Causal Inference with Treatment Measurement Error

A Non-parametric Instrumental Variable Approach

Z., Limor Gultchin, Arthur Gretton, Matt Kusner, Ricardo Silva

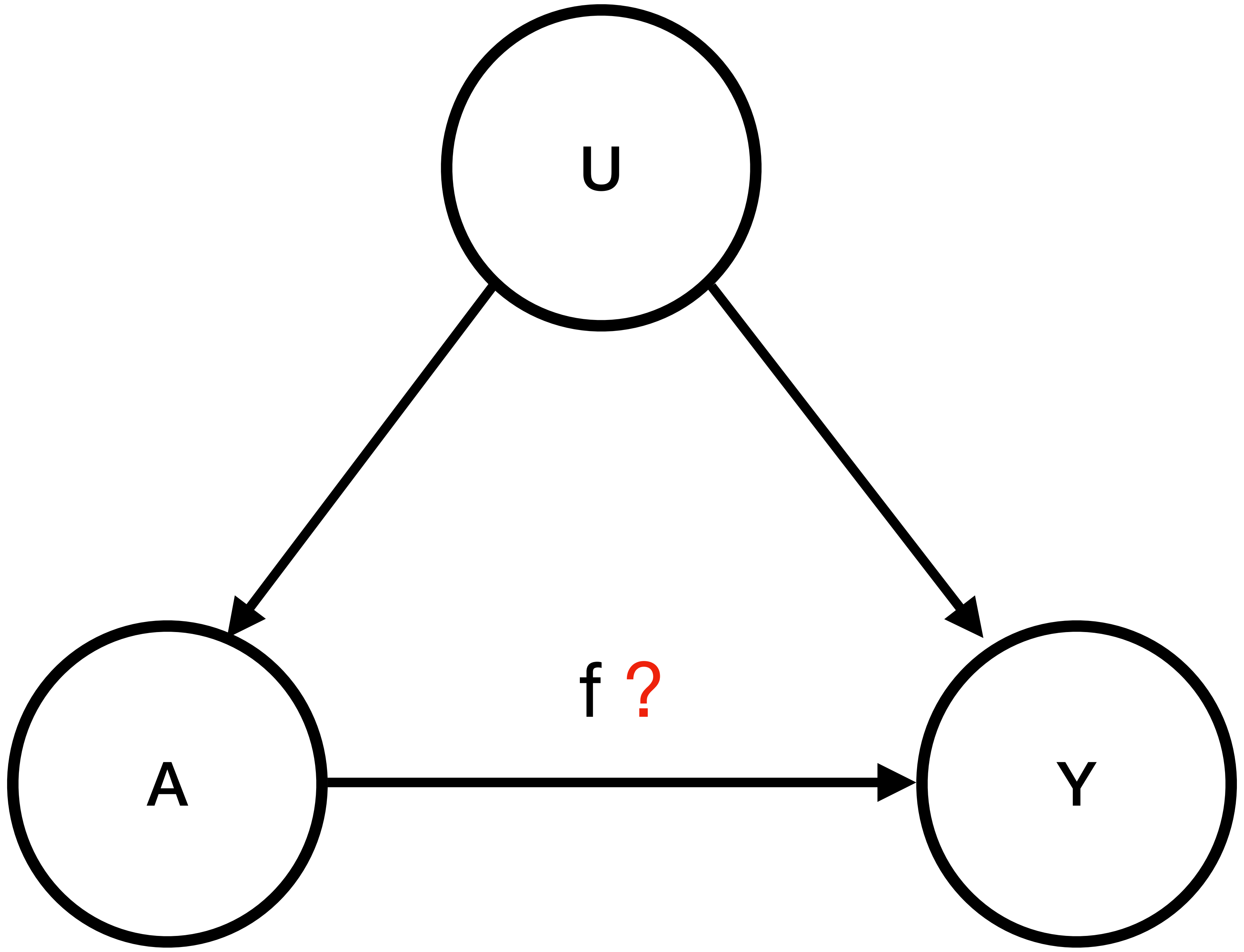
Oxford talk 10.2022

Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence | Oral presentation

Causal inference in the real world

Image source: Google image

Cognitive ability

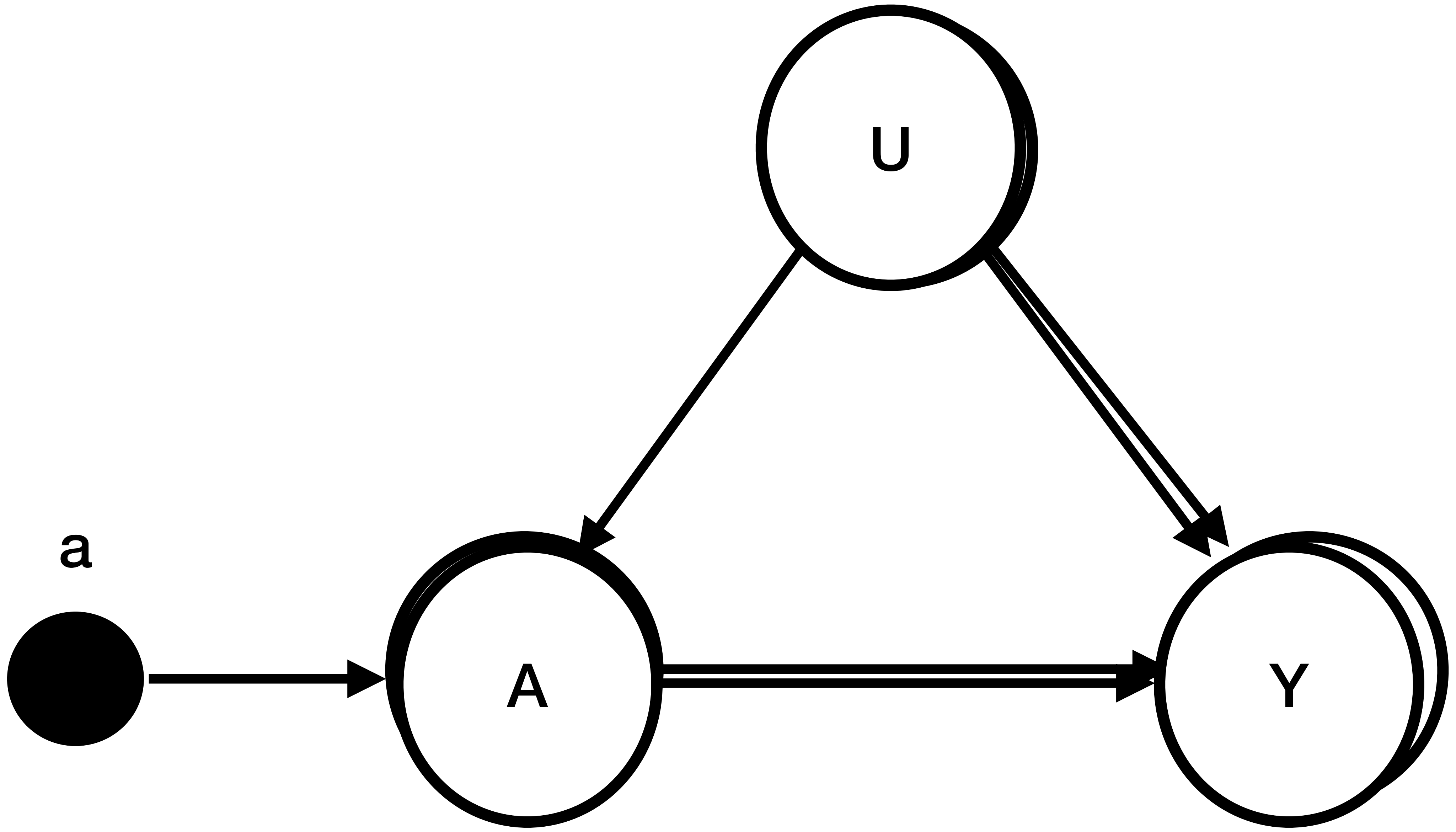


Grade retention

Cognitive outcome

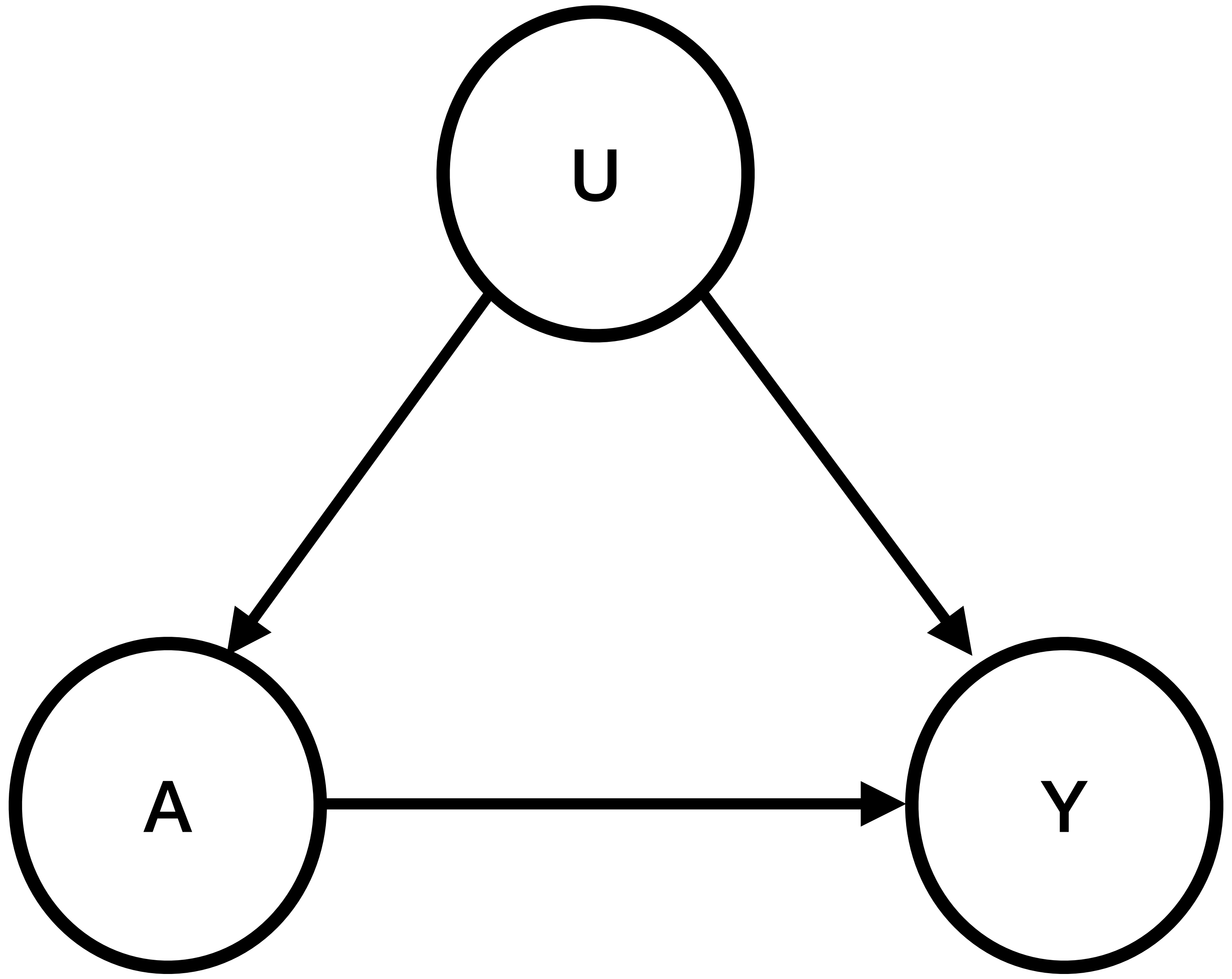
Fruehwirth, Navarro, Takahashi. 2016: How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects.

