Estimation of Individual Treatment Effect in Latent Confounder Models via Adversarial Learning

CHANGHEE LEE[1], NICHOLAS MASTRONARDE[2], AND MIHAELA VAN DER SCHAAR[3,1]
[1]UNIVERSITY OF CALIFORNIA LOS ANGELES (UCLA), [2] UNIVERSITY AT BUFFALO, [3] UNIVERSITY OF OXFORD

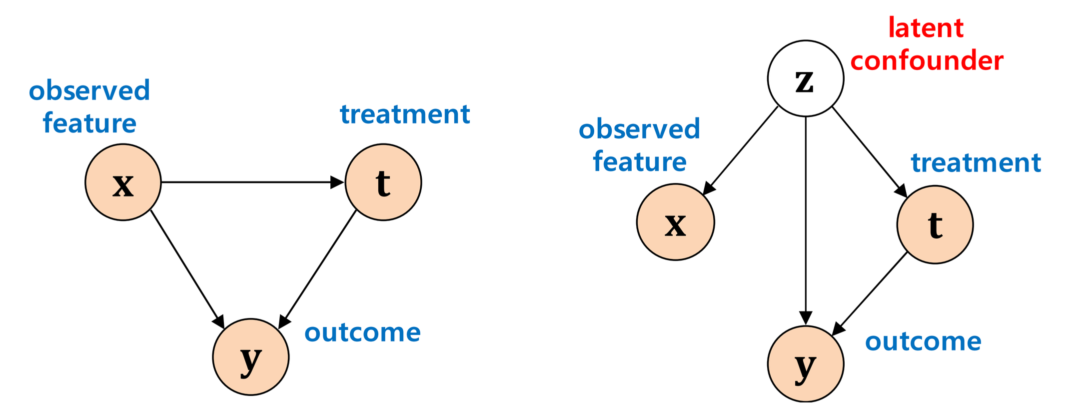
1. INTRODUCTION

Estimating Individual Treatment Effect (ITE)

- A key challenge with observational data
  → Will treatment A help patient B recover?
- Most previous work relies on the *un-confoundedness assumption*, which posits that all the confounders are measurable; see Figure 1(a)

Latent Confounder Model

- In practice, there are often unmeasurable (latent) confounders; see Figure 1(b)
  → Socio-economic status affects medications available to a patient and her health
- If not appropriately accounted for, the estimated ITE will be subject to *confounding bias*



(a) No latent confounding (b) Latent confounding

Figure 1: Causal diagrams

2. PROBLEM FORMULATION

Observational Dataset: $\mathcal{D} = \{(\mathbf{x}_i, t_i, \mathbf{y}_i)\}_{i=1}^N$

- $\mathbf{x}_i$: feature vector
- t_i: treatment (we assume $t \in \{0,1\}$)
- $\mathbf{y}_i$: outcome vector
- $\mathbf{z}_i$: latent confounder that is not in $\mathcal{D}$

Objective:

- **Estimate ITE** without confounding bias:

$$ITE(\mathbf{x}) = \mathbb{E}[\mathbf{y}|\mathbf{x}, do(t=1)] - \mathbb{E}[\mathbf{y}|\mathbf{x}, do(t=0)]$$

How to Account for Latent Confounding?

- We assume the latent confounder model in Figure 1(b); $\mathbf{x}$ is treated as a *proxy variable* that provides a noisy view of $\mathbf{z}$
- We can identify $p(\mathbf{y}|\mathbf{x}, do(t=1))$ (or, similarly, $p(\mathbf{y}|\mathbf{x}, do(t=0))$) by

$$p(\mathbf{y}|\mathbf{x}, do(t=1)) = \int_{\mathbf{z}} p(\mathbf{y}|\mathbf{z}, \mathbf{x}, do(t=1)) p(\mathbf{z}|\mathbf{x}, do(t=1)) d\mathbf{z}$$
$$= \int_{\mathbf{z}} p(\mathbf{y}|\mathbf{z}, \mathbf{x}, t=1) p(\mathbf{z}|\mathbf{x}) d\mathbf{z},$$

- We adopt an **adversarial learning framework** to learn $p(\mathbf{y}|\mathbf{z}, \mathbf{x}, t)$ and $p(\mathbf{z}|\mathbf{x})$

6. PERFORMANCE METRIC

Precision in Estimation of Heterogeneous Effect (PEHE)

- A commonly used metric to quantify the *goodness* of ITE estimation

$$\epsilon_{\text{PEHE}} = \frac{1}{N}\sum_{i=1}^{N}\Big((y_i(1) - y_i(0)) - (\hat{y}_i(1) - \hat{y}_i(0))\Big)^2$$

7. BENCHMARKS

- CFR_{WASS}: counterfactual reg. w/ Wasserstein
- CMGP: causal multi-task Gaussian process
- CEVAE: causal effect VAE (CEVAE)

