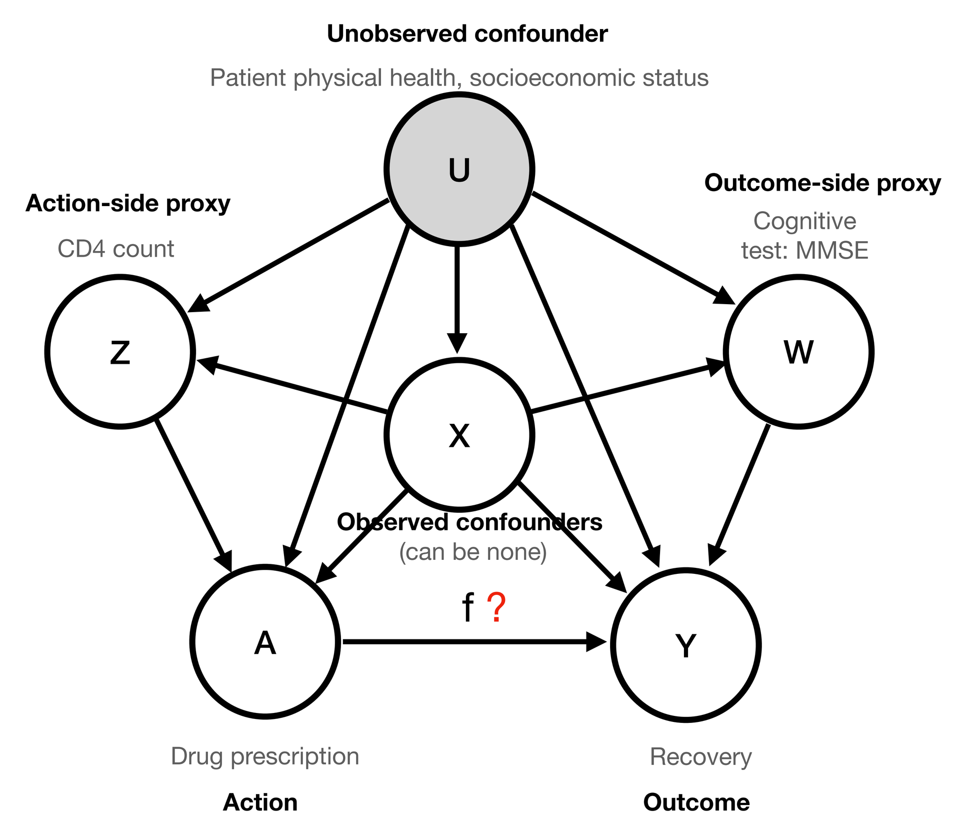
Relaxing Observability Assumption in Causal Inference with Kernel Methods

Yuchen Zhu [1] Limor Gultchin [2] Arthur Gretton [1] Anna Korba [3] Matt Kusner [1] Afsaneh Mastouri [1] Krikamol Muandet [4] Ricardo Silva [1]

[1]UCL [2]University of Oxford [3]ENSAE/ CREST [4]CISPA

Proximal Causal Learning



Data: $\{a, z, x, w, y\} \sim P(A, Z, X, W, Y)$.

The Proximal Problem

$$\mathbb{E}[Y|A,X,Z] = \int_{\mathcal{W}} h(A,X,W)\, dF(W|A,X,Z)$$

Learn h, solution to above Fredholm integral equation. It follows:

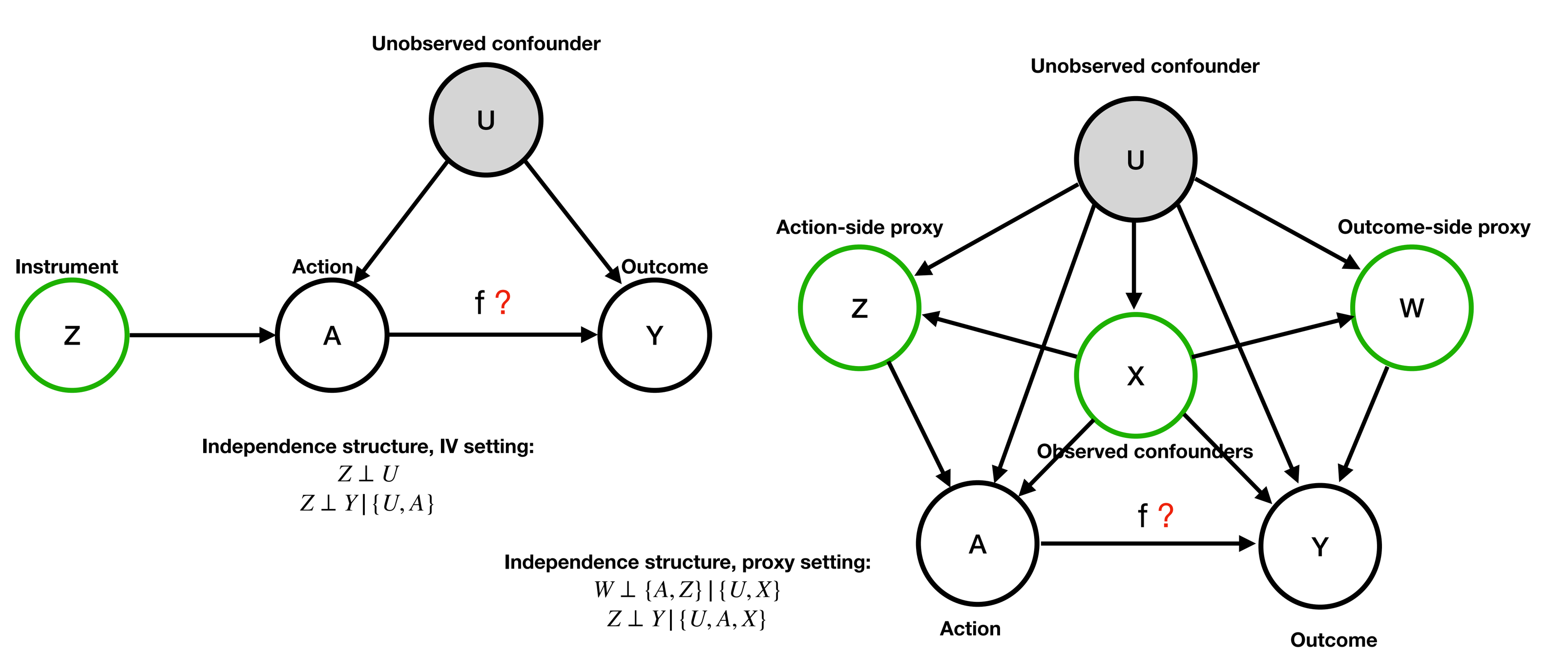
$$\mathbb{E}[Y|do(A)=a] = \mathbb{E}_{X,W}\, h(a,X,W)$$

Algorithm: Proximal Maximum Moment Restriction (PMMR)