Target quantity from observational data



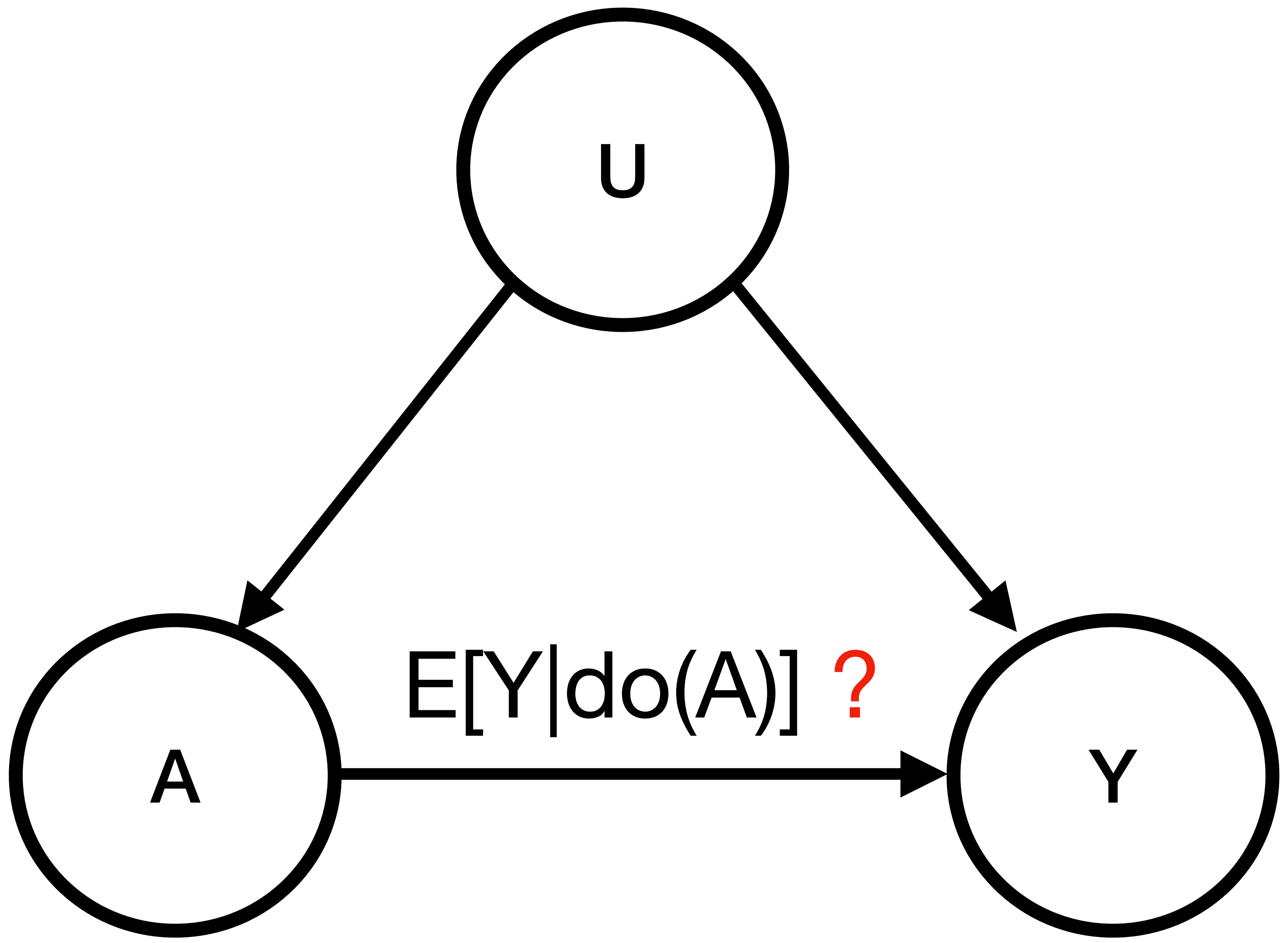
$E[Y|do(A := a)]$

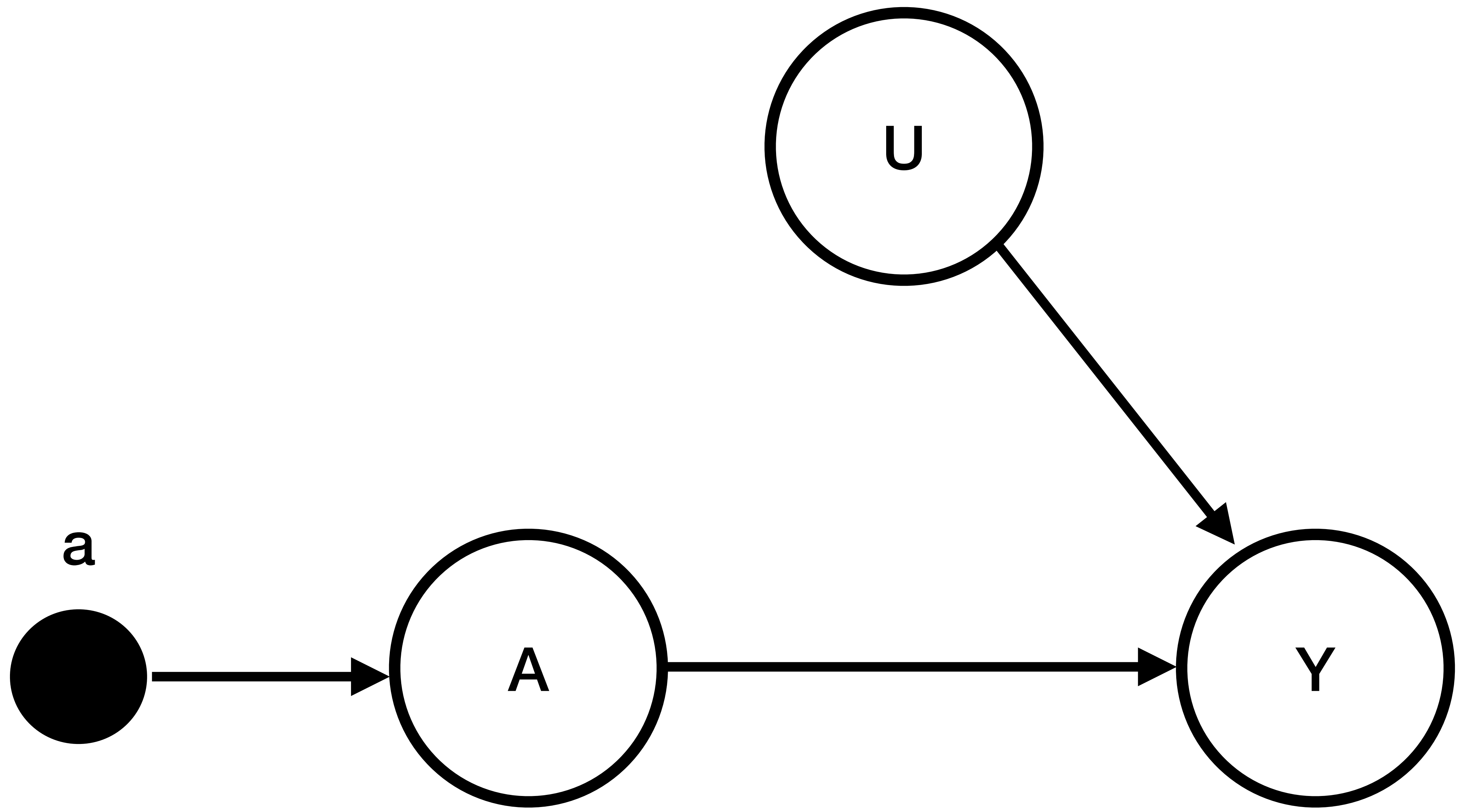
Observed data



Observed data

Recap: Observed confounders



