ml-mr: Software for nonparametric and nonlinear Mendelian randomization estimation using machine learning

Marc-André Legault*, Jason Hartford, Michael Lu, Archer Y. Yang, Joëlle Pineau

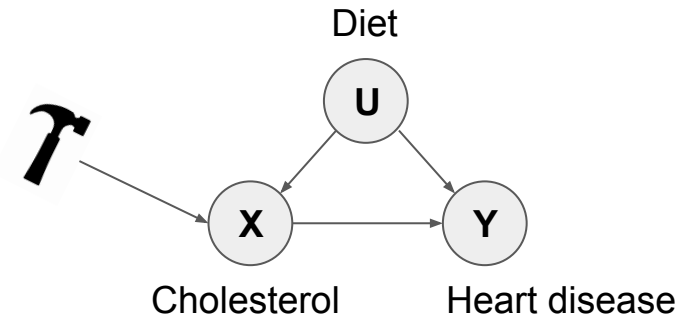
Introduction

Estimating causal effects is difficult because of confounders

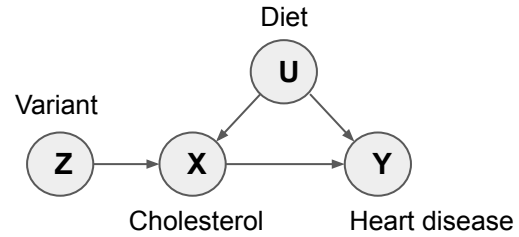
- Variables causing both the exposure and the outcome
- May be unmeasured or incompletely measured
- The observed X → Y effect will include the contribution that goes “through” the confounding variables (U)

Instrumental variables (IVs) influence the exposure independently from the confounders

- Genetic variants (Mendelian randomization)
- Random allocation to pharmacological intervention in clinical trials
- …



Introduction



Example of machine learning MR estimator: **DeepIV**

The **two-stage least squares** estimator (conventional IV estimator)

Stage 1

Predict X based on Z using a linear regression

$$\mathbb{E}[X|Z] = Z\beta$$

$$\hat{x} = Z\hat{\beta}$$

Stage 2

Regress the predicted values of the exposure on the outcome

$$\mathbb{E}[Y|\hat{X}] = \hat{X}\theta$$

DeepIV estimator (Hartford *et al.* ICLR 2017)

Stage 1

Estimate the conditional density of X given Z

$$\hat{F}_\phi(x|z)$$

Arbitrary density model parametrized by ϕ

Stage 2

Sample from x | z and use the samples to estimate the X → Y function

$$y = f_\theta(\dot{x}),\ \ \dot{x} \sim \hat{F}_\phi(x|z)$$

This step is akin to traditional supervised learning

Objectives / Research Question

1. Provide high quality implementations of machine learning instrumental variable estimators to be used by bioinformaticians / biostatisticians
    - Streamline machine learning model selection and optimization
    - Quantify uncertainty
    - Facilitate model comparison and evaluation
2. Do machine learning instrumental variable estimators work for MR?
    - Effect of heritability, sample size, etc.
    - What metrics should we use for model selection?
    - What new do they bring?

Results - Overview of *ml-mr*

Simulation

- Pre-implemented simulation models from MR studies (*e.g.* Tian *et al.* 2023, Burgess *et al.* 2014)
- Facilitate algorithm comparison and large scale simulation studies

Estimation

- **Unified programmatic and command-line interfaces to fit machine learning IV estimators**
- Support logging and best practices: sample splitting, early stopping, model checkpointing
- **Uncertainty quantification using conformal inference**

ML-MR 

Evaluation

- Evaluate simulation models with respect to true function
- Compare models, compute performance metrics, extract metadata

Sweep

- **Grid or probabilistic sampling of hyperparameter configurations**
- Multiprocessing support
- Find optimal parameters

github.com/legaultmarc/ml-mr

Results - Overview of *ml-mr*

github.com/legaultmarc/ml-mr

Simulation

```
sim = Simulation(n, prefix)

sim.parameters["h2"] = h2
sim.add_variable(Normal("u", 0, 1))
…
sim.resample()
sim.save(fn)  # CSV + metadata
sim.save_pickle(fn)  # Python pickle file
```