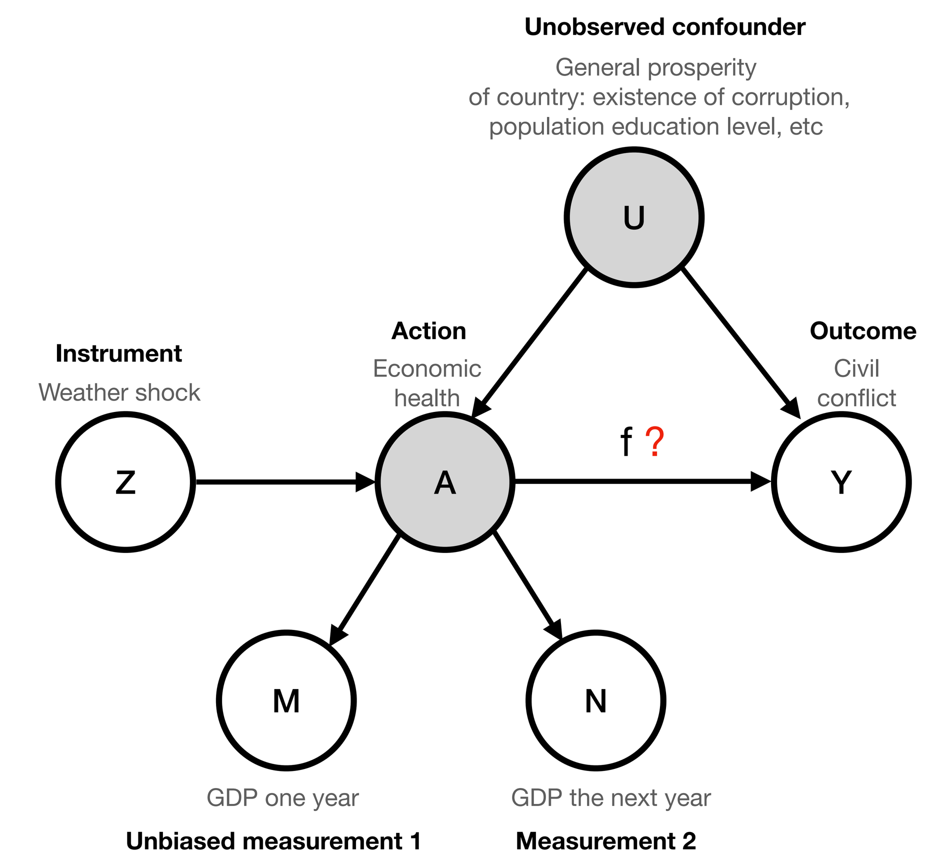
Transfer minimising the discrepancy between the two sides of (1) into working with a weighted regression objective:

$$\min_h \mathbb{E}[(Y - h(A,W,X))(Y' - h(A',W',X'))k((A,Z,X),(A',Z',X'))]$$

Comparison between IV and Proximal settings



Causal Inference with Action Measurement Error



Data: $\{z, m, n, y\} \sim P(Z, M, N, Y)$.

Identifying The Characteristic Function $\psi_{A|Z}$

$$\overbrace{\mathbb{E}_{\mathcal{P}_{A|z}}[e^{i\alpha X}](\alpha)}^{\psi_{A|Z}(\alpha)} = \exp\left(\int_0^\alpha i\frac{\mathbb{E}[Me^{i\nu N}|z]}{\mathbb{E}[e^{i\nu N}|z]}d\nu\right) \qquad (\star)$$

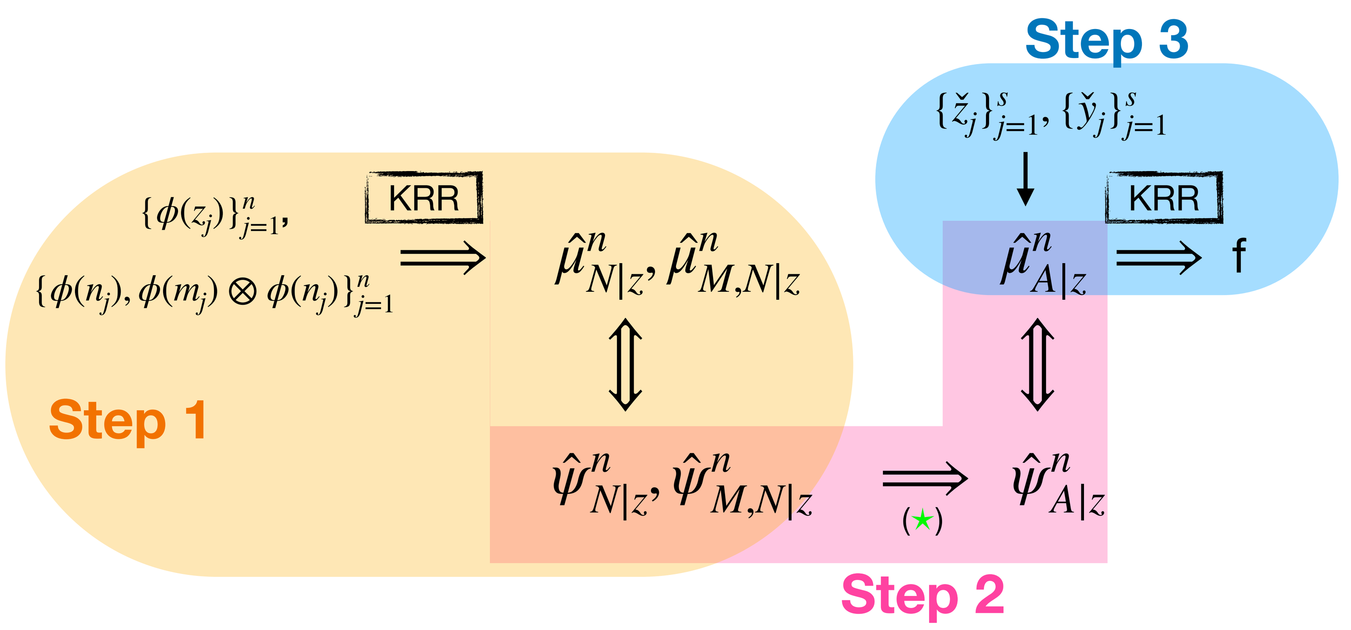
From Characteristic Function to Mean Embeddings