3. CEGAN ARCHITECTURE & COMPONENTS

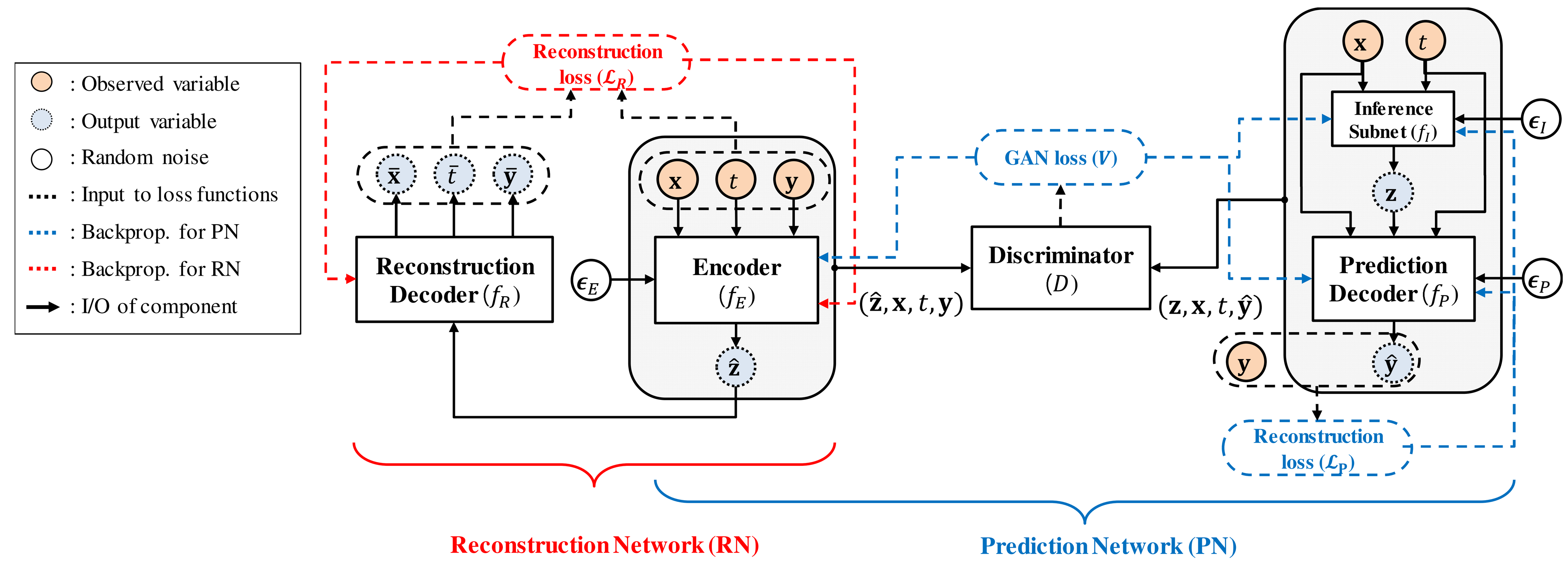


Figure 2: CEGAN architecture

Prediction Network:

- **Generator (G):**
  - Comprises encoder (f_E), inference subnet (f_I), and prediction decoder (f_P) which output $\hat{\mathbf{z}} \sim q_E(\mathbf{z}|\mathbf{x}, t, \mathbf{y})$, $\mathbf{z} \sim q_I(\mathbf{z}|\mathbf{x}, t)$, $\hat{\mathbf{y}} \sim q_P(\mathbf{y}|\mathbf{z}, \mathbf{x}, t)$ (via universal approximator technique)
  - Constructs samples of tuples $(\mathbf{z}, \mathbf{x}, t, \mathbf{y})$ drawn from two joint distributions, i.e., $q_E(\mathbf{z}, \mathbf{x}, t, \mathbf{y}) = p_d(\mathbf{x}, t, \mathbf{y}) q_E(\mathbf{z}|\mathbf{x}, t, \mathbf{y})$ and $q_P(\mathbf{z}, \mathbf{x}, t, \mathbf{y}) = p_d(\mathbf{x}, t) q_I(\mathbf{z}|\mathbf{x}, t) q_P(\mathbf{y}|\mathbf{z}, \mathbf{x}, t)$
  - Tries to fool the discriminator into believing the tuples are drawn from the same distribution
- **Discriminator (D):**
  - Distinguishes between tuples $(\mathbf{z}, \mathbf{x}, t, \mathbf{y})$ that are drawn from $q_E(\mathbf{z}, \mathbf{x}, t, \mathbf{y})$ and $q_P(\mathbf{z}, \mathbf{x}, t, \mathbf{y})$

Reconstruction Network:

- Comprises the same encoder (f_E) and reconstruction decoder (f_R)
- Nudge f_E to learn a meaningful mapping by reconstructing its original input

CEGAN *matches the two distribution* by playing an *adversarial game* between G and D.