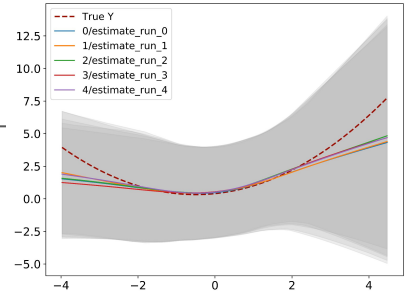
Estimation

```
ml-mr estimation \
  quantile_iv \
  --q 10 \
  --data filename.tsv.gz \
  --exposure x --outcome y --instruments v{1..20} \
  --wandb-project my_fit
```

ML-MR

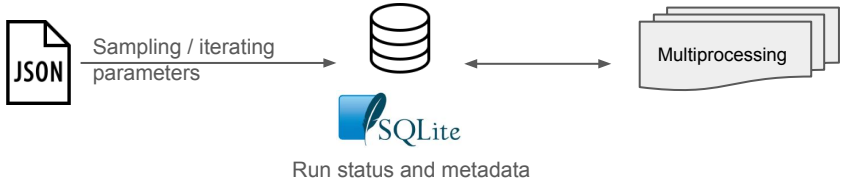
Evaluation

```
ml-mr  \
    evaluation \
    --input dfiv_estimate \
    --true-function 'lambda x: (0.3 * (x + 2) * (x - 1) + 1)' \
    --plot
```



