Causal Inference with Conditional Instruments Using Deep Generative Models

Debo Cheng[1,2*], **Ziqi Xu**[2*], **Jiuyong Li**[2], **Lin Liu**[2], **Jixue Liu**[2] **and Thuc Duy Le**[2]

[1] School of Computer Science and Engineering, Guangxi Normal University, Guilin, 541004, China
[2] UniSA STEM, University of South Australia, Adelaide, SA, 5095, Australia
{debo.cheng,ziqi.xu}@mymail.unisa.edu.au, {jiuyong.li,lin.liu,jixue.liu,thuc.le}@unisa.edu.au

Abstract

The instrumental variable (IV) approach is a widely used way to estimate the causal effects of a treatment on an outcome of interest from observational data with latent confounders. A standard IV is expected to be related to the treatment variable and independent of all other variables in the system. However, it is challenging to search for a standard IV from data directly due to the strict conditions. The conditional IV (CIV) method has been proposed to allow a variable to be an instrument conditioning on a set of variables, allowing a wider choice of possible IVs and enabling broader practical applications of the IV approach. Nevertheless, there is not a data-driven method to discover a CIV and its conditioning set directly from data. To fill this gap, in this paper, we propose to learn the representations of the information of a CIV and its conditioning set from data with latent confounders for average causal effect estimation. By taking advantage of deep generative models, we develop a novel data-driven approach for simultaneously learning the representation of a CIV from measured variables and generating the representation of its conditioning set given measured variables. Extensive experiments on synthetic and real-world datasets show that our method outperforms the existing IV methods.

Introduction

Estimating the causal effect of a treatment (a.k.a., intervention, exposure, or action) on an outcome of interest, is a fundamental area of research (Pearl 2009; Pearl and Mackenzie 2018). Randomised controlled trials (RCTs) are considered the gold standard for causal effect estimation. However, RCTs are often difficult or impossible to conduct due to ethical issues and/or high costs (Imbens and Rubin 2015). Thus, it is important to estimate causal effects from observational data. For the case when there are no latent or unmeasured confounders[1] (i.e., the unconfoundedness assumption (Imbens and Rubin 2015) holds), many methods have been developed for causal effects estimation from data (Yao et al. 2021; Guo et al. 2020). In contrast, for the more realistic and challenging case when there are latent confounders in data, only a handful of data-driven methods have been developed.

Two types of approaches, the instrumental variable (IV) approach (Hernán and Robins 2006) and the proxy approach (Kuroki and Pearl 2014; Miao, Geng, and Tchetgen Tchetgen 2018) can be used for estimating causal effects from data with latent confounders. The proxy approach relies on the assumption that the measured variables are noisy measurements of the latent confounders. However, recent work (Rissanen and Marttinen 2021) has shown that it is difficult to recover a blocking set (i.e., a set that satisfies the back-door criterion (Pearl 2009)) for removing the confounding bias of the latent confounders even though there exists a rich set of proxy variables. Instead, the IV approach aims to avoid the spurious associations caused by latent confounders with the aid of a valid IV. In this work, we explore the direction of the IV approach.

The traditional IV approach requires a standard IV (denoted as S here) which satisfies the following three conditions (Hernán and Robins 2006): (i) S is correlated with the treatment T (a.k.a., *relevance condition*), (ii) S affects the outcome Y only through the treatment T (a.k.a., *exclusion restriction*), and (iii) there is no confounding bias between S and Y (a.k.a., *unconfounded instrument*). However, the last two conditions are too strict to be satisfied in practice. A conditional IV (CIV, see Definition 1 for its formal definition), which requires some measured variables as its conditioning set, has more relaxed conditions than a standard IV (Brito and Pearl 2002). Hence in this paper, we focus on data-driven methods based on CIVs.

It is challenging to determine a CIV and its conditioning set directly from observational data with latent confounders since a CIV and its conditioning set are not distinguishable by statistical tests due to latent confounders (Spirtes et al. 2000; Pearl 2009). The example in Fig. 1 illustrates the challenge. The causal directed acyclic graph (DAG) in Fig. 1 shows an underlying data generation mechanism, where T and Y are the treatment and outcome variables respectively, U_C is a latent confounder, and S is a valid CIV for which W_S is the conditioning set. In a dataset generated from this mechanism, it is impossible to distinguish the roles of S (as a CIV) and W_S (as the conditioning set of S).

Recently, the deep generative model based on variational autoencoder (VAE) (Kingma and Welling 2014; Sohn, Yan, and Lee 2015) has achieved many successes in areas such as causal representation learning (Schölkopf et al.

*These authors contributed equally.

[1]A confounder is a variable that causally affects both the treatment and the outcome.

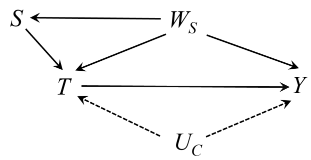

Figure 1: An exemplar causal DAG shows the indistinguish ability between a CIV and its conditioning set in the data with a latent confounder U_C. T and Y are treatment and outcome variables. S and W_S are a CIV and its conditioning set, and they are indistinguishable in the data since both are dependent on T and Y and conditional dependent on Y given T. Dashed edges indicate that U_C cannot be measured. As from the DAG, we know that in the dataset generated, both S and W_S are associated with T, given any other observed variable(s). Moreover, while S is associated with Y given any other observed variable(s), W_S is also associated with Y given any other observed variable(s) due to the unmeasured confounder U_C.

2021; Schölkopf 2022) and individual causal effect estimation (Hassanpour and Greiner 2019; Zhang, Liu, and Li 2021). In this work, we will leverage the strength of VAE to address the challenge of discovering CIVs and their conditioning sets from data with latent confounders.