4. EXPERIMENTS: SEMI-SYNTHETIC

TWINS Dataset:

- Based on records of twin births in the USA from 1989-1991
- Artificially create a binary treatment: $t = 1$ ($t = 0$) denotes being born the heavier (lighter)
- Outcome corresponds to the mortality of each of the twins in their first year

Data Generation Process:

- Select `GESTAT` (i.e. the gestational age in weeks) as the latent confounder z.
- Assign binary treatment $t_i \sim \text{Bern}(\sigma(wz_i))$, where $w \sim \mathcal{N}(10, 0.1^2)$.
- Choose outcome of the heavier twin, $y_i(1)$, if $t_i = 1$ and that of the lighter twin, $y_i(0)$, if $t_i = 0$.

Table 1: Comparison of $\sqrt{\epsilon_{\text{PEHE}}}$ (mean ± std)

Method	*no latent confounding*		*latent confounding*	
	In-sample	Out-sample	In-sample	Out-sample
LR-1	0.365±0.00	0.367±0.00	0.413±0.01	0.423±0.02
LR-2	0.404±0.02	0.411±0.02	0.442±0.02	0.454±0.02
kNN	0.486±0.02	0.506±0.02	0.492±0.02	0.515±0.02
CForest	**0.356±0.01**	0.372±0.01	0.417±0.02	0.429±0.02
CMGP	0.367±0.01	0.365±0.01	0.430±0.05	0.438±0.05
CFR_{WASS}	0.371±0.03	0.371±0.03	0.427±0.05	0.438±0.05
CEVAE	0.363±0.00	0.364±0.00	0.423±0.00	0.428±0.00
CEGAN	0.363±0.00	**0.362±0.00**	**0.369±0.00**	**0.369±0.00**

- *"no latent confounding"* includes `GESTAT` in the observational data $\mathcal{D}$
  - Causal model → Figure 1(a)
- *"latent confounding"* excludes `GESTAT` from the observational data $\mathcal{D}$
  - Causal model → Figure 1(b)

5. EXPERIMENTS: SYNTHETIC

Toy Example:

- To assess the robustness of CEGAN to latent confouderers (due to noise in the proxy variables)

Data Generation Process:

- Assume latent confounding model in Figure 1(b):

$$z_{ij} \sim \mathcal{N}(3(\mu - 1), 1^2), \quad j = 1, \dots, d_z,$$
$$\mu \sim \text{Bern}(0.5), \quad \mathbf{n} \sim \mathcal{N}(0, \zeta^2 \mathbf{I})$$
$$\mathbf{x}_i | \mathbf{z}_i = \mathbf{z}_i + \mathbf{n},$$
$$t_i | \mathbf{z}_i \sim \text{Bern}(\sigma(0.25 \cdot z_{id_z})),$$
$$y_i | \mathbf{z}_i, t_i = \sigma(\mathbf{1}^T \mathbf{z}_i + (2t_i - 1))$$

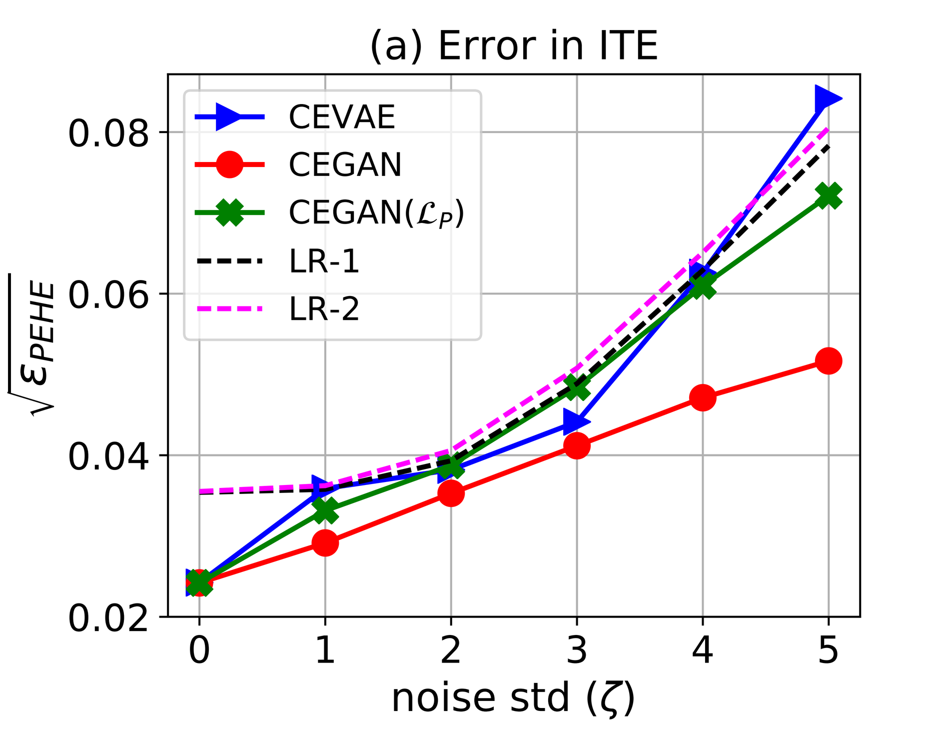


Figure 3: $\sqrt{\epsilon_{\text{PEHE}}}$ with respect to noise level ζ