Sweep

```
ml-mr sweep config.json
```

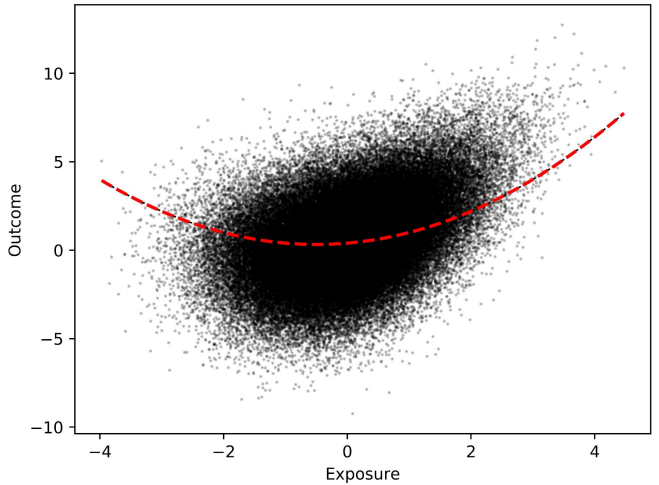


Results

Simulation study for a “simple” case

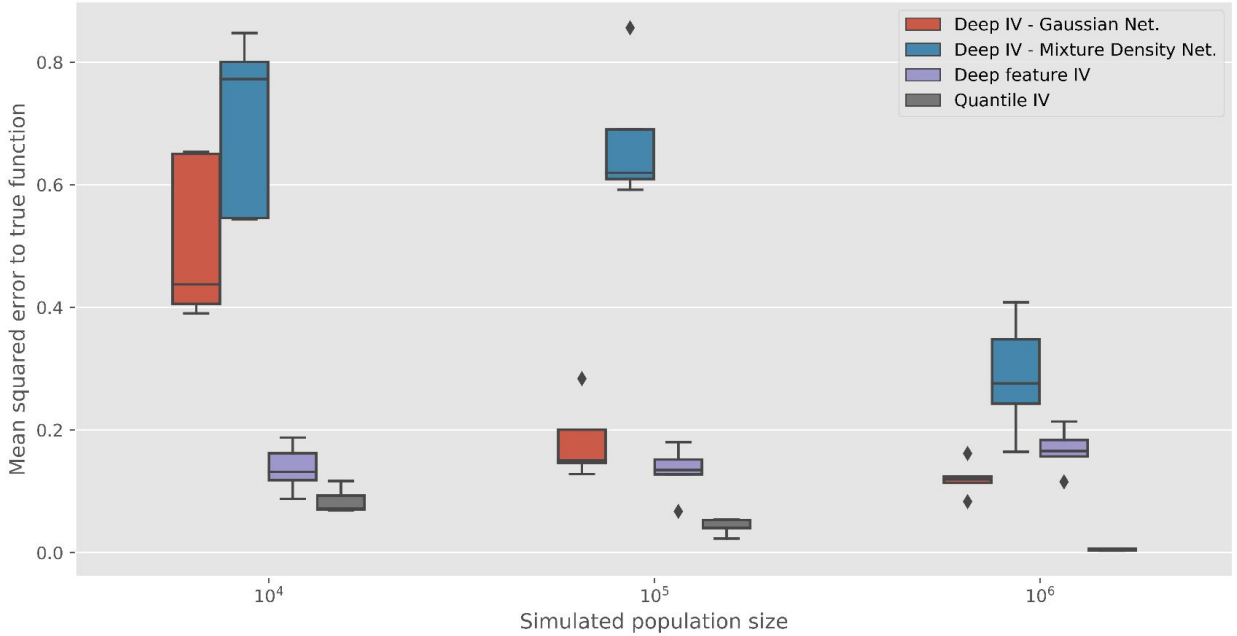
- Quadratic true function
- 20 independent variants (IVs)
  - Sampled allele frequencies between 20% and 45%
- 3 simulation parameters:
  - Heritability in the exposure explained by the IVs (h^2)
  - Population size (n)
  - Confounding strength (effect between latent confounder and exposure, β_{UX})
- 5 Simulation replicates for each configuration

Parameter	**Simulated values**
Variance in X explained (h^2)	0.2, **0.5**, 0.8
Population size (n)	10^4, **10^5**, 10^6
Confounding strength (β_{UX})	-0.5, -0.25, 0.25, **0.35**, 0.5



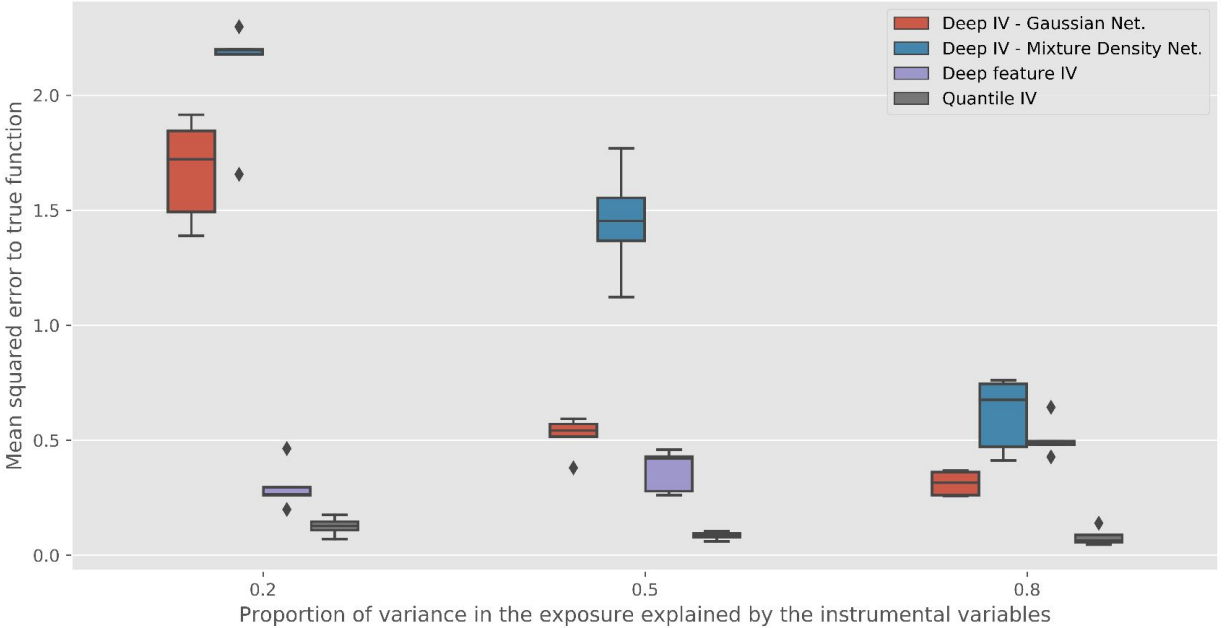
Model comparison

- We show the **best model**'s MSE across hyperparameter sweep
- The distribution represented by the **boxplots are over simulation replicates**
- **Deep feature IV and Quantile IV are less sensitive to sample size**
- Good performance even with **10,000** samples



