From Latent to Deep Latent Variable Models

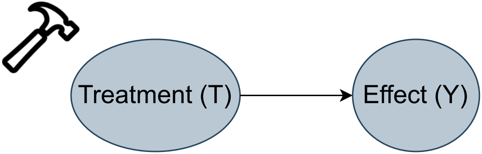
A Study On Causal Effect Inference with CEVAE

Valeria De Stasio, Christian Faccio, Giovanni Lucarelli

June 17, 2025

Objective: estimating causal effects

Estimate how a medical treatment T affects the health Y of a *random* patient.



$$P(Y = y|\text{do}(T = t)) = P(Y = y|T = t)$$

$$\text{ITE} = \mathbb{E}[(Y|\text{do}(T = 1)] - \mathbb{E}[(Y|\text{do}(T = 0)]$$

This is usually the case in a **randomized controlled trial** (RCT), where the treatment is randomly assigned to the patients.

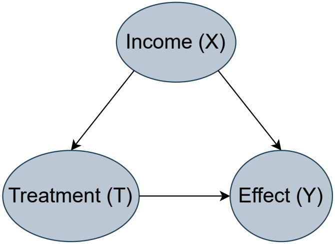
Confounder

In a observational study:

- **no** control over the treatment T assignment
  $\rightarrow$ there may be a confounder X (e.g. Income) that influences both variables

$$P(Y|\mathrm{do}(t)) \neq P(Y|t)$$

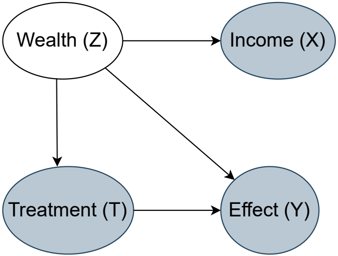
$$P(Y|\mathrm{do}(t), x) = P(Y|t, x)$$



If the confounder is observed even a linear rergession can do the job!

Latent confounder

But what if the confounder Z (e.g. Wealth) is **not** observed and the observed X is only a proxy of it?



Idea: estimate the latent variable and condition on it

Vanilla Latent Variable Model

- Assume parametric relationships between variables

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

$$X_j \sim \mathcal{N}(a\,z + b,\ \mathrm{diag}(\sigma_X^2))$$

$$T \sim \mathrm{Be}(\sigma(c\,z))$$

$$Y \sim \mathcal{N}(e\,t + f\,z,\ \sigma_Y^2)$$

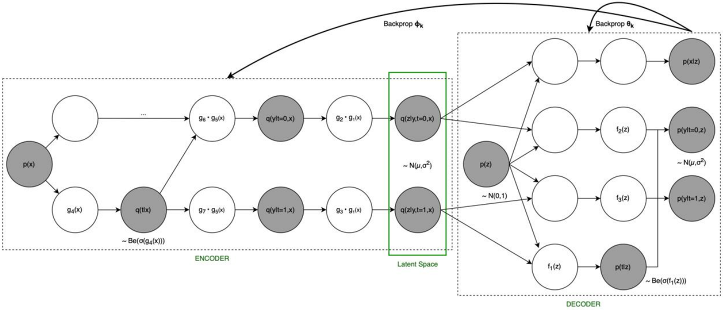
- Infer parameters using **Mean Field Variational Inference**
- For a new data point x:
  - infer: $z_0 = \mathbb{E}[Z|x, T = 0]$, $z_1 = \mathbb{E}[Z|x, T = 1]$
  - compute ITE $= \mathbb{E}[Y|T = 1, z_1] - \mathbb{E}[Y|T = 0, z_0]$
- Problem: we need to train a new model at test time!