For a treatment T and outcome Y, and a set of measured pretreatment variables $\mathbf{X}$, and given the assumption that there exists at least one CIV in $\mathbf{X}$, we propose the causal representation learning scheme (as shown in Fig. 2) to learn the representation $\mathbf{Z}_T$ of $\mathbf{X}$ and generate the latent representation $\mathbf{Z_C}$ conditioning on $\mathbf{X}$ respectively, and we prove that the obtained $\mathbf{Z}_T$ is a valid CIV conditioning on $\mathbf{Z_C}$. We then develop a VAE-based method, named CIV.VAE (Conditional IV approach based on VAE model) to conduct causal representation learning to obtain $\mathbf{Z}_T$ and $\mathbf{Z_C}$ for unbiased average causal effect estimation.

The contributions of this work are summarised as follows.

- We propose to tackle the problem of discovering a CIV and its conditioning set from data in the presence of latent confounders with causal representation learning. As far as we know, this is the first work using a data-driven approach and causal representation learning for identifying CIVs and their conditioning sets.
- We develop a novel VAE based method, CIV.VAE, to learn the representation of CIVs and generate the representation of their conditioning sets for causal effect estimation from data in the presence of latent confounders.
- Extensive experiments on a wide range of synthetic and real-world datasets show that the causal effects estimated using the CIVs and conditioning sets obtained by CIV.VAE have the smallest estimation error compared with the state-of-the-art causal effect estimators.

The rest of the paper is organised as follows. We firstly introduce background knowledge in Preliminary. Secondly, the details of our proposed CIV.VAE method is presented. Thirdly, we discuss the experimental setup, datasets and experimental results. Fourthly, we review the related works. Finally, we conclude our work.

Preliminary

In this section, we briefly introduce some important notations, definitions and assumptions used in the paper.

Notations & Definitions

We use uppercase and lowercase letters to represent variables and their values, respectively. Bold-faced uppercase and lowercase letters are used to indicate a set of variables and a value assignment of the set, respectively.

Let T be a binary treatment variable ($T = 1$ for treated and $T = 0$ for control) and Y be the outcome of interest. Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG with nodes $\mathbf{V} = \mathbf{X} \cup \mathbf{U} \cup \{T, Y\}$ and directed edges $\mathbf{E}$. $\mathbf{X}$ is the set of measured pretreatment variables, and $\mathbf{U} = \mathbf{U_C} \cup \mathbf{U}'$ is the set of unmeasured confounders, where $\mathbf{U_C}$ denotes the latent confounders between the pair T and Y, and $\mathbf{U}'$ represents the latent confounders between other variable pairs. In a causal DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, an edge (directed) represents the causal relationship between two nodes. For example, $X_i \rightarrow X_j$ in $\mathcal{G}$ indicates that X_i is a cause of X_j, and X_j is an effect of X_i. In a DAG $\mathcal{G}$, a path between X_i and X_j consists of a sequence of distinct nodes $\langle X_i, \dots, X_j \rangle$ with every pair of successive nodes being adjacent. More definitions regarding graphical causal modelling such as d-separation, collider, Markov property and faithfulness can be found in the supplement.

In this work, we would like to query the average causal effect (ACE) of treatment variable T on outcome variable Y, referred to as $ACE(T, Y)$, from a dataset containing a set of pretreatment variables $\mathbf{X}$, treatment T and outcome Y, and assuming that $\mathbf{X}$ contains at least one CIV and its conditioning set and there exists a set of latent confounders $\mathbf{U_C}$ affecting both T and Y and $\mathbf{U_C} \neq \emptyset$. There may exist other latent confounders (denoted as $\mathbf{U}'$) which affect pairs of variables other than (T, Y), and $\mathbf{U}'$ can be an $\emptyset$.

Conditional Instrumental Variable (CIV)

The IV approach is a powerful method for removing the confounding bias caused by the latent confounders affecting both treatment and outcome. As discussed in the Introduction, the last two conditions of a standard IV are too strict to be satisfied in real-world applications. In contrast, a conditional IV (CIV) , as defined below, has more relaxed conditions than a standard IV (Brito and Pearl 2002; Pearl 2009).

Definition 1 (Conditional IV (Pearl 2009))**.** *Given a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with $\mathbf{V} = \mathbf{X} \cup \mathbf{U} \cup \{T, Y\}$, a variable S is said to be a CIV w.r.t., $T \rightarrow Y$ if there exists a set of measured variables $\mathbf{W} \subseteq \mathbf{X} \setminus \{S\}$ and $\mathbf{W} \neq \emptyset$ such that (i) $S \not\perp\!\!\!\perp_d T | \mathbf{W}$, (ii) $S \perp\!\!\!\perp_d Y | \mathbf{W}$ in $\mathcal{G}_{\underline{T}}$, where $\mathcal{G}_{\underline{T}}$ is a manipulated DAG obtained by removing $T \rightarrow Y$ from $\mathcal{G}$, and (iii) $\forall W \in \mathbf{W}$, W is not a descendant of Y.*

In the above definition, $\not\perp\!\!\!\perp_d$ and $\perp\!\!\!\perp_d$ indicate d-connection and d-separation respectively in a DAG (Pearl 2009). Definition 1 can be easily generalised to a set of CIVs $\mathbf{S}$ with the corresponding conditioning set $\mathbf{W}$. Under the pretreatment assumption (i.e. all variables in $\mathbf{X}$ are pretreatment

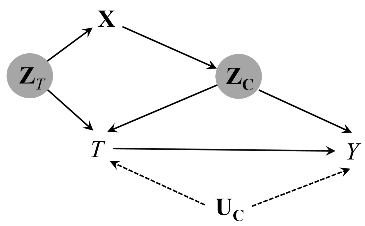

Figure 2: A causal graph representing the proposed causal representation learning scheme for discovering CIVs and their conditioning sets. T, Y, $\mathbf{X}$, and $\mathbf{U_C}$ are the treatment, the outcome, the set of measured pretreatment variables, and latent confounders between T and Y, respectively. The two grey circles denote the representation $\mathbf{Z}_T$ containing the CIV information in $\mathbf{X}$ and $\mathbf{Z_C}$ holding the information of the conditioning set of $\mathbf{Z_T}$ generated given $\mathbf{X}$.

variables), condition (iii) is always satisfied, so we only need to test conditions (i) and (ii). Given $X \in \mathbf{X}$, if a set $\mathbf{W} \subseteq \mathbf{X} \setminus \{X\}$ is found to block all paths between X and Y; and given $\mathbf{W}$, X and T are still dependent, then $\mathbf{W}$ instrumentalises X to be a CIV. Note that finding such a $\mathbf{W}$ from a causal DAG is proved to be NP-complete even with the pretreatment assumption (Van der Zander, Liśkiewicz, and Textor 2015).