Model comparison

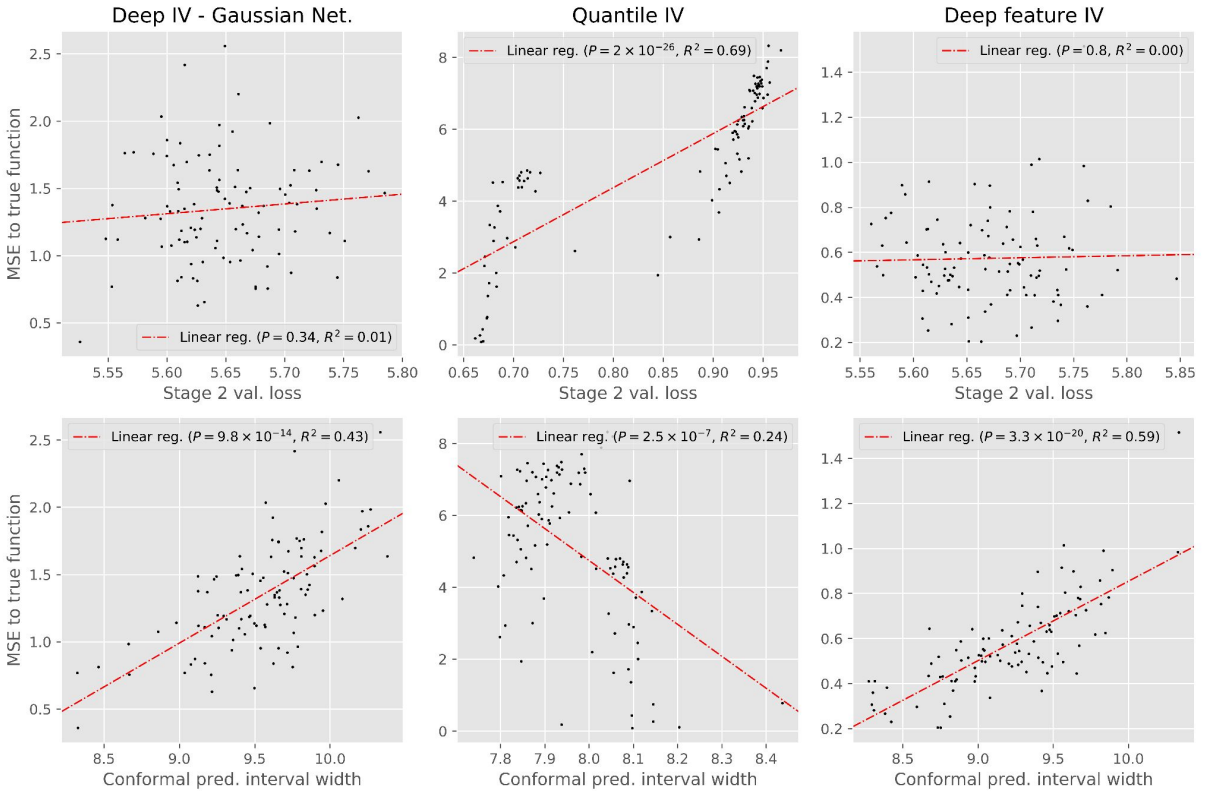
- Deep feature IV and Quantile IV also perform well when 20% of the variance in the exposure is explained by the IVs



Results - Model selection

How can we select the best model when the true function is unknown?