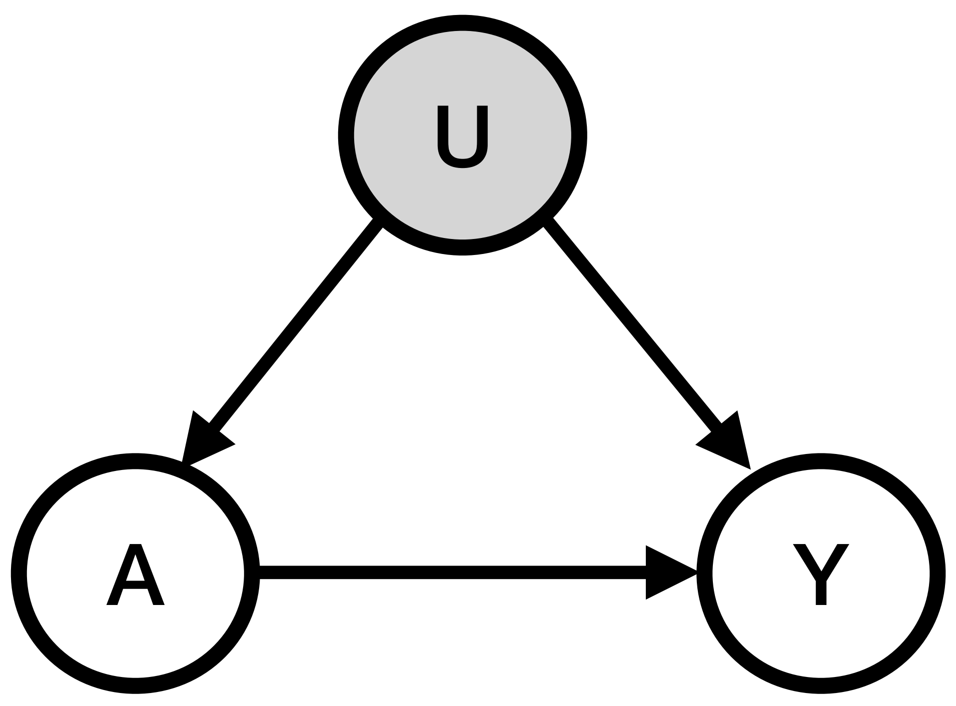


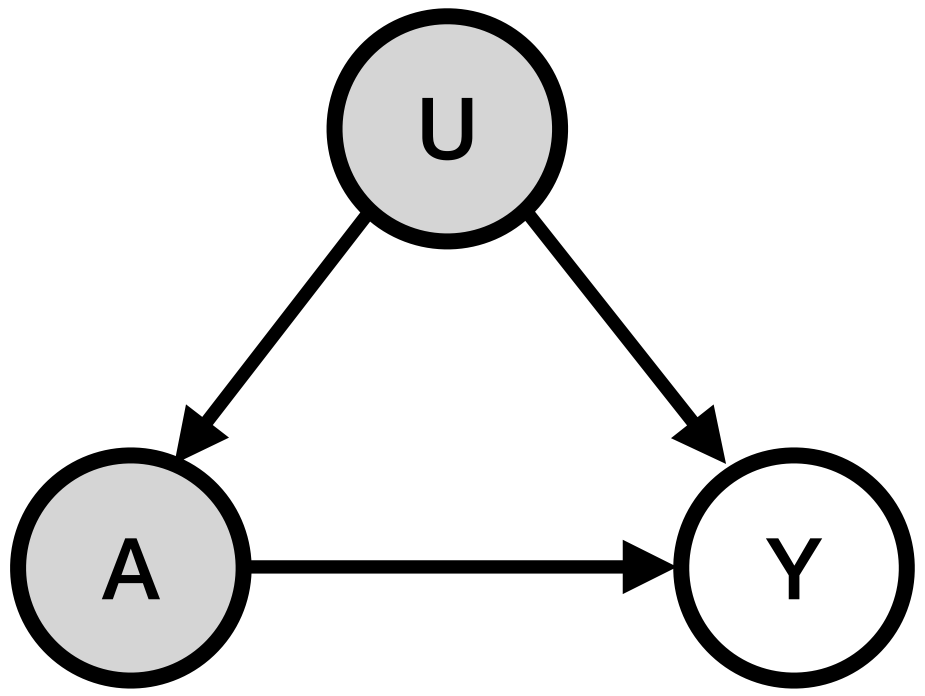
Backdoor adjustment: $\mathbb{E}[Y \,|\, do(a)] = \sum_{i=1}^{n} \mathbb{E}[Y \,|\, A = a, U = i]\mathbb{P}(U = i)$