In this work, we utilise the two stage least squares method (Angrist and Imbens 1995) as the ACE estimator. In a linear system, once we have a valid CIV S and its conditioning set $\mathbf{W}$, $ACE(T,Y)$ can be estimated by $\sigma_{s*y*\mathbf{w}}/\sigma_{s*t*\mathbf{w}}$, where $\sigma_{s*y*\mathbf{w}}$ and $\sigma_{s*t*\mathbf{W}}$ are the estimated causal effect of S on Y conditioning on $\mathbf{W}$ and the causal effect of S on T conditioning on $\mathbf{W}$, respectively.

The Proposed CIV.VAE Model

The Causal Representation Learning Scheme for CIV Discovery

In this work, we aim to simultaneously learn the latent representation $\mathbf{Z}_T$ of $\mathbf{X}$ and generate the latent representation $\mathbf{Z_C}$ given $\mathbf{X}$. Here $\mathbf{Z}_T$ contains the instrumental information that only influences T but not Y, and $\mathbf{Z_C}$ denotes the representation of the confounders that affect both T and Y. We assume that Fig. 2 is the underlying generative model (i.e., the underlying causal DAG) and it shows our proposed causal representation learning scheme. Specifically, a set of pretreatment variables $\mathbf{X}$ are generated from $\mathbf{Z}_T$ and $\mathbf{Z}_T$ captures the information of $\mathbf{S}$ in Fig. 1. The latent confounding representation $\mathbf{Z_C}$ is generated based on $\mathbf{X}$ and captures the information of $\mathbf{X} \setminus \mathbf{S}$ in Fig. 1.

If the two disjoint representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$ can be inferred from data with latent confounders, the following proposed theorem guarantees that $\mathbf{Z}_T$ is a valid CIV with $\mathbf{Z_C}$ being its conditioning set. Using $\mathbf{Z}_T$ and $\mathbf{Z_C}$, we can obtain unbiased causal effect estimation from data with latent confounders.

Theorem 1. *Given a causal DAG $\mathcal{G}$=$(\mathbf{X} \cup \mathbf{U} \cup \{T, Y\}, \mathbf{E})$ where $\mathbf{X}$ is the set of pretreatment variables. T and Y are the treatment and outcome respectively, and there exists $T \rightarrow Y$. $\mathbf{U} = \mathbf{U_C} \cup \mathbf{U}'$ is the set of latent confounders between T and Y (denoted as $\mathbf{U_C}$) and latent confounders between measured variables in $\mathbf{X}$ (denoted as $\mathbf{U}'$), and $\mathbf{E}$ is the set of edges between the variables. Suppose that there exists a set of CIVs $\mathbf{S} \subset \mathbf{X}$ with $|\mathbf{S}| \geq 1$ and its conditioning set $\mathbf{W} \subseteq \mathbf{X} \setminus \mathbf{S}$ with $|\mathbf{W}| \geq 1$. If the latent representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$ as shown in Fig. 2 can be learned from data, then $\mathbf{Z}_T$ is a CIV conditioning on $\mathbf{Z_C}$ for estimating the causal effect of T on Y.*

Proof. We prove that the two representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$ satisfy the three conditions of Definition 1 based on the causal DAG shown in Fig. 2. The conditions $|\mathbf{S}| \geq 1$ and $|\mathbf{W}| \geq 1$ are to ensure that there exists at least a pair of CIV and its conditioning set in $\mathbf{X}$. In $\mathcal{G}$, (1) $\mathbf{Z}_T$ is a set of parents of the treatment T, so $\mathbf{Z}_T \not\perp\!\!\!\perp_d T|\mathbf{Z_C}$ and the first condition of Definition 1 is satisfied; (2) the spurious association between $\mathbf{Z}_T$ and Y are caused by these paths, $\mathbf{Z}_T \rightarrow \mathbf{X} \rightarrow \mathbf{Z_C} \rightarrow Y$, $\mathbf{Z}_T \rightarrow T \leftarrow \mathbf{U_C} \rightarrow Y$, $\mathbf{Z}_T \rightarrow T \leftarrow \mathbf{Z_C} \rightarrow Y$ and $\mathbf{Z}_T \rightarrow \mathbf{X} \rightarrow \mathbf{Z_C} \rightarrow T \leftarrow \mathbf{U_C} \rightarrow Y$ in the manipulated DAG $\mathcal{G}_{\underline{T}}$. The first path is blocked by $\mathbf{Z_C}$ and the other three paths are blocked by $\emptyset$ since these paths contain a collider T, i.e., $\mathbf{Z}_T \perp\!\!\!\perp_d Y|\mathbf{Z_C}$ in $\mathcal{G}_{\underline{T}}$ and the second condition of Definition 1 holds; (3) $\mathbf{Z_C}$ is generated based on $\mathbf{X}$ and $\mathbf{X}$ contains only pretreatment variables, so $\mathbf{Z_C}$ is not a descendant of Y, i.e., the third condition of Definition 1 holds. Hence, the two representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$ satisfy Definition 1, i.e., $\mathbf{Z_C}$ instrumentalises $\mathbf{Z}_T$ such that $\mathbf{Z}_T$ is a valid CIV for estimating the causal effect of T on Y. □

Theorem 1 permits us to develop a data-driven method based on deep generative models (Kingma, Welling et al. 2019) to learn the representations of a CIV and its conditioning set directly from observational data. In the next subsection, we introduce our proposed data-driven method, CIV.VAE for learning the two representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$.