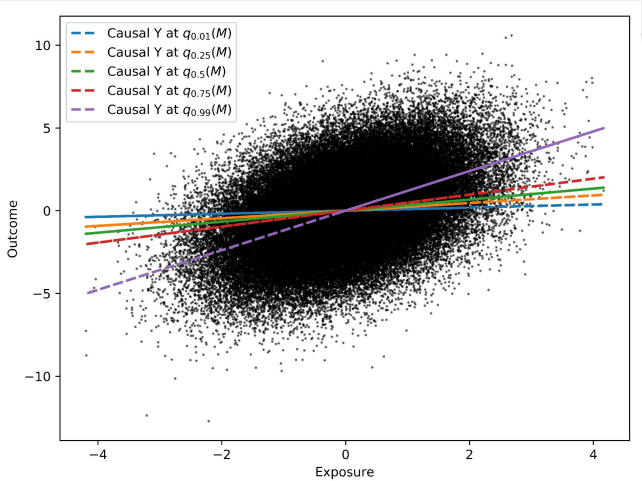
- For Deep IV and Deep Feature IV, it's a **bad idea to select model based on 2nd stage loss** (plot 1st row)
- For these models, we should rely on the mean **conformal prediction interval width** (plot 2nd row)
- Quantile IV uses a different uncertainty measure
  (Simultaneous Quantile Regression, Tagasovska (2018))



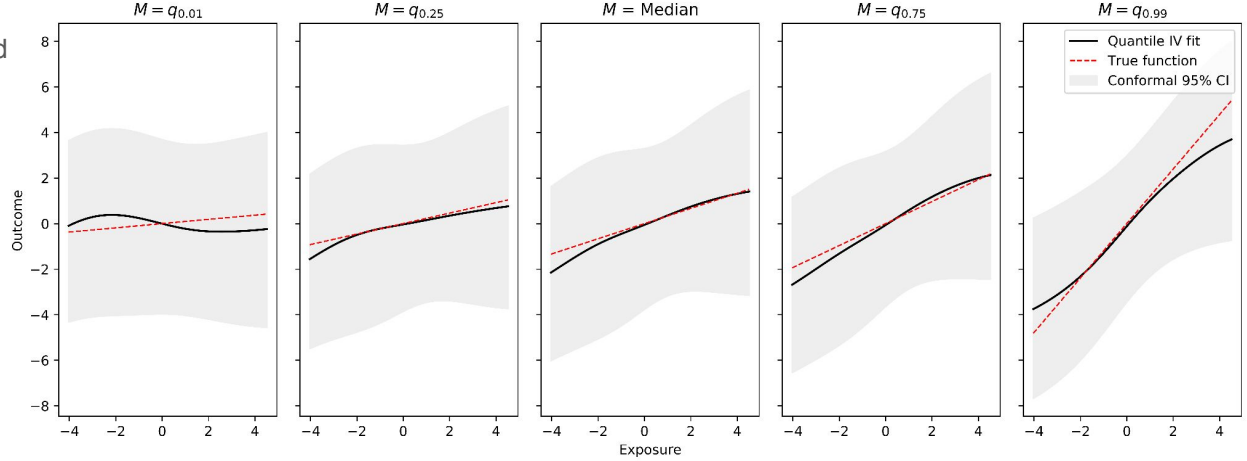
Results - Conditional Average Treatment Effects

Machine learning MR methods also **estimate conditional effects**

- **We introduce an effect modifier (M)**
- The effect is linear within strata of M, but the causal effect changes **nonlinearly** based on the effect modifier
- Think of the **effect of a drug on a health outcome for individuals with different drug response**



Quantile IV fit presented at different levels of the modifier variable.



