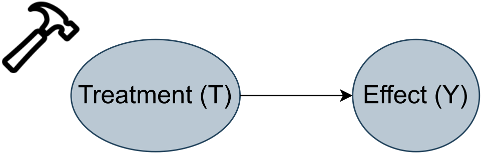
From Latent to Deep Latent Variable Models

A Study On Causal Effect Inference with CEVAE

Valeria De Stasio, Christian Faccio, Giovanni Lucarelli

June 17, 2025

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



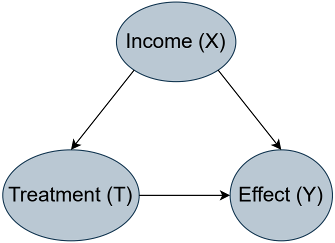
$$P(Y = y|\mathrm{do}(T = t)) = P(Y = y|T = t)$$

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

In a **observational** study no control over the treatment T assignment $\rightarrow$ there may be a confounder X (e.g. Income) that influences both T and Y.

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

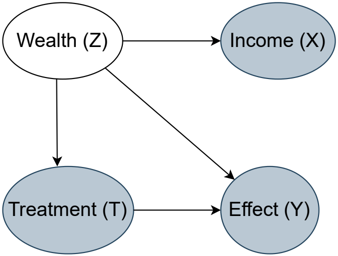
$$P(Y|\text{do}(t), x) = P(Y|t, x)$$



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

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

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

Methodology

Two different synthetic datasets, relationships between variables are:

- **Linear** functions
- **Non-linear** functions

To address this problem:

- Linear Regression
- Latent Variable Model (LVM)
- Deep Latent Variable Model (CEVAE)

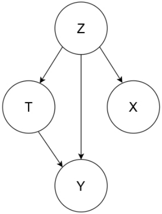
- Generative model:

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

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

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

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



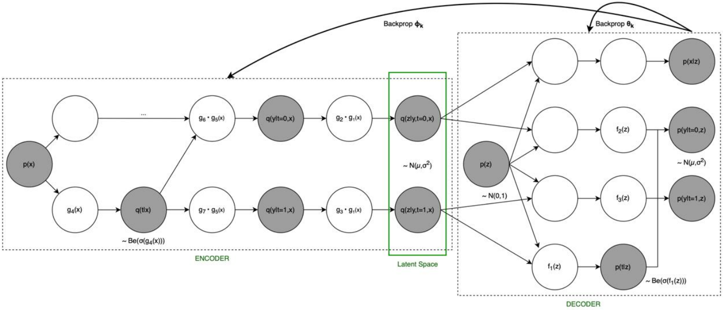
- **SVI** on both the generative parameters and the posterior distribution

Latent Variable Model (LVM):

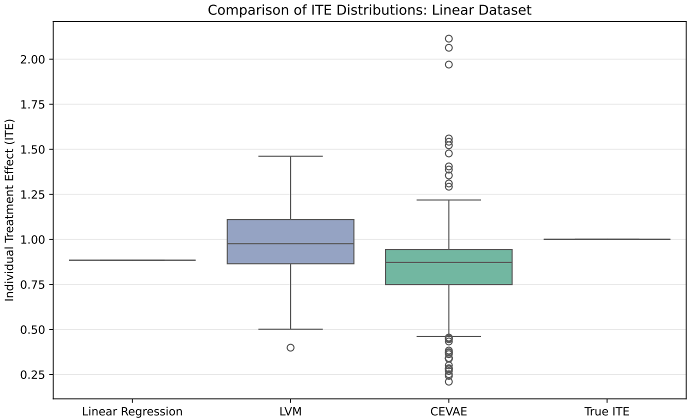
- Linear functions
- SVI + mini-SVI for new x

Deep Latent Variable Model (CEVAE):

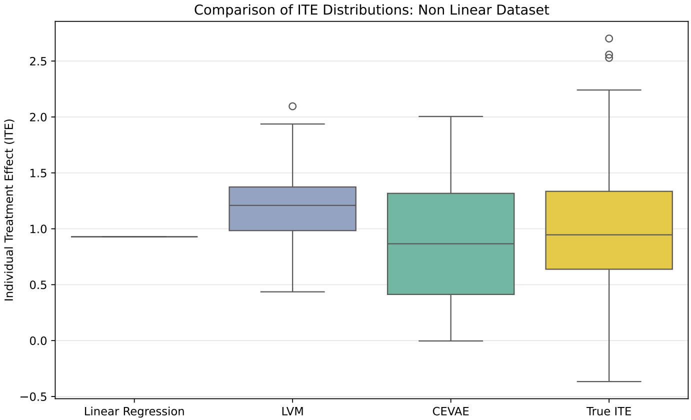
- MLP functions
- SVI + Amortized inference



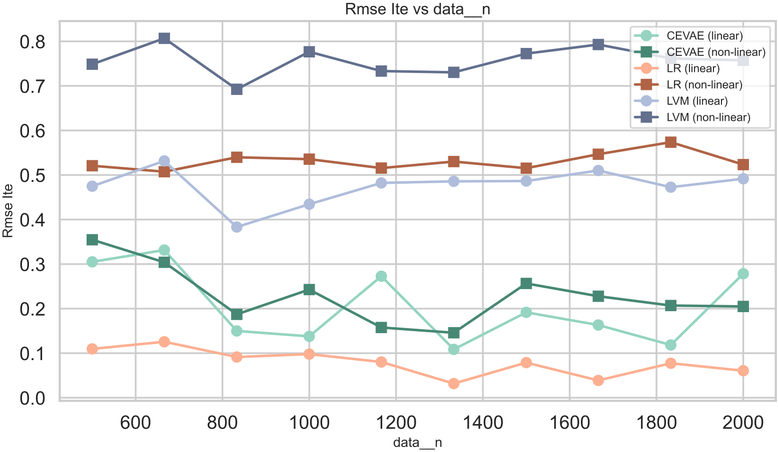
Results



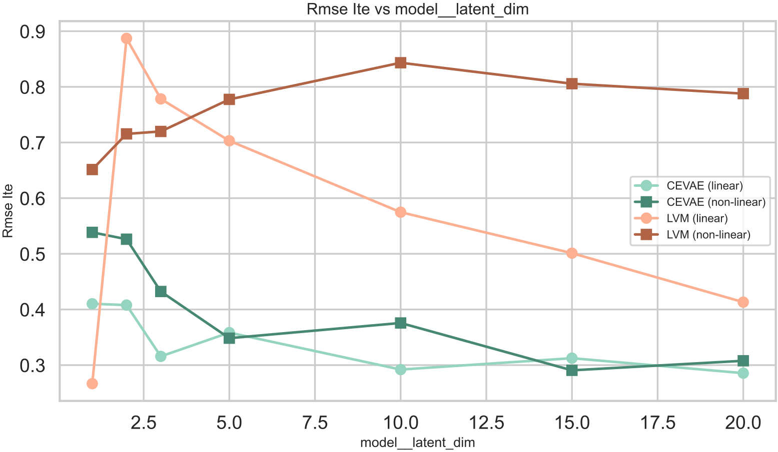
Comparison of ITE Distributions: Linear Dataset
Individual Treatment Effect (ITE)
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
Linear Regression
LVM
CEVAE
True ITE


Comparison of ITE Distributions: Non Linear Dataset
Individual Treatment Effect (ITE)
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Linear Regression
LVM
CEVAE
True ITE

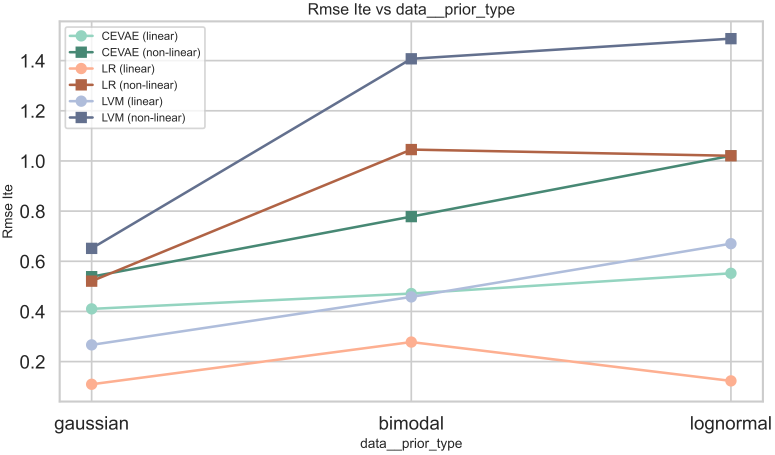
Increasing the sample size



Increasing latent dimension



Latent distribution misspecified

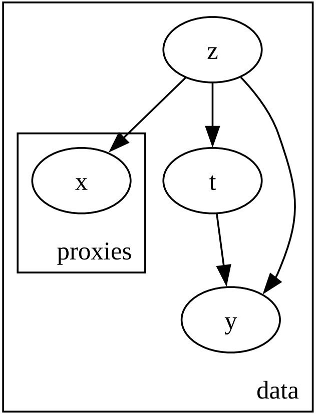


Conclusions

- CEVAE performs better on non-linear synthetic data, as more flexible
- Overcomplicates simple problems, e.g. linear dataset
- Does not need strong assumptions, being more robust than the LVM
- Next steps – Test on real datasets, explore more sophisticate variant of CEVAE (e.g. *TEDVAE*, *DCEVAE*, *ICEVAE*)

Thank You!

Synthetic linear dataset



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

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

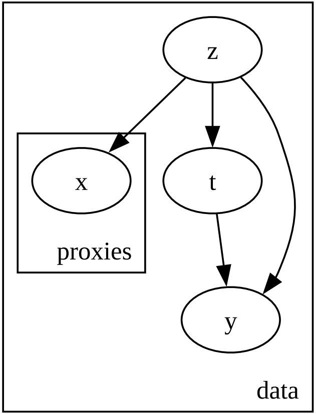
$$X_j \sim \mathcal{N}(a_j z, \sigma_X^2)$$

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

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

Remark: ITE is constant for all x!

Synthetic non linear dataset



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

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

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

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

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

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

Remark: ITE depends on interaction between t and z!