Deep Latent Variable Model: CEVAE

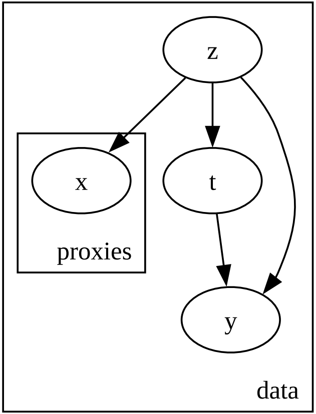
- Assume parametric relationships between variables through Neural Networks
- **CEVAE** = VAE + causal semantics for latent Z

$$(X,\ T,\ Y) \longrightarrow\ Z\ \longrightarrow\ (X,\ T,\ Y)$$

- Counterfactuals in one forward pass $\Rightarrow$ **amortized inference, no test-time training**



Synthetic linear dataset



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

$$a_j \sim \mathcal{U}(-10,10)$$

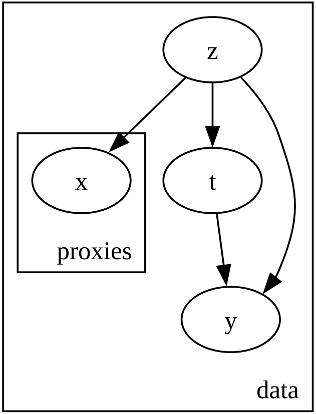
$$X_j \sim \mathcal{N}(a_j z, \sigma_X^2)$$

$$T|Z \sim \mathrm{Be}(\sigma(\beta z))$$

$$Y|T,Z \sim \mathcal{N}(z+t, \sigma_Y^2)$$

Remark: ITE is constant for all x!

Synthetic non linear dataset



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

$$a \sim \mathcal{U}(-10, 10)^d$$

$$X_1, ..., X_d \sim \mathcal{N}(a \tanh(z), \sigma_X^2 \mathbb{I})$$

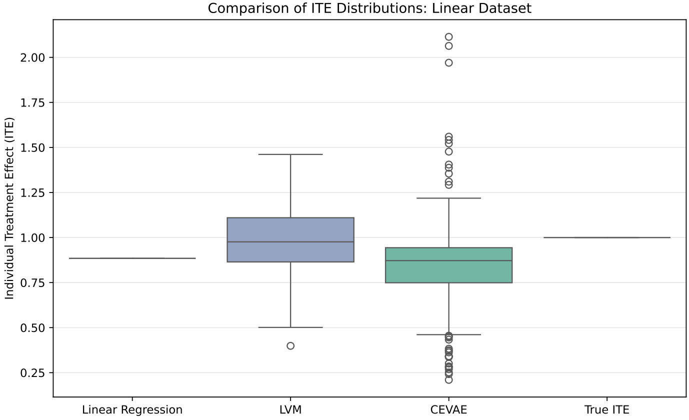
$$T|Z \sim \mathrm{Be}(\sigma(\beta z))$$

$$Y|T, Z \sim \mathcal{N}\left(\mathrm{NonLin}(z, t), \sigma_Y^2\right)$$

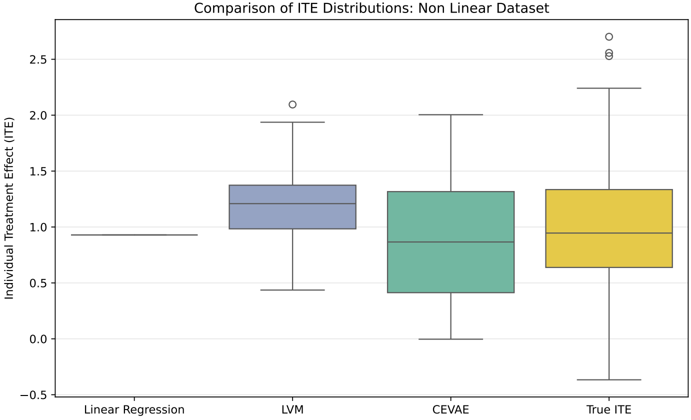
$$\mathrm{NonLin}(z, t) = \sin(z) + \frac{1}{2}z + t\left(1 + \frac{1}{2}z\right)$$

Remark: ITE depends on interaction between t and z!

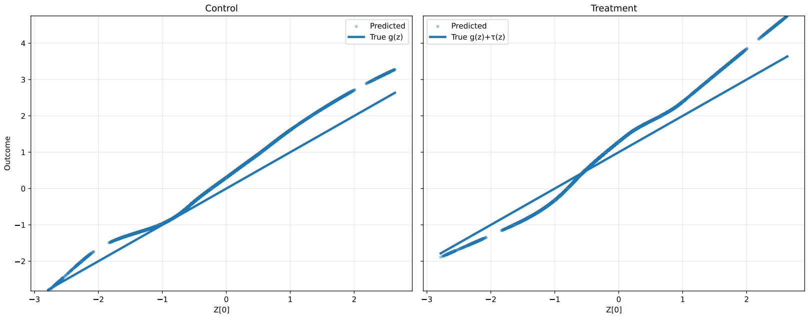
Results (1/2): Linear Dataset



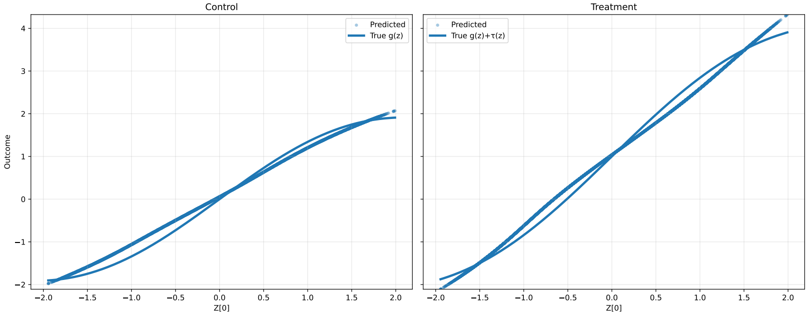
Results (2/2): Non Linear Dataset



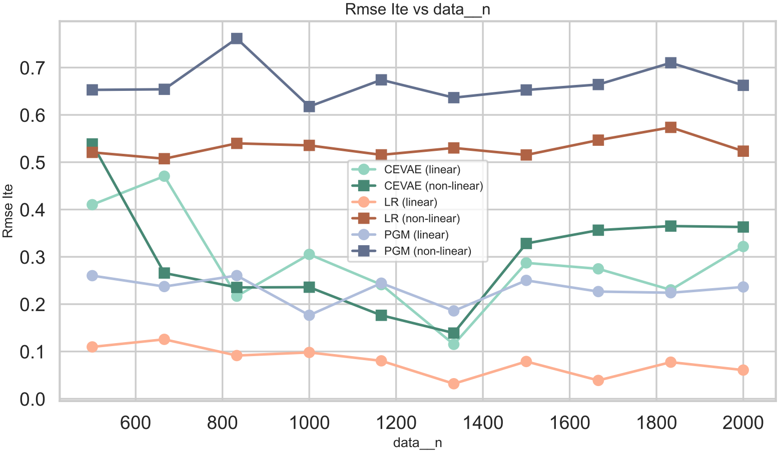
Internal Representation: Linear Dataset



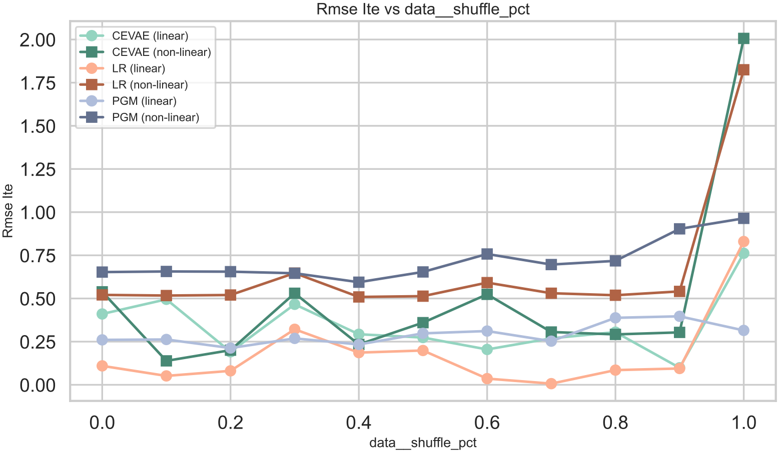
Internal Representation: Non Linear Dataset



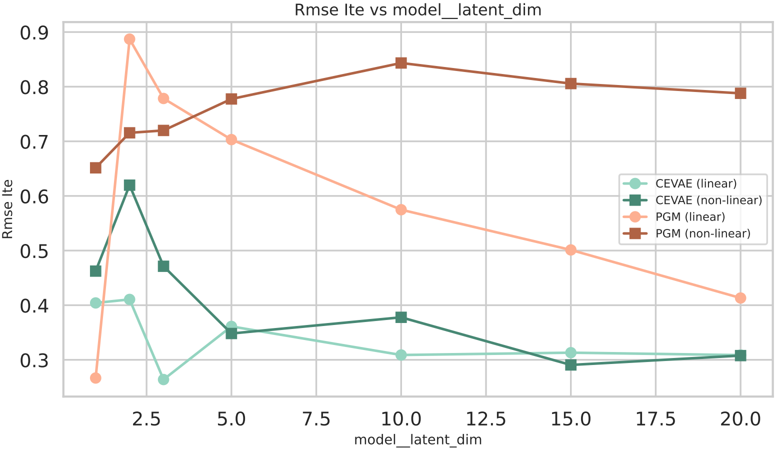
Experiment 1: Increasing the sample size



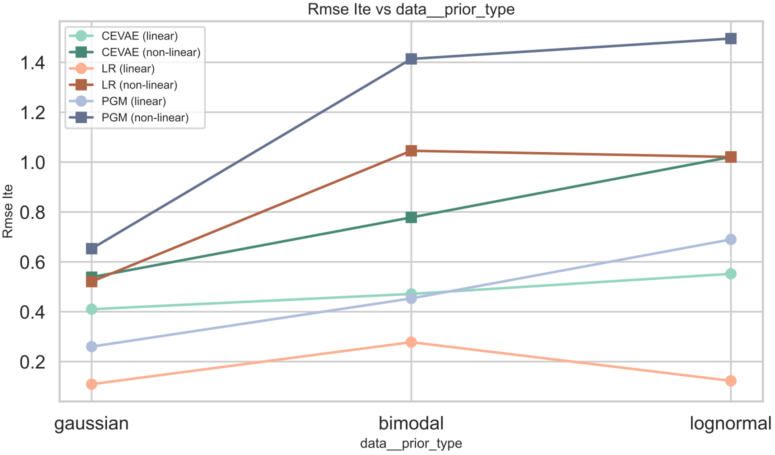
Experiment 2: Increasing decorrelation among proxies



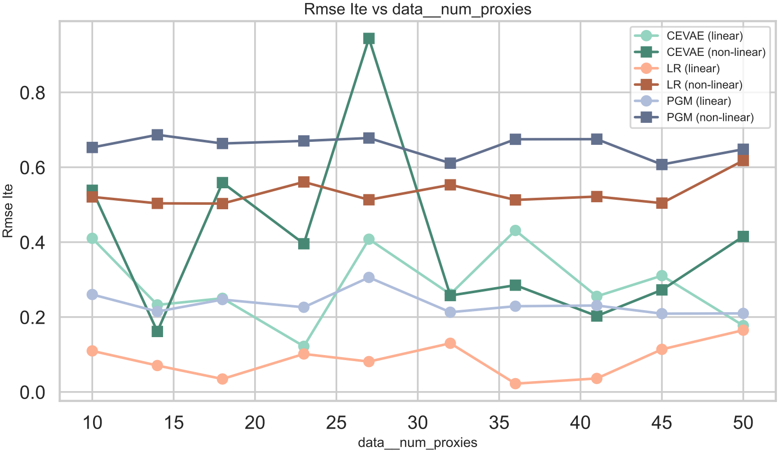
Experiment 3: Increasing latent dimension



Experiment 4: latent distribution misspecified



Experiment 5: Increasing the number of proxies



Conclusions

Pros

- Captures *non-linear* causal links
- Counterfactuals in one forward pass
- Fewer hand-crafted assumptions

Cons

- Computationally intensive training
- Sensitive to the hyperparameters

Thank You!