Conclusion

- ***ml-mr* is the first software package providing user-friendly implementations of machine learning instrumental variable estimators for MR**
- We characterized the performance of the IV estimators in realistic MR scenarios
- We showed the potential of machine learning IV estimators to estimate **conditional** treatment effects

Future work

- Better understand model selection in this challenging context
- *ml-mr* for drug target validation
- *ml-mr* is still under development and we're constantly improving it

Acknowledgements

McGill

- Joëlle Pineau (research director)
- Archer Y. Yang
- Michael Lu

Mila & Recursion

- Jason Hartford

ml-mr is available on github

github.com/legaultmarc/ml-mr

For questions, contact me

marc-andre.legault@mcgill.ca

@legaultmarc

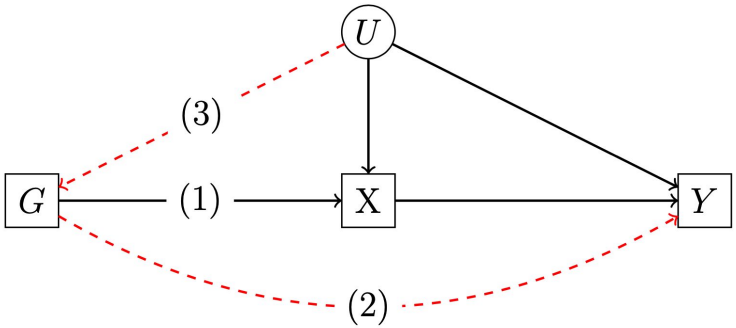
marclegault.com

SUPPLEMENTARY SLIDES

Introduction

Mendelian randomization (MR) studies use genetic variants as instrumental variables

They allow estimating causal effects even in the presence of unmeasured confounding, but rely on important assumptions



The conventional instrumental variable assumptions used in MR are:

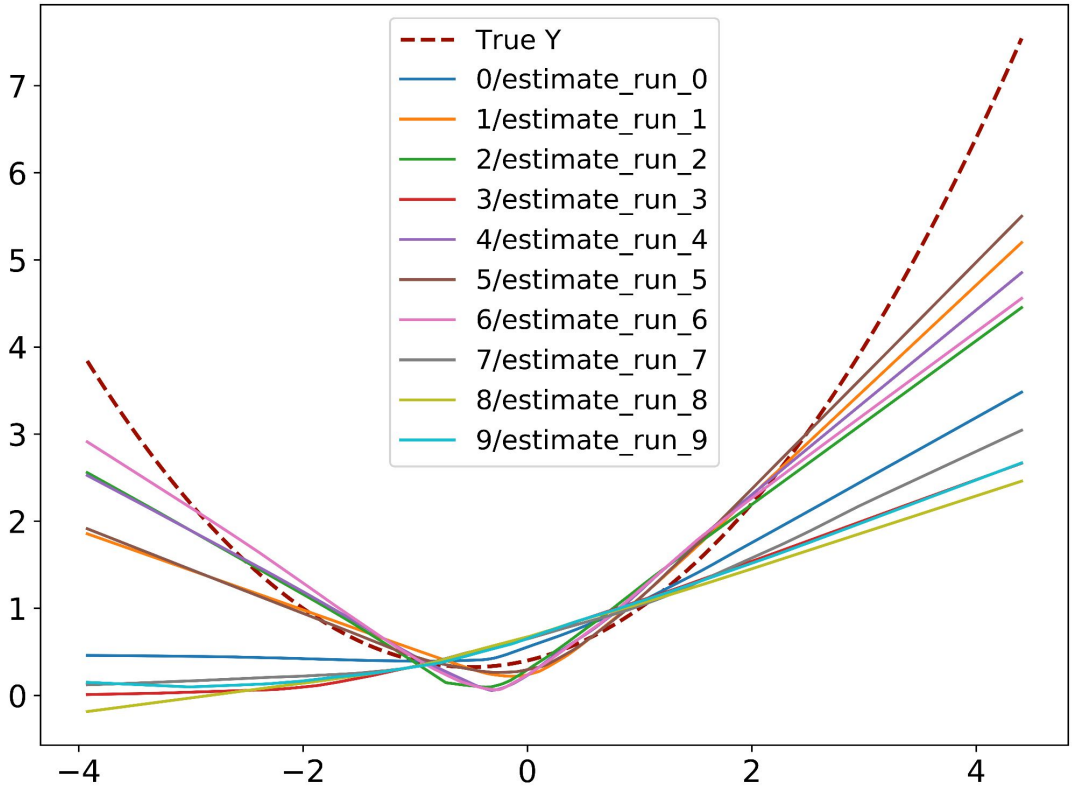
1. **Relevance**: The genetic variant has an effect on the exposure
2. **Exclusion**: The genetic variant has no direct effect on the outcome (i.e. effects that do not go through the exposure)
3. **Unconfoundedness**: There are no confounders of the instrument-outcome relationship

In addition: Homogeneity or monotonicity assumptions and parametric assumptions!