VAE-based Representation Learning of $\mathbf{Z}_T$ and $\mathbf{Z_C}$

Fig. 3 shows the CIV.VAE architecture we have developed for learning the latent representations of the CIV $\mathbf{Z}_T$ and its conditioning set $\mathbf{Z_C}$. CIV.VAE comprises an inference network and a generative network as shown in Fig. 3(a) and Fig. 3(b) respectively. CIV.VAE utilises the inference network and the generative network to approximate the posterior distributions of $p(\mathbf{Z}_T|\mathbf{X})$ and $p(\mathbf{Z_C}|\mathbf{X})$ for the two latent representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$ which indicate the latent CIV representation and the representation of its conditioning set, respectively.

In the inference network, two separate encoders $q(\mathbf{Z}_T|\mathbf{X})$ and $q(\mathbf{Z_C}|\mathbf{X})$ are utilised as variational posteriors over the latent representations. In the generative model, these latent representations are utilised by a single decoder $p(\mathbf{X}|\mathbf{Z}_T, \mathbf{Z_C})$ for the reconstruction of $\mathbf{X}$. Based on the standard VAE framework in literature (Kingma and Welling 2014; Kingma, Welling et al. 2019), the prior distribution of

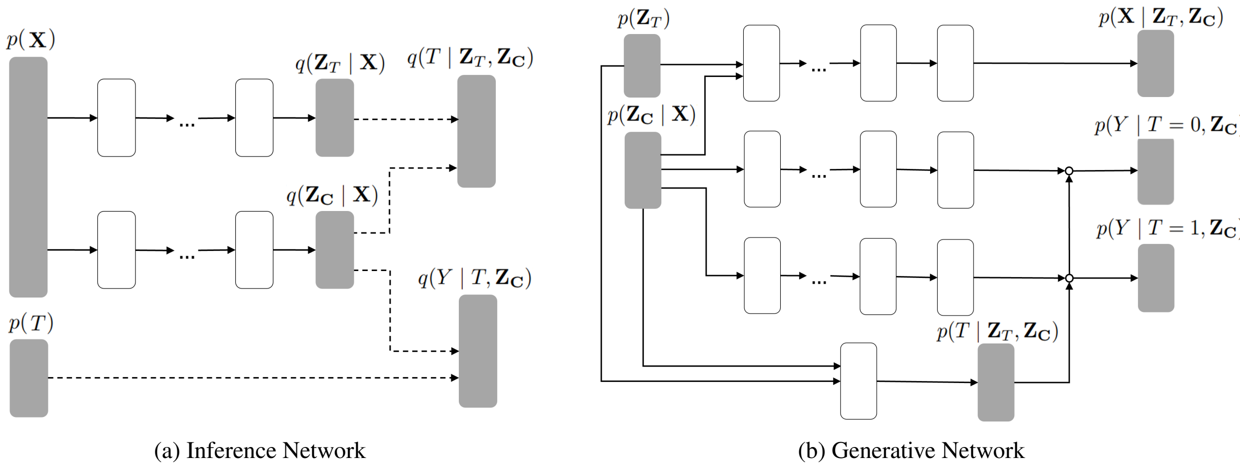

Figure 3: The proposed CIV.VAE architecture which consists of the inference network and the generative network for learning the latent representations of CIV $\mathbf{Z}_T$ and its conditioning set $\mathbf{Z_C}$. A grey box denotes the drawing of samples from the respective distribution, a white box indicates the parameterised deterministic neural network transitions, and a circle indicates switching paths according to the value of T. In the inference network, the dashed arrows indicate the two auxiliary predictors $q(T|\mathbf{Z}_T, \mathbf{Z_C})$ and $q(Y|T, \mathbf{Z_C})$.

$p(\mathbf{Z}_T)$ is sampled from a Gaussian distribution as follows:

$$p(\mathbf{Z}_T) = \prod_{i=1}^{D_{\mathbf{Z}_T}} \mathcal{N}(Z_{T_i}|0, 1). \tag{1}$$

where $D_{\mathbf{Z}_T}$ is the dimension of $\mathbf{Z}_T$.

Specifically, in the inference model, the variational approximations of the posterior distributions are as follows:

$$\begin{aligned} q(\mathbf{Z}_T|\mathbf{X}) &= \prod_{i=1}^{D_{\mathbf{Z}_T}} \mathcal{N}(\mu = \hat{\mu}_{\mathbf{Z}_{T_i}}, \sigma^2 = \hat{\sigma}^2_{\mathbf{Z}_{T_i}}); \\ q(\mathbf{Z_C}|\mathbf{X}) &= \prod_{i=1}^{D_{\mathbf{Z_C}}} \mathcal{N}(\mu = \hat{\mu}_{\mathbf{Z_C}_i}, \sigma^2 = \hat{\sigma}^2_{\mathbf{Z_C}_i}) \end{aligned} \tag{2}$$

where $\hat{\mu}_{\mathbf{Z}_T}, \hat{\mu}_{\mathbf{Z_C}}$ and $\hat{\sigma}^2_{\mathbf{Z}_T}, \hat{\sigma}^2_{\mathbf{Z_C}}$ are the means and variances of the Gaussian distributions parameterised by neural networks. $D_{\mathbf{Z_C}}$ is the dimension of $\mathbf{Z_C}$.

In the generative model, according to the conditional variational autoencoder (CVAE) network (Sohn, Yan, and Lee 2015), we use Monte Carlo (MC) sampling to obtain $\mathbf{Z_C}$ conditioning on the set of pretreatment variables $\mathbf{X}$:

$$\mathbf{Z_C} \backsim p(\mathbf{Z_C}|\mathbf{X}) \tag{3}$$

The generative models for T and $\mathbf{X}$ are described as:

$$\begin{aligned} p(T|\mathbf{Z}_T, \mathbf{Z_C}) &= Bern(\sigma(g_1(\mathbf{Z}_T, \mathbf{Z_C}))); \\ p(\mathbf{X}|\mathbf{Z}_T, \mathbf{Z_C}) &= \prod_{i=1}^{D_{\mathbf{X}}} p(X_i|\mathbf{Z}_T, \mathbf{Z_C}), \end{aligned} \tag{4}$$

where $g_1(\cdot)$ is the function parameterised by neural networks and $\sigma(\cdot)$ is the logistic function.