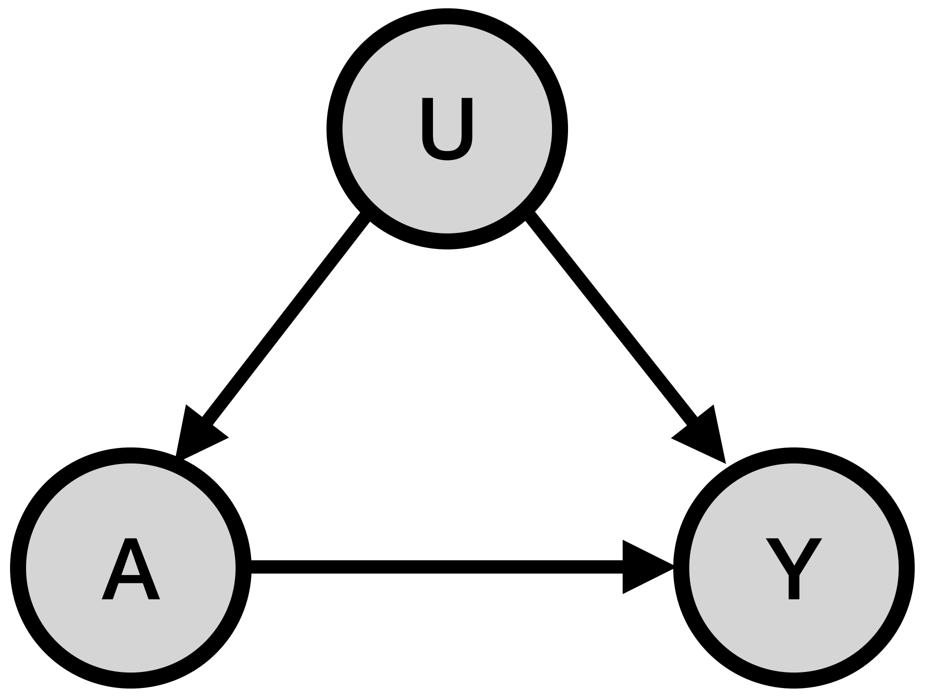
What are the obstacles?

- Disentanglement issues from (observed) U.
  - e.g. U might lead to imbalanced datasets.
- Non-identifiability from latent variables.

<— More on this next

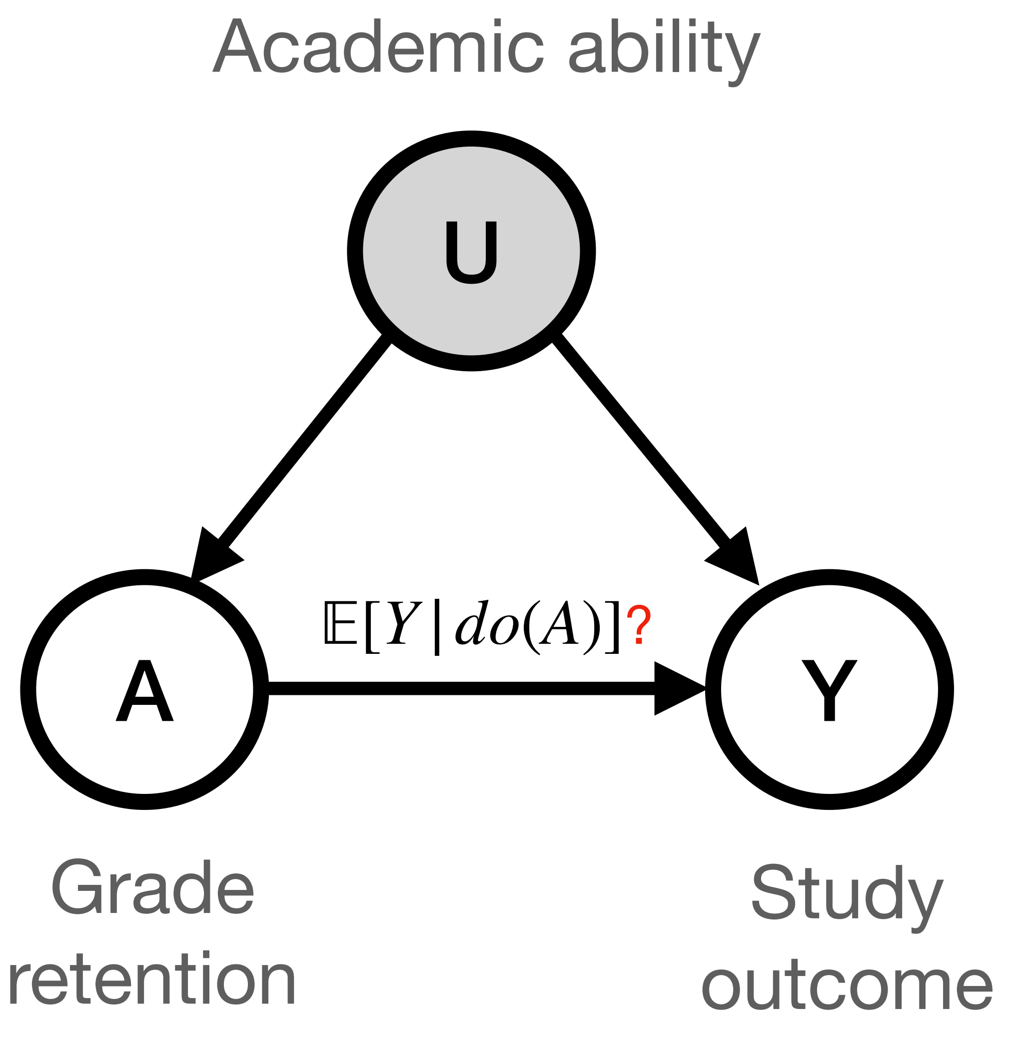




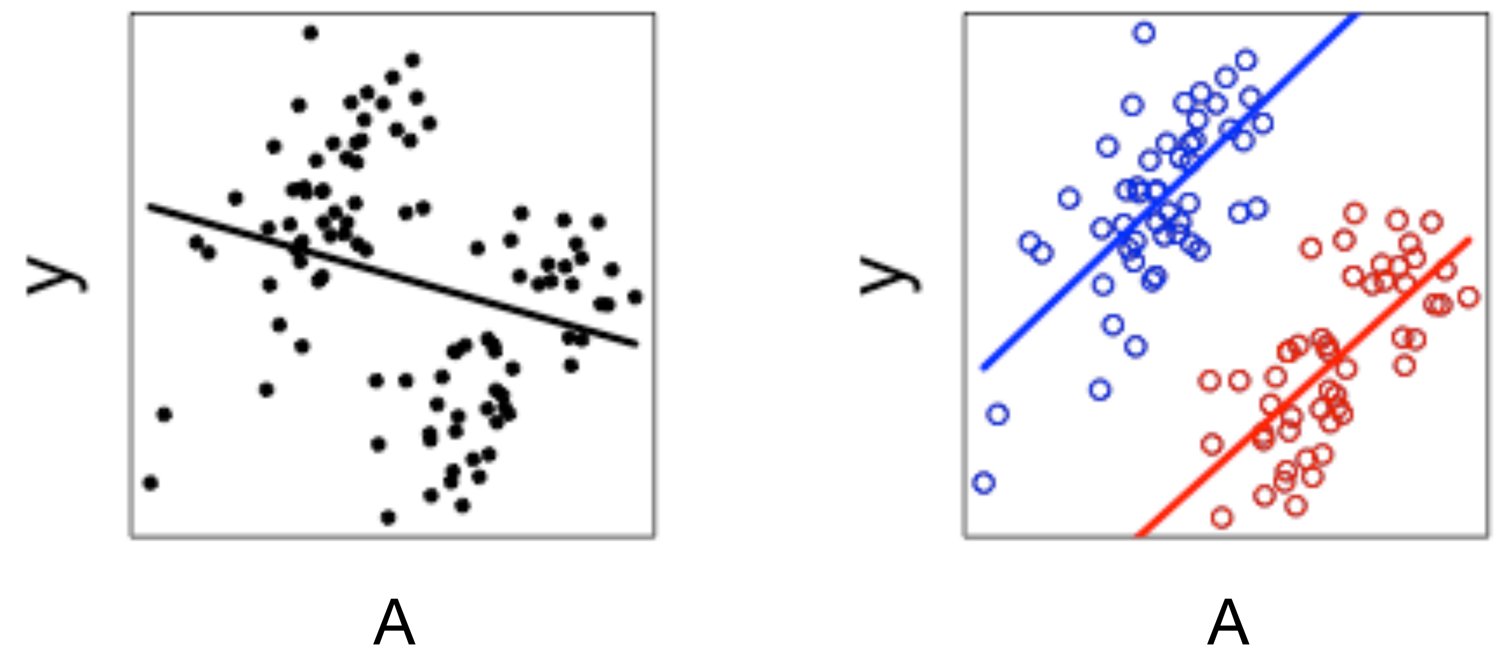


Why relax observability assumptions?