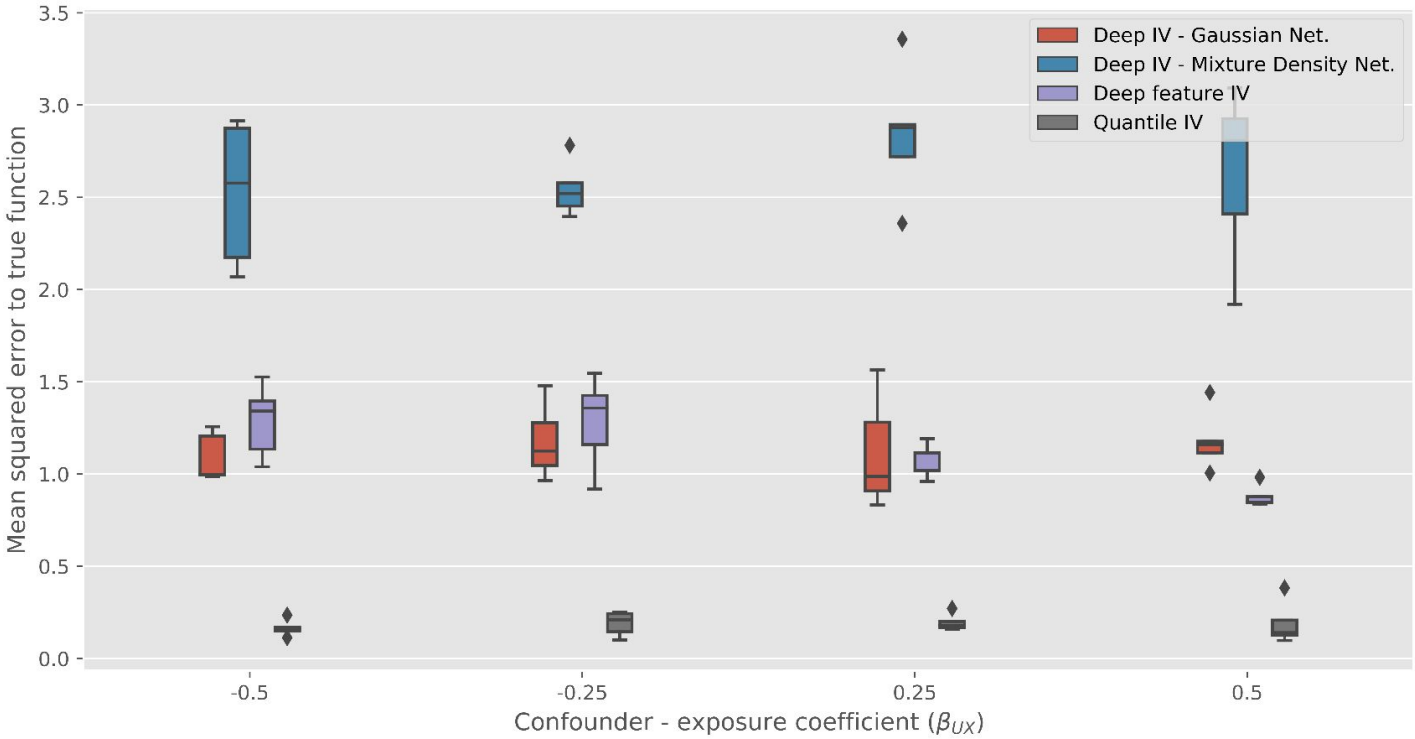
Variability in the estimated function across a hyperparameter sweep (10 runs) for DFIV

Simulation model:

- Quadratic (homogeneous)
- h2 = 0.2
- n = 10^5
- beta_ux = 0.35



Model comparison



quantile_iv sensitivity to q