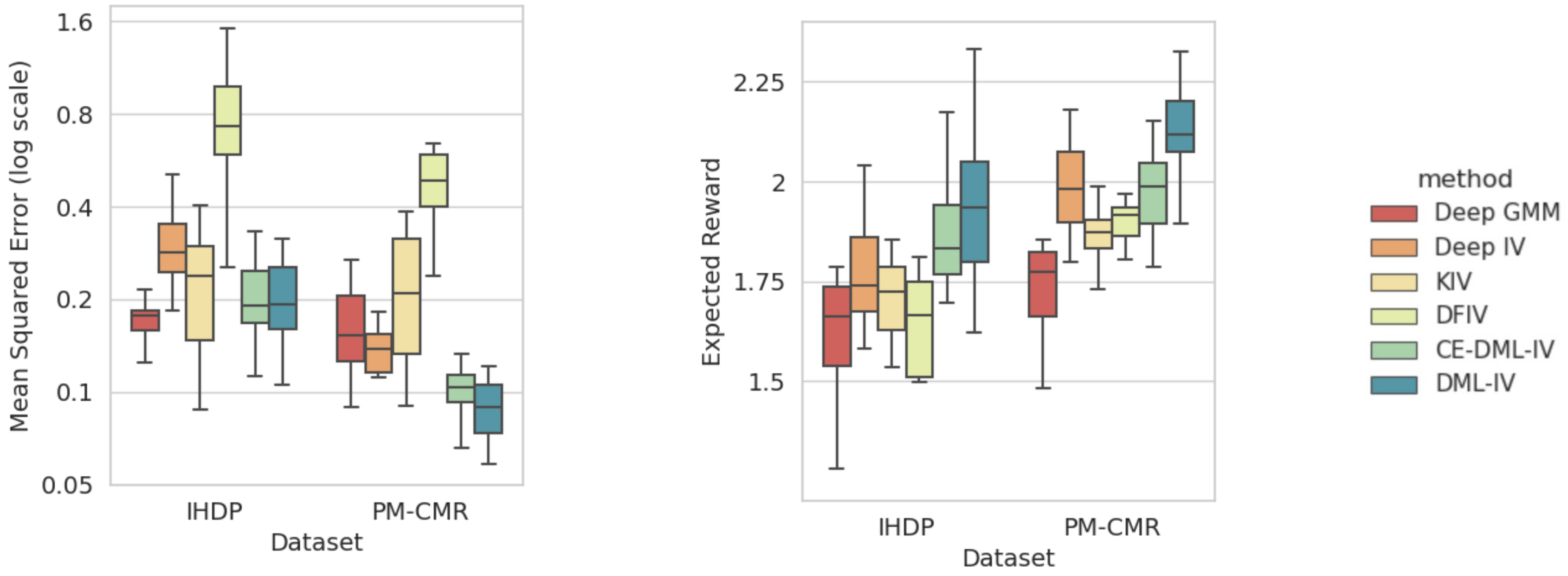




Experiments

Real-world datasets:

The true counterfactual prediction function is rarely available for real-world data. Therefore, in line with previous approaches, we instead consider two semi-synthetic real-world datasets: Infant Health and Development Program (IHDP) dataset and the PM-CMR (impact of PM2.5 particle level on the cardiovascular mortality rate), where only the outcome is generated synthetically.



Conclusion

- We have proposed a novel method for instrumental variable regression, DML-IV. By leveraging IVs and DML on offline data
- DML-IV can learn counterfactual predictions and effective decision policies with fast convergence rate and suboptimality guarantees by mitigating the regularisation and overfitting biases of DL
- Superior performance against SOTA IV-regression methods
- DML-IV is practical and can be used to solve real-world problems if **IVs are available!**