$$\hat{\mu}^n_{A|z}(y) = \textstyle\sum_{j=1}^n \hat{\gamma}^n_j(z)k(a_j, y),\ \hat{\psi}^n_{A|z}(\alpha) := \sum_{j=1}^n \hat{\gamma}^n_j(z)e^{i\alpha a_j}.$$

$$\hat{\gamma}^n_j(z) = (K_{ZZ} + n\hat{\lambda}^n I)^{-1}K_{Zz}$$

Theorem. $\hat{\mu}^n_{A|Z} \to^n \mu_{A|Z}$ iff $\hat{\psi}^n_{A|Z} \to^n \psi_{A|Z}$ in IFT of kernel.

Algorithm: Merrorment-Error Kernel Instrumental Variable Regression (MEKIV)



Contributions

Proximal Maximum Moment Restriction

1. A kernel-based nonparametric estimation algorithm given by duality of the loss objective; can be applied to a more general class of inverse problems that involve a solution to a Fredholm integral equation.
2. Derive convergence guarantees for the proposed algorithm.

Measurement-Error Kernel Instrumental Regression

1. A provably consistent, kernel-based nonparametric estimation algorithm for estimating the structural function under measurement error in the action variable.
2. Connect characteristic function estimation with mean embedding learning.

Results on (semi-)synthetic datasets. Top: PMMR; Bottom: MEKIV

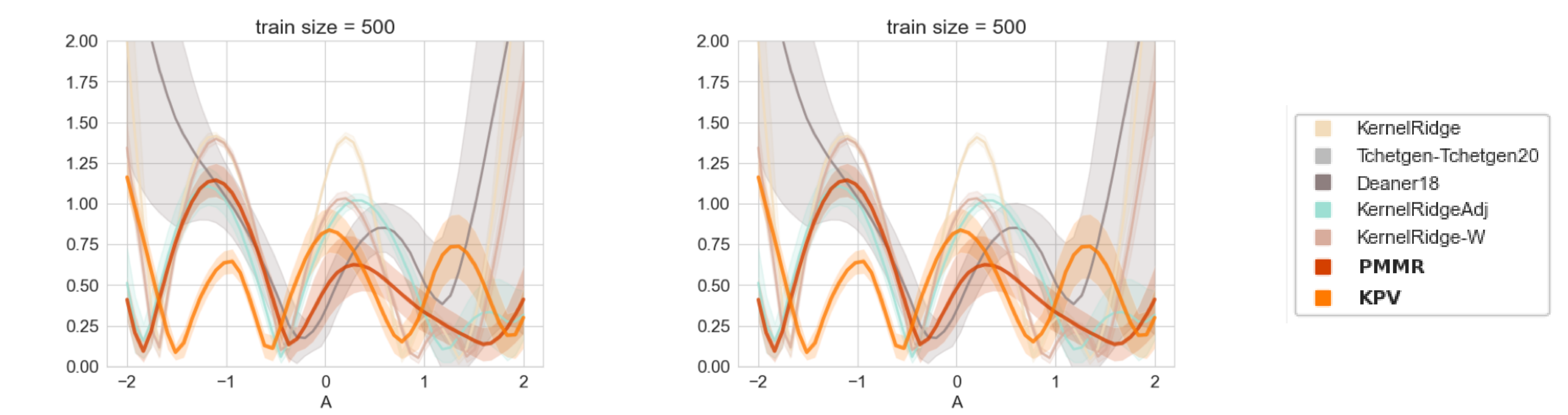


Figure 1:PMMR results on a synthetic dataset: to evaluate the efficacy of our method on a dataset designed to be absent of an observed valid adjustment set.

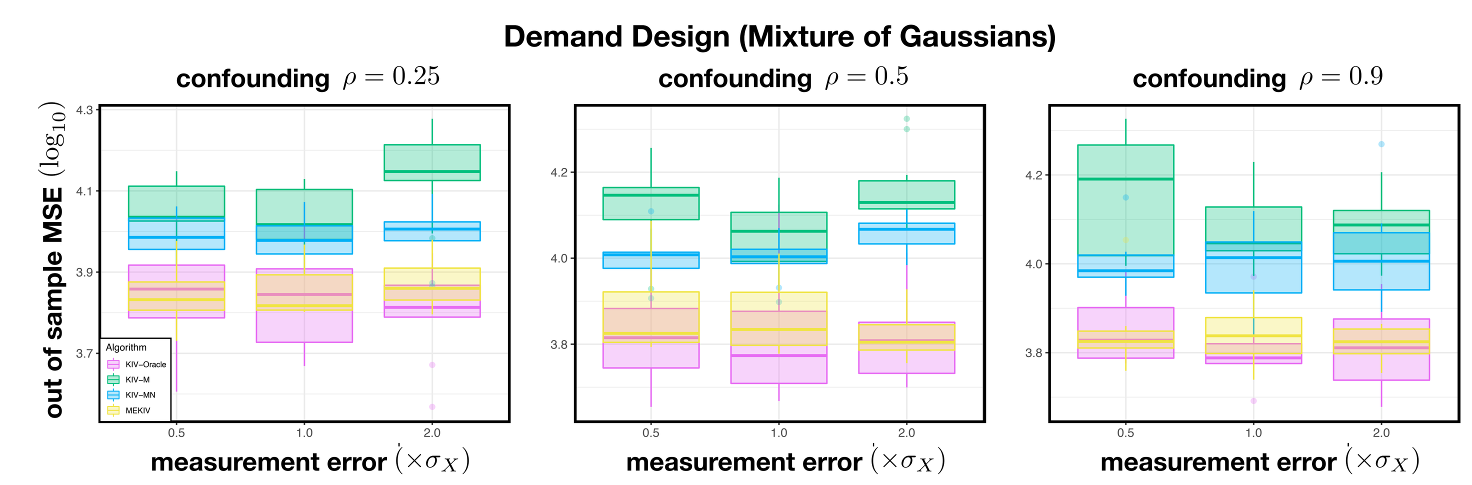


Figure 2:MEKIV results on a semi-synthetic dataset with true actions masked from model.

References

[1] Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt J. Kusner, Arthur Gretton, and Krikamol Muandet. Proximal causal learning with kernels: Two-stage estimation and moment restriction. In *ICML*, 2021.

[2] Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 2018.

[3] Susanne M. Schennach. Estimation of nonlinear models with measurement error. *Econometrica*, 2004.

[4] Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*, 2020.

[5] Yuchen Zhu, Limor Gultchin, Arthur Gretton, Matt J. Kusner, and Ricardo Silva. Causal inference with treatment measurement error: a nonparametric instrumental variable approach. In *UAI*, 2022.