For continuous Y, in the generative model, it is modelled as a Gaussian distribution with its mean and variance parameterised it by the mutually exclusive neural networks that defines $p(Y|T = 0, \mathbf{Z_C})$ and $p(Y|T = 1, \mathbf{Z_C})$, respectively. Specifically, the model is defined as:

$$\begin{aligned} p(Y|T, \mathbf{Z_C}) &= \mathcal{N}(\mu = \hat{\mu}_Y, \sigma^2 = \hat{\sigma}^2_Y), \\ \hat{\mu}_Y &= T \cdot g_2(\mathbf{Z_C}) + (1 - T) \cdot g_3(\mathbf{Z_C}), \\ \hat{\sigma}^2_Y &= T \cdot g_4(\mathbf{Z_C}) + (1 - T) \cdot g_5(\mathbf{Z_C}). \end{aligned} \tag{5}$$

where $g_2(\cdot), g_3(\cdot), g_4(\cdot)$ and $g_5(\cdot)$ are the functions parameterised by neural networks. For binary Y, we model it with a Bernoulli distribution. The specific model is:

$$p(Y|T, \mathbf{Z_C}) = Bern(\sigma(g_6(T, \mathbf{Z_C}))). \tag{6}$$

where $g_6(\cdot)$ is a function parameterised by neural networks. For inference, the parameters can be optimised by maximising the evidence lower bound (ELBO):

$$\begin{aligned} \mathcal{M} = \; & \mathbb{E}_q[\log p(\mathbf{X}|\mathbf{Z}_T, \mathbf{Z_C})] - D_{KL}[q(\mathbf{Z}_T|\mathbf{X})||p(\mathbf{Z}_T)] \\ & - D_{KL}[q(\mathbf{Z_C}|\mathbf{X})||p(\mathbf{Z_C}|\mathbf{X})]. \end{aligned} \tag{7}$$

Note that the decoder (generative network) $p(\mathbf{Z_C}|\mathbf{X})$ conditioning on $\mathbf{X}$ is used to encourage as much information as possible from $\mathbf{X}$ is captured in the CIV.VAE method. To improve the learning of the latent representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$) and enable that the treatment T can be predicted from the two representations and the outcome Y can be predicted from $\mathbf{Z_C}$ and T, we add two auxiliary predictors to the variational ELBO in Eq.(7) as designed by the works (Louizos et al. 2017; Zhang, Liu, and Li 2021). Consequently, the proposed objective function of CIV.VAE is expressed as:

$$\begin{aligned} \mathcal{L}_{CIV.VAE} = \; & - \mathcal{M} + \alpha \mathbb{E}_q[\log q(T|\mathbf{Z}_T, \mathbf{Z_C})] \\ & + \beta \mathbb{E}_q[\log q(Y|T, \mathbf{Z_C})], \end{aligned} \tag{8}$$

where α and β are the weights for the auxiliary predictors.

For calculating $ACE(W, Y)$, we draw $\mathbf{Z}_T$ and $\mathbf{Z_C}$ from the trained CIV.VAE method, and utilise both learned latent representations $\mathbf{Z}_T$ and $\mathbf{Z_C}$ in the Instrumental Variable (CIV) method (Angrist and Imbens 1995) as described in the "Conditional Instrumental Variable (CIV)" subsection, where s and w are replaced by $\mathbf{Z}_T$ and $\mathbf{Z_C}$ respectively.

The main advantage of CIV.VAE is that it simultaneously learns the latent CIV representation $\mathbf{Z}_T$ and the latent representation of the conditioning set $\mathbf{Z_C}$ for $\mathbf{Z}_T$ without specifying a CIV and its conditioning set by domain knowledge, and it provides a practical solution to the challenge (described in the Introduction) of distinguishing a CIV and its conditioning set from data with latent confounders. Hence, CIV.VAE is expected to have wider applications. CIV.VAE only relies on two practical assumptions, the pretreatment variable assumption and the existence of at least one CIV and its conditioning set in $\mathbf{X}$.

The main difference between CIV.VAE and the two other VAE-based causal effect estimators, CEVAE (Louizos et al. 2017) and TEDVAE (Zhang, Liu, and Li 2021) is that CIV.VAE builds on conditioning VAE for learning $\mathbf{Z}_T$ as the CIV and $\mathbf{Z_C}$ as its conditioning set that blocks the confounding bias between $\mathbf{Z}_T$ and Y, whereas CEVAE and TEDVAE are to recover the set that blocks the confounding bias between T and Y. Moreover, CIV.VAE method belongs to the IV approach, while CEVAE and TEDVAE methods are confounding adjustment methods.

Experiments

In this section, we evaluate the performance of CIV.VAE for the task of estimating the average causal effect of T on Y. The experiments are divided into two parts: evaluation with simulated data and evaluation with real-world data. For the first part, we use the causal DAG in Fig. 1 in the supplement to generate synthetic datasets with latent confounders. For the second part, we use three real-world datasets, Schoolingreturns (Card 1993), 401k (Wooldridge 2010) and Sachs (Sachs, Perez et al. 2005), which have reference causal effect values available in literature. The three datasets are widely used in the evaluation of IV-based methods (Abadie 2003; Wooldridge 2010; Silva and Shimizu 2017). Note that Schoolingreturns and 401k each have a nominated CIV for the causal effect estimation, but the corresponding conditioning set is unknown, and there is not a nominated IV in Sachs. Our CIV.VAE method do not use a known CIV, instead we learn the representations of CIVs and their conditioning sets.

Experiment Setup

We compare CIV.VAE with three types of causal effect estimators: (1) IV-based estimators with a given IV, (2) IV-based estimators without a given IV and (3) VAE-based causal effect estimators. Five of the IV-based estimators, TSLS (two-stage least squares) regression (Angrist and Imbens 1995), FIVR (the causal random forest method for IV regression) (Athey, Tibshirani, and Wager 2019), the deepIV (a deep learning based IV estimator) (Hartford, Lewis et al. 2017), OrthoIV (orthogonal machine learning based IV estimator) (Syrgkanis, Lei et al. 2019) and DMLIV (double machine learning based IV estimator) (Chernozhukov et al. 2018), each of which needs a given IV; whereas the other IV-based estimators, IV.Tetrad (Silva and Shimizu 2017) and sisVIVE (some invalid some valid IV estimator) (Kang et al. 2016) do not need a given IV. The two VAE-based causal effect estimators are CEVAE (causal effect variational autoencoder) (Louizos et al. 2017) and TEDVAE (treatment effect by disentangled variational autoencoder) (Zhang, Liu, and Li 2021). We choose the two VAE-based estimators as our baseline because CIV.VAE also builds on the VAE model.

Evaluation Metrics. For synthetic datasets with the true causal effect $ACE(T, Y)$, the absolute error: $\varepsilon_{ACE} = \left|\hat{ACE}(T, Y) - ACE(T, Y)\right|$ is used to evaluate the performance of all estimators. For multiple replications, we report the average results with STD (standard deviation). For the three real-world datasets, all estimators are evaluated against the reference causal effect values in the literature.

Implementation Details. We use *Python* and the libraries including *pytorch* (Paszke, Gross et al. 2019), *pyro* (Bingham, Chen et al. 2019) and *scikit-learn* (Pedregosa et al. 2011) to implement our CIV.VAE method. We provide the details of our CIV.VAE implementation and the parameters setting in the supplement. TSLS is implemented by using the functions *glm* and *ivglm* in the *R* packages *stats* and *ivtools* (Sjolander and Martinussen 2019) respectively. FIVR is coded by employing the function *instrumental_forest* in the *R* package *grf* (Athey, Tibshirani, and Wager 2019). The program of DeepIV is retrieved from the authors' GitHub [2]. The implementations of OrthoIV and DMLIV are from the *Python* package *encoml*. IV.Tetrad is obtained from the authors' site [3]. The implementation of CEVAE is obtained from the *Python* library *pyro* (Bingham, Chen et al. 2019) and the implementation of TEDVAE is downloaded from the authors' GitHub [4].

Evaluation with Simulated Data

We use the causal DAG provided in Fig. 1 in the supplement to generate a set of synthetic datasets with a range of sample sizes: 2k, 4k, 6k, 8k, 10k and 20k. The set of $\mathbf{X}$ is $\{S, X_1, X_2, X_3, X_4, X_5\}$ and the set $\mathbf{U}$ (latent confounders) consists of $\{U, U_1, U_2, U_3, U_4\}$ where $T \leftarrow U \rightarrow Y$ and $\mathbf{U}' = \{U_1, U_2, U_3, U_4\}$. The true $ACE(T, Y)$ of all synthetic datasets is 2. Due to page limitation, more details of the data generation process are provided in the supplement.

To avoid the bias brought by data generation, 30 synthetic datasets for each sample size are generated in our experiments. We utilise the CIV S in the underlying causal DAG as a standard IV for TSLS. Moreover, the CIV S and the set $\mathbf{X} \setminus \{S\}$ in the underlying causal DAG are the true CIV and the corresponding conditioning set, respectively, and they both are taken as input for the four estimators, FIVR,

[2]https://github.com/jhartford/DeepIV

[3]http://www.homepages.ucl.ac.uk/~ucgtrbd/code/iv_discovery

[4]https://github.com/WeijiaZhang24/TEDVAE

Methods		Samples					
		2k	4k	6k	8k	10k	20k
Known IV	TSLS	11.12±1.74	11.12±1.38	10.87±0.83	11.08±0.94	11.34±0.82	11.06±0.55
	FIVR	1.58±0.97	1.10±0.61	0.64±0.50	0.62±0.37	0.63±0.42	0.33±0.21
	DeepIV	1.64±0.19	1.47±0.21	1.53±0.21	1.53±0.23	1.43±0.32	1.33±0.22
	OrthIV	3.57±2.90	1.91±1.19	1.29±1.49	1.51±1.25	1.29±0.93	0.71±0.52
	DMLIV	3.53±2.63	2.11±1.70	1.19±1.33	1.49±1.32	1.12±0.85	0.71±0.58
Unknown IV	sisVIVE	1.37±0.79	1.61±0.97	1.70±1.01	1.39±0.61	1.56±0.81	2.08±0.89
	IV.Tetrad	2.89±3.77	2.15±3.45	2.73±3.92	2.86±3.90	2.12±3.60	2.90±3.98
VAE-based	CEVAE	**1.34±0.23**	1.14±0.25	1.23±0.29	1.17±0.36	1.17±0.39	1.13±0.58
	TEDVAE	1.68±0.27	1.65±0.18	1.70±0.14	1.68±0.11	1.70±0.10	1.70±0.09
CIV.VAE		1.94±1.45	**0.95±0.42**	**0.42±0.34**	**0.33±0.20**	**0.26±0.19**	**0.22±0.17**

Table 1: The table summarises the estimated errors ε_{ACE} (Mean±STD) over 30 synthetic datasets in each sample size. The lowest estimated errors are marked in boldface. Note that CIV.VAE relies on the least domain knowledge among all estimators and obtain the smallest ε_{ACE} among all methods compared.

DeepIV, OrthIV and DMLIV. The estimation errors of all estimators on all synthetic datasets are reported in Table 1.

