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 unconfoundedness 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

2. PROBLEM FORMULATION

Observational Dataset: $D = \{(x_i, t_i, y_i)\}_{i=1}^N$
• $x_i$: feature vector
• $t_i$: treatment (we assume $t \in \{0, 1\}$)
• $y_i$: outcome vector
• $z_i$: latent confounder that is not in $D$

Objective: Estimate ITE without confounding bias: $ITE(x) = \mathbb{E}[y|x, do(t = 1)] - \mathbb{E}[y|x, do(t = 0)]$

How to Account for Latent Confounding?
• We assume the latent confounder model in Figure 1(b): $x$ is treated as a proxy variable that provides a noisy view of $z$
• We can identify $p(y|x, do(t = 1))$ (or, similarly, $p(y|x, do(t = 0))$) by
$$p(y|x, do(t = 1)) = \int p(y|x, do(t = 1)) p(z|x, do(t = 1)) dz$$
• We adopt an adversarial learning framework to learn $p(y|x, z, t)$ and $p(z|x)$

3. CEGAN ARCHITECTURE & COMPONENTS

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

Reconstruction Network (RN)
• 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 gestation (i.e. the gestational age in weeks) as the latent confounder $z$
• Assign binary treatment $t_i \sim \text{Bern}(\sigma(wz_i))$, where $w \sim \mathcal{N}(0, 0.1^2)$
• 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 $\sqrt{\text{MSE}}$ (mean ± std)

<table>
<thead>
<tr>
<th>Method</th>
<th>no latent confounding</th>
<th>latent confounding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-sample</td>
<td>Out-sample</td>
</tr>
<tr>
<td>LR-1</td>
<td>0.365±0.00</td>
<td>0.367±0.00</td>
</tr>
<tr>
<td>LR-2</td>
<td>0.404±0.02</td>
<td>0.411±0.02</td>
</tr>
<tr>
<td>kNN</td>
<td>0.495±0.02</td>
<td>0.506±0.02</td>
</tr>
<tr>
<td>CForest</td>
<td>0.356±0.01</td>
<td>0.372±0.01</td>
</tr>
<tr>
<td>CMGP</td>
<td>0.367±0.01</td>
<td>0.365±0.01</td>
</tr>
<tr>
<td>CFRWASS</td>
<td>0.371±0.03</td>
<td>0.371±0.03</td>
</tr>
<tr>
<td>CEVAE</td>
<td>0.363±0.00</td>
<td>0.364±0.00</td>
</tr>
<tr>
<td>CEGAN</td>
<td>0.363±0.00</td>
<td>0.362±0.00</td>
</tr>
</tbody>
</table>

“no latent confounding” includes gestation in the observational data $D$
• Causal model $\rightarrow$ Figure 1(a)
• “latent confounding” excludes gestation from the observational data $D$
• Causal model $\rightarrow$ Figure 1(b)

5. EXPERIMENTS: SYNTHETIC

Toy Example:
• To assess the robustness of CEGAN to latent confounders (due to noise in the proxy variables)

Data Generation Process:
• Assume latent confounding model in Figure 1(b):
  $z_{ij} \sim \mathcal{N}(3(\mu - 1), 1^2)$, $j = 1, \ldots, d_z$
  $\mu \sim \text{Bern}(0.5)$
  $n \sim \mathcal{N}(0, \zeta^2 I)$
  $x_{ij} | z_i = n_i$
  $t_i | z_i \sim \text{Bern}(\sigma(0.25z_i, 1 - \sigma(0.25z_i, 1)))$
  $y_i | x_i, t_i = \sigma(1^T z_i + (2t_i - 1))$

Figure 3: $\sqrt{\text{MSE}}$ with respect to noise level $\zeta$