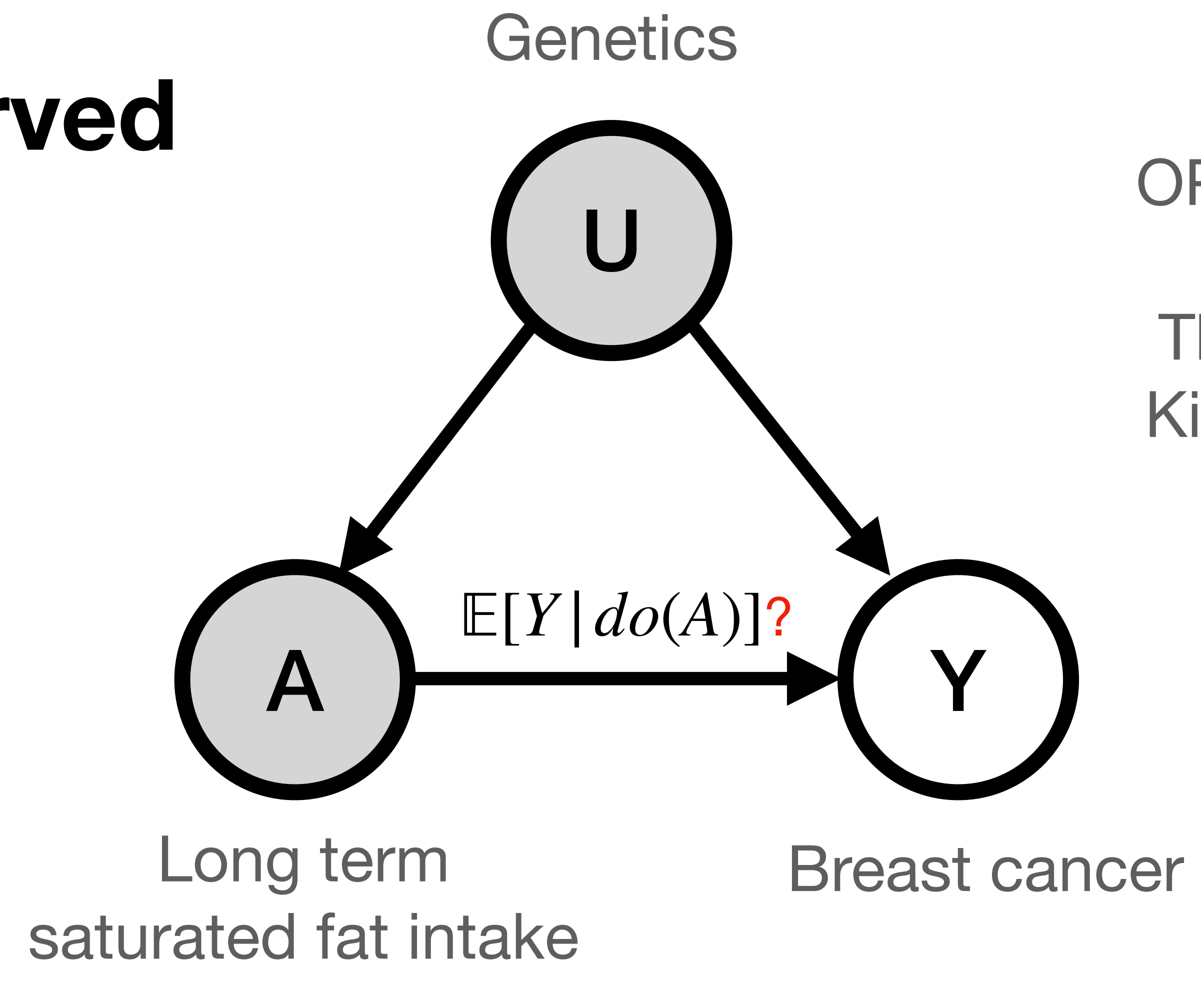
Unobserved confounders:



Simpson's paradox:

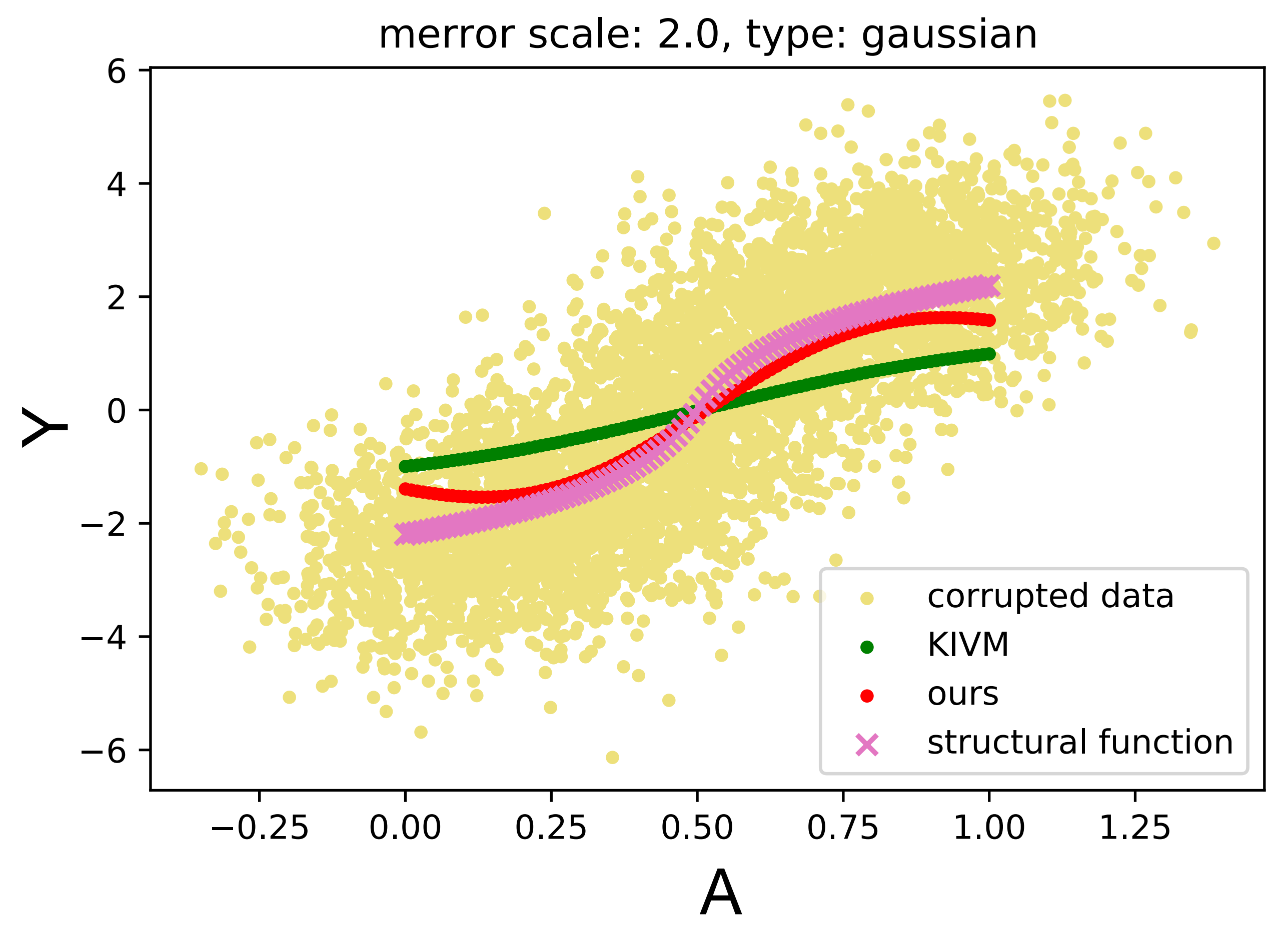


Action observed with error:

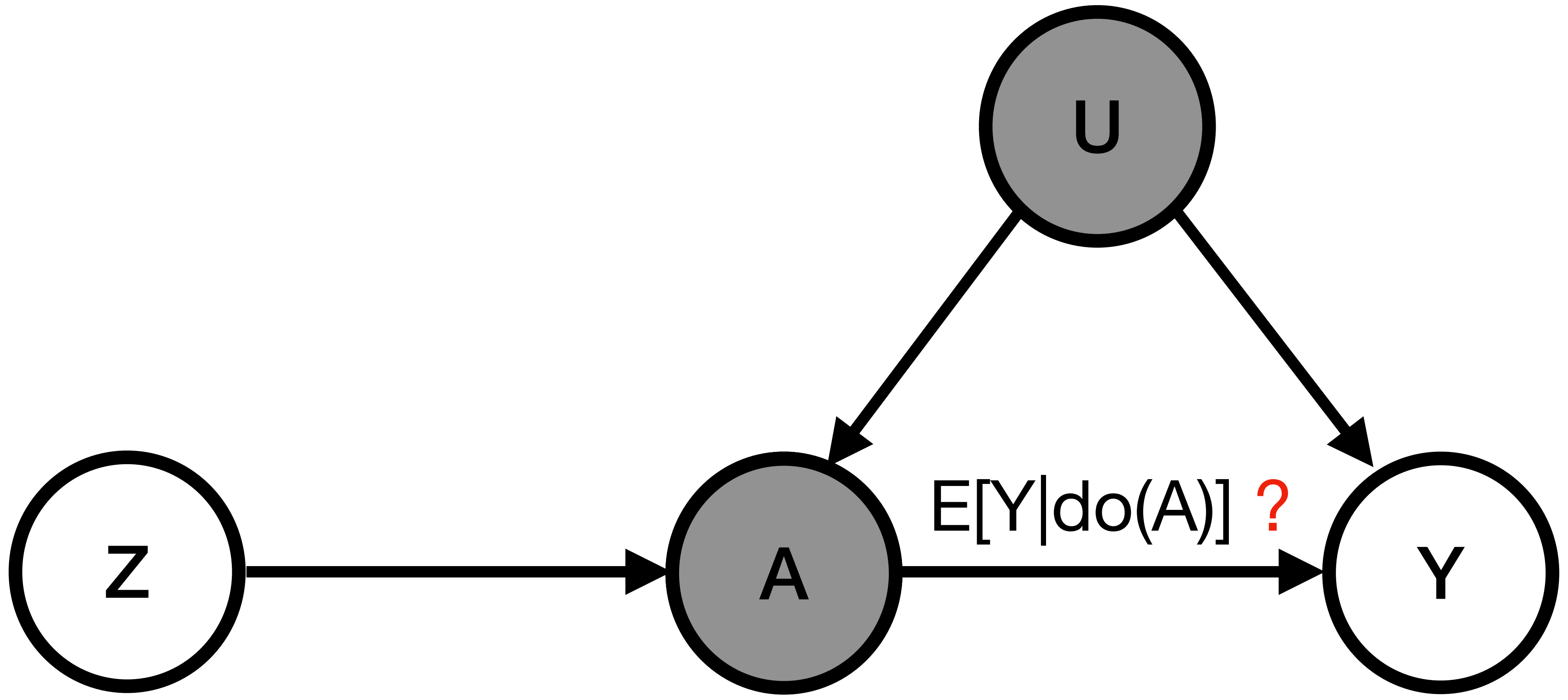


OPEN study: Subar, Thompson, Kipnis, et al. 2001

Mask interesting relationships:



Recap: Identification with instrumental variables



Identification:

$$Y = f(A) + \epsilon \quad \mathbb{E}[\epsilon | Z] = 0$$

$$f(A) = \mathbb{E}[Y | do(A)]$$

$$\mathbb{E}[Y|Z] = \int_{\mathscr{A}} f(a)p(a|Z)da$$

???

But if $f(a) = \theta^T \phi(a)$, then simplies to

$$\mathbb{E}[Y|Z] = \theta^T \mathbb{E}[\phi(A)|Z]$$

Linear case:

$$Y = \beta A + \epsilon_Y \quad \epsilon_Y \perp Z$$

$$A = \gamma Z + \epsilon_A \quad \epsilon_A \perp Z$$

$$\implies Y = \beta\gamma Z + \beta\epsilon_A + \epsilon_Y$$

(Strong) Assumptions:

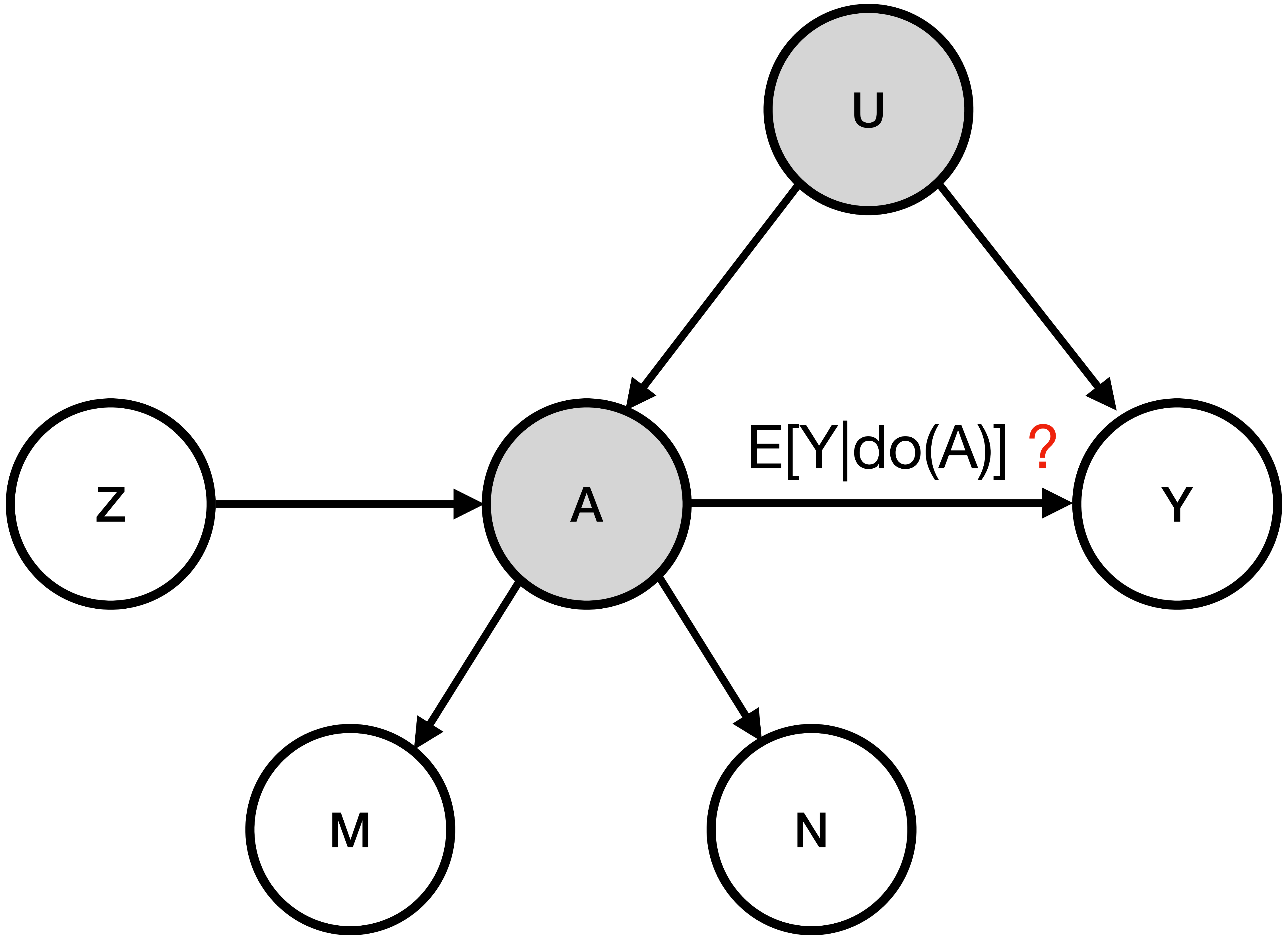
- **Additive error model**
- $(Z ! \perp A)_G$
- $(Z \perp Y)_{G_{\bar{A}}}$

To induce well-posedness:

- Assume f in RKHS.
- Tikhonov regularisation.

False IV: using same 'IV' for several different actions.

Measurement error on action variables



Measurement error on action variables - overview

Measurement error assumptions

Additive noise on measurements:

$M = A + \Delta M$

$N = A + \Delta N$

Independence/uncorrelatedness:

$\mathbb{E}[\Delta M \,|\, A, \Delta N] = 0$

$\Delta N \perp A$

Instrumental identification assumptions

Additive noise on nonlinear response:

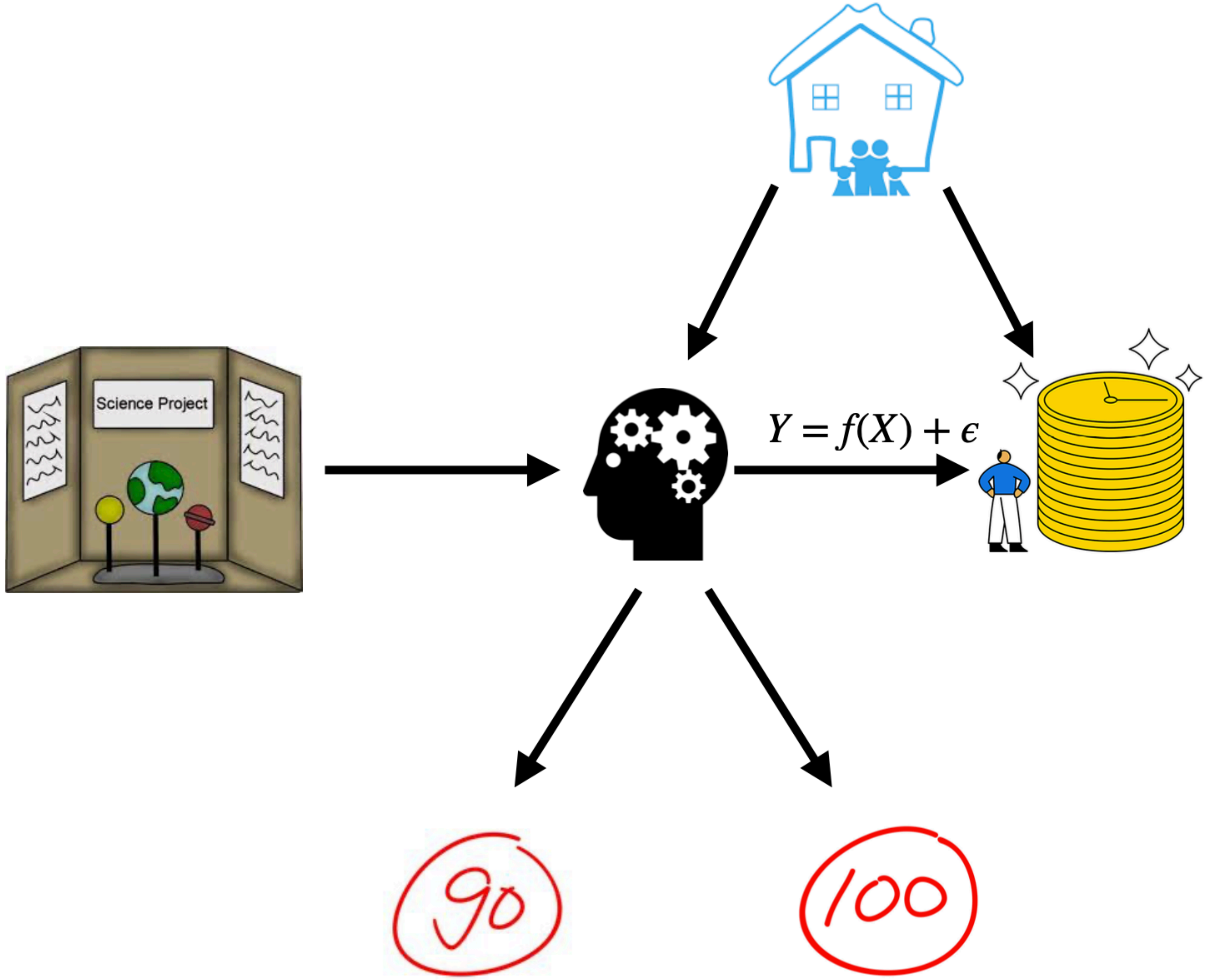
$Y = f(A) + \epsilon \quad \mathbb{E}[\epsilon \,|\, Z]$

IV structural assumptions:

$(Z! \perp A)_G$

$(Z \perp Y)_{G_{\bar{A}}}$

Application scenario



Measurement error on action variables - overview

Additive noise:

$$M = A + \Delta M$$

$$N = A + \Delta N$$

Independence/uncorrelatedness:

$$\mathbb{E}[\Delta M \,|\, A, \Delta N] = 0$$

$$\Delta N \perp A$$

Kotlarski's:

$$\overbrace{\mathbb{E}[e^{i\alpha A} \,|\, z]}^{\psi_{A|z}(\alpha)} = \exp\left(\int_0^\alpha i \frac{\mathbb{E}[M e^{i\nu N} \,|\, z]}{\mathbb{E}[e^{i\nu N} \,|\, z]} d\nu \right) \quad (1)$$

Parametric assumption:

- Use $\psi_{A|Z}$ to get expectations of basis functions by:
- Convolve FT of basis function and characteristic function.

Can we do better? Yes!

From $\hat{\psi}^n_{A|z}(\alpha)$ to $\hat{\mu}^n_{A|z}(y)$

$$\overbrace{\mathbb{E}_{\mathcal{P}_{A|z}}[e^{i\alpha X}](\alpha)}^{\psi_{A|Z}(\alpha)} = \exp\left(\int_0^\alpha i\frac{\mathbb{E}[Me^{i\nu N}|z]}{\mathbb{E}[e^{i\nu N}|z]}d\nu\right) \quad (1)$$

Have $\hat{\mu}^n_{A|z}(y) = \sum_{j=1}^{n} \hat{\gamma}^n_j(z)k(a_j, y)$.

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

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

Theorem 1. With translation-invariant, characteristic kernel:
$\hat{\mu}^n_{A|Z} \to^n \mu_{A|Z}$ iff $\hat{\psi}^n_{A|Z} \to^n \psi_{A|Z}$ in IFT of kernel.

Assume $f \in \mathcal{H}_A$:

$$\mathbb{E}[Y|Z] = \mathbb{E}[f(A)|Z] = \langle f, \mu_{A|Z}\rangle_{\mathcal{H}_A}$$

How to get $\mu_{A|Z}$?

Approximate $\hat{\mu}^n_{A|Z}$ via $\hat{\psi}^n_{A|Z}$ (eq.1):

Both only depend on $\hat{\gamma}^n_j(Z)$ and $\{a_j\}^n_{j=1}$!

Measurement Error KIV

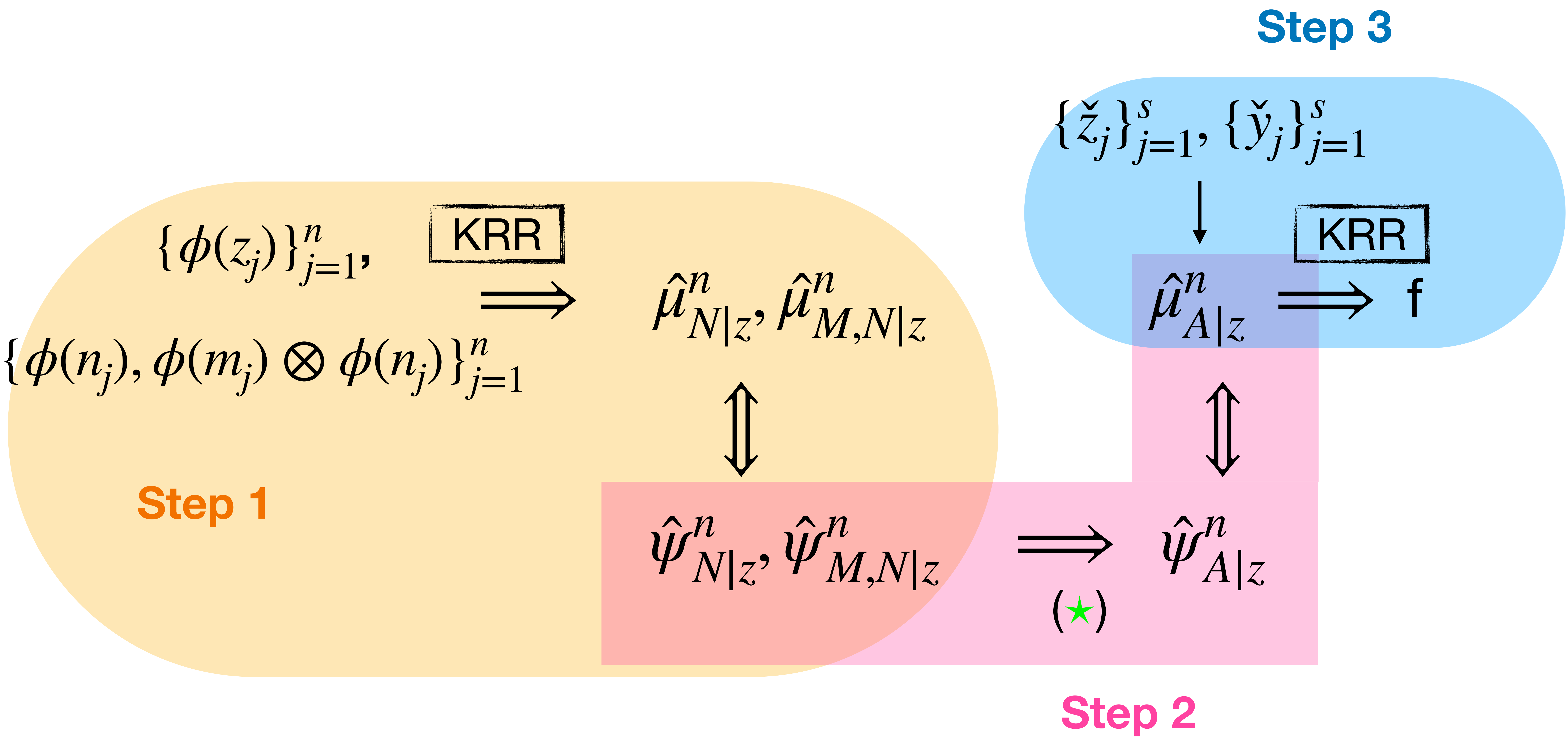
To obtain $\hat{\psi}^n_{A|z}$:

$$\overbrace{\mathbb{E}_{\mathscr{P}_{A|z}}[e^{i\alpha X}](\alpha)}^{\psi_{A|z}(\alpha)} = \exp\left(\int_0^\alpha i\frac{\overbrace{\mathbb{E}[Me^{i\nu N}\,|\,z]}^{\frac{\partial}{\partial v}\psi_{M,N|z}(v,\nu)\big|_{v=0}}}{\underbrace{\mathbb{E}[e^{i\nu N}\,|\,z]}_{\psi_{N|z}(\nu)}}d\nu\right) \quad (1)$$

1. **Differentiate wrt α to remove integral.**
2. **Replace with sample estimates.**

$$\frac{\frac{d}{d\alpha}\hat{\psi}^n_{A|z}(\alpha)}{\hat{\psi}^n_{A|z}(\alpha)} = \frac{\frac{\partial}{\partial v}\hat{\psi}^n_{M,N|z}(v,\alpha)\Big|_{v=0}}{\hat{\psi}^n_{N|z}(\alpha)} \quad (2)$$

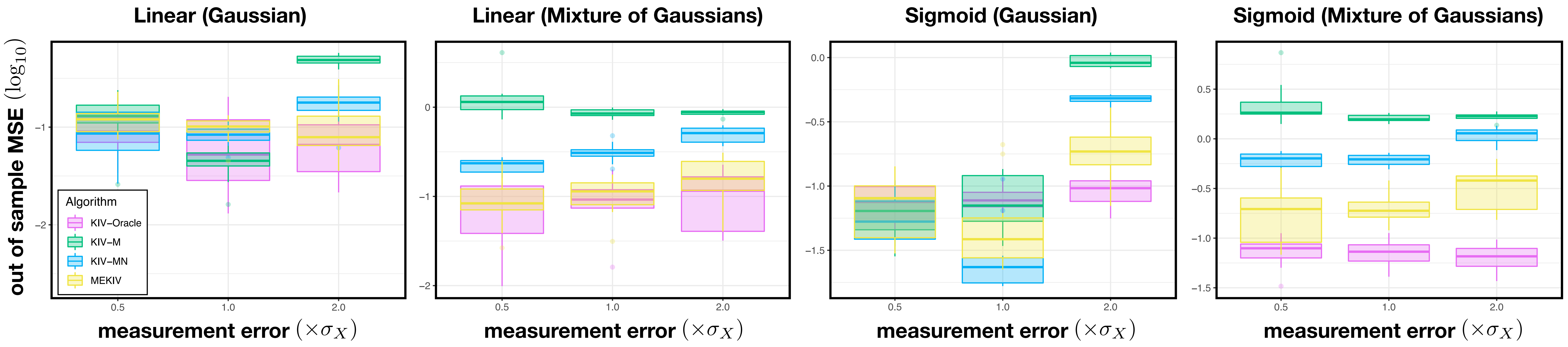
Measurement Error KIV



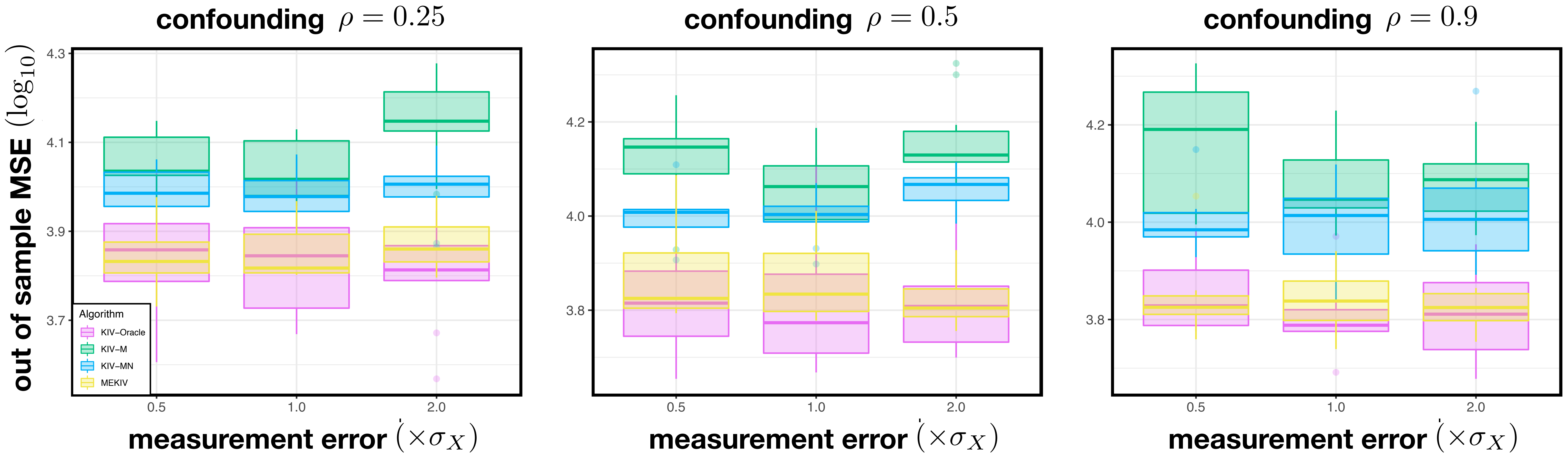
Advantages of MEKIV

- Free of distributional assumptions.
- Only modelling the conditional mean embedding, and not the full distribution.
- No need to train large number of hyperparameters.

MEKIV results



Demand Design (Mixture of Gaussians)



Future directions

- Relax the additive error assumption.
- Extend to sequential settings.

Thanks for listening! :)