Results. From Table 1, we see that CIV.VAE obtains the smallest ε_{ACE} across almost all datasets compared with the other estimators. Note that CIV.VAE relies on the least domain knowledge, i.e. does not require a specific IV (whereas TSLS, FIVR, DeepIV, OrthIV and DMLIV do) or the conditioning set (whereas IV.Tetrad does), or a rich set of IVs (whereas sisVIVE does) or a rich set of proxy variables (whereas CEVAE does) or the unconfoundedness assumption (whereas TEDVAE does).

There are five other observations from Table 1: (1) the baseline IV estimator, TSLS, has the largest estimation errors because the confounding bias between S and Y caused by confounders and latent confounders is not blocked at all even though it uses the CIV S as its known IV. (2) FIVR obtains the best performance in the first type of IV-based methods, i.e., TSLS, DeepIV, OrthIV and DMLIV, but has larger estimation errors than CIV.VAE. (3) two IV-based estimators without needing a given IV, sisVIVE and IV.Tetrad obtain constant estimation errors with little variation across all synthetic datasets and have larger estimation errors than CIV.VAE. (4) the two VAE-based causal effect estimators in the third type of comparison methods, CEVAE and TEDVAE obtain smaller estimation errors than the second type of methods, but both have larger estimation errors than CIV.VAE. (5) As the sample size increases, CIV.VAE consistently gets the lowest bias compared with all causal effect estimators except for the 2k sample size. This indicates that CIV.VAE requires relatively large sample size to learn the two representations $\mathbf{Z}_T$ and $\mathbf{Z}_\mathbf{C}$ such that their distributions are close to the true distributions.

Therefore, the experimental results on synthetic datasets show that CIV.VAE has the capability to learn high quality CIV and conditioning set representations for causal effect estimation from data with latent confounders.

Experiments on Real-world Datasets

In this section, we conduct experiments on three benchmark real-world datasets, Schoolingreturns (Card 1993), 401(k) (Verbeek 2008) and Sachs (Sachs, Perez et al. 2005) for which the empirical causal effects available and widely accepted. Note that the first two datasets each have a known IV based on domain knowledge, but Sachs does not have a known IV. The detailed descriptions of the three real-world datasets are introduced in the supplement.

Schoolingreturns. This dataset consists of 3,010 records and 19 variables. The treatment variable is the education level of a person. The outcome variable is raw wages in 1976 (in cents per hour). The goal of collecting this dataset is to study the causal effect of the education level on wages. In the work (Card 1993), *nearcollege* (geographical proximity to a college) is nominated as the known IV, and the estimated $ACE(T,Y) = 0.1329$ with 95% confidence interval (0.0484, 0.2175) from the works (Verbeek 2008) as the reference causal effect.

401(k). The dataset contains 9,275 individuals and 11 variables. The dataset is from the survey of income and program participation (SIPP) (Verbeek 2008). The treatment is *p401k* (a binary indicated variable of participation in 401 (k)), and the outcome is *pira* (a binary indicated variable, $pira$ = 1 denotes participation in IRA). *e401k* (a binary indicated variable of eligibility for 401 (k)) is used as an IV, $ACE(T,Y) = 0.0712$ with 95% confidence interval (0.047, 0.095) (Verbeek 2008) as the reference causal effect.

Sachs. The dataset contains 853 samples and 11 variables (Sachs, Perez et al. 2005). The treatment is Erk (the manipulation of concentration levels of a molecule). The outcome is the concentration of Akt. Note that there is not a nominated IV. In this work, we take the reported $ACE(T,Y) = 1.4301$ with 95% confidence interval (0.05, 3.23) in the work (Silva and Shimizu 2017) (i.e., IV.Tetrad's estimated causal effect) as the reference causal effect.

Results. All results on the three datasets are reported in Table 2. From the results in Table 2, we see that (1) the causal effects estimated by CIV.VAE are within the 95% confidence interval of their empirical results, and on Sachs, the estimated causal effect by CIV.VAE is the closest to IV.Tetrad's result; (2) the estimated causal effects by

Estimators	TSLS	FIVR	DeepIV	OrthIV	DMLIV	sisVIVE	IV.Tetrad	CEVAE	TEDVAE	CIV.VAE
Schoolingreturns	0.5042	1.1513	-0.0444	1.3189	1.2806	0.0254	**0.0643**	**0.0956**	-0.1082	**0.1034**
401(k)	0.1500	**0.0746**	-	0.1502	0.1503	1.5172	1.2484	0.0384	0.0283	**0.0752**
Sachs	-	-	-	-	-	**0.4356**	**1.4301**	**0.2542**	**0.2553**	**1.5133**

Table 2: The estimated causal effects of all estimators on the three real-world datasets. We highlight the estimated causal effects within the 95% confidence interval on Schoolingreturns and 401(k). '-' is used for the corresponding IV-based estimator on Sachs because there is not a known IV. Note that DeepIV cannot work on 401(k) and is also marked as '-'.

IV.Tetrad and CEVAE on Schoolingreturns, and the estimation by FIVR on 401(k) are in the 95% empirical interval, but not on all three datasets. The other compared estimators do not perform well on both datasets with empirical intervals. (3) There is not a known IV on Sachs, so IV-based estimators requiring a given IV do not work on this dataset. Note that the estimated average causal effects by CIV.VAE, sisVIVE, CEVAE and TEDVAE are in the empirical interval and they work well on Sachs.

The experiments on the three real-world datasets further confirm that CIV.VAE is able to infer the conditional IV representation $\mathbf{Z}_T$ and the corresponding conditioning representation $\mathbf{Z}_\mathbf{C}$ from data in the presence of latent confounders for unbiased average causal effect estimation.

In a word, CIV.VAE, without knowing a CIV and the corresponding conditioning set, performs better than the state-of-the-art IV-based and two VAE-based estimators on the three real-world datasets.

Limitations. CIV.VAE relies on the assumptions of the pretreatment variables and the existence of at least one CIV, and it also relies on that VAE correctly identifies the latent variables. However, work by (Khemakhem et al. 2020) shows that identifying latent variables is not guaranteed by VAE. This means that, when some of the assumptions are not satisfied or VAE identifiability, CIV.VAE may provide an unreliable conclusion. To avoid the potential negative impact, it is better to choose other causal effect estimators to cross check or conducting a sensitive analysis (Imbens and Rubin 2015; Hartford et al. 2021). Furthermore, the identifiable VAE framework (a.k.a. iVAE) in (Khemakhem et al. 2020) provides some ideas for improving the identifiability of the VAE-based model, and it can be used to improve the reliability of CIV.VAE.

Related Work

We review the work closely related to our proposed method, including IV-based methods requiring a given IV and data-driven IV estimators without a known IV.

IV-based Estimators Requiring a Given IV. Several IV-based causal effect estimators have been proposed for average causal effect estimation when there is a known IV, such as causal random forest based IV regression (FIVR) (Athey, Tibshirani, and Wager 2019), generalised method of moments based IV estimator (GMM) (Bennett, Kallus, and Schnabel 2019), deep ensemble method based IV approach (DeepIV) (Hartford, Lewis et al. 2017) and kernel IV regression (KIV) (Singh, Sahani, and Gretton 2019). Different from these IV-based estimators, CIV.VAE does not require a given IV and the conditioning set by domain knowledge.

Data-driven IV-based Estimators without a Known IV. In most real-world applications, there is not a known IV. A few data-driven IV estimators have been developed for discovering a valid IV (Yuan, Wu et al. 2022) or a synthesising IV (Burgess and Thompson 2013) or eliminating the influence of invalid IVs by using statistical strategies (Kang et al. 2016; Guo, Kang et al. 2018; Hartford et al. 2021). For instance, the tetrad constraint is utilised by IV.Tetrad (Silva and Shimizu 2017) to validate the validity of a pair of CIVs for estimating causal effects from data in presence of latent confounders. However, it requires that there exists at least a pair of CIVs and assumes that the set of all the remained variables is the conditioning set. Kuang et al., (Kuang, Sala et al. 2020) proposed the Ivy method to synthesise an IV by combining a set of IV candidates for determining all invalid IVs or dependencies. The sisVIVE method (Kang et al. 2016) was developed to estimate causal effects when the majority assumption holds (i.e., at least a half of the covariates are valid IVs). Under the majority assumption, the ModeIV estimator (Hartford et al. 2021) is developed by employing a deep learning based IV estimator (Hartford, Lewis et al. 2017). Unlike this type of data-driven IV-based estimators, CIV.VAE takes the advantages of deep generative model to learn the latent IV representation and the conditioning set representation from data.

Conclusion

Latent confounders are a crucial challenge for causal inference in practice. IV-based approach provides an effective way to circumvent the latent confounding problem. However, for data-driven causal inference, standard IV is not feasible due to the strict conditions. CIVs shed light on data-driven IV-based causal inference, but in many cases it is impossible to distinguish a CIV and its conditioning set from data using traditional methods. In this paper, by leveraging the VAE model, we have designed the CIV.VAE method to learn the representations of a CIV and its conditioning set from data with latent confounders. We have conducted extensive experiments on synthetic and three real-world datasets, and the experimental results demonstrate the capability and the validity of our proposed CIV.VAE model against the state-of-the-art estimators in causal effect estimation with data containing latent confounders.

Acknowledgments

We wish to acknowledge the support from the Australian Research Council (under grant DP200101210).

References

Abadie, A. 2003. Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2): 231–263.

Angrist, J. D.; and Imbens, G. W. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430): 431–442.

Athey, S.; Tibshirani, J.; and Wager, S. 2019. Generalized random forests. *The Annals of Statistics*, 47(2): 1148–1178.

Bennett, A.; Kallus, N.; and Schnabel, T. 2019. Deep generalized method of moments for instrumental variable analysis. In *International Conference on Neural Information Processing Systems*, 3564–3574.

Bingham, E.; Chen, J. P.; et al. 2019. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1): 973–978.

Brito, C.; and Pearl, J. 2002. Generalized instrumental variables. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, 85–93.

Burgess, S.; and Thompson, S. G. 2013. Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology*, 42(4): 1134–1144.

Card, D. 1993. Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *Econometrica*, 69(9): 1127–1160.

Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; and Robins, J. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1): C1–C68.

Guo, R.; Cheng, L.; Li, J.; Hahn, P. R.; and Liu, H. 2020. A survey of learning causality with data: Problems and methods. *ACM Computing Surveys (CSUR)*, 53(4): 1–37.

Guo, Z.; Kang, H.; et al. 2018. Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4): 793–815.

Hartford, J.; Lewis, G.; et al. 2017. Deep IV: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, 1414–1423.

Hartford, J. S.; Veitch, V.; Sridhar, D.; and Leyton-Brown, K. 2021. Valid causal inference with (some) invalid instruments. In *International Conference on Machine Learning*, 4096–4106. PMLR.

Hassanpour, N.; and Greiner, R. 2019. Learning disentangled representations for counterfactual regression. In *International Conference on Learning Representations*, 1–11.

Hernán, M. A.; and Robins, J. M. 2006. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 360–372.

Imbens, G. W.; and Rubin, D. B. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kang, H.; Zhang, A.; Cai, T. T.; and Small, D. S. 2016. Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American statistical Association*, 111(513): 132–144.

Khemakhem, I.; Kingma, D.; Monti, R.; and Hyvarinen, A. 2020. Variational autoencoders and nonlinear ica: A unifying framework. In *International Conference on Artificial Intelligence and Statistics*, 2207–2217. PMLR.

Kingma, D. P.; and Welling, M. 2014. Auto-encoding variational bayes. In *International Conference on Learning Representations*.

Kingma, D. P.; Welling, M.; et al. 2019. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4): 307–392.

Kuang, Z.; Sala, F.; et al. 2020. Ivy: Instrumental variable synthesis for causal inference. In *International Conference on Artificial Intelligence and Statistics*, 398–410.

Kuroki, M.; and Pearl, J. 2014. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2): 423–437.

Louizos, C.; Shalit, U.; Mooij, J. M.; Sontag, D.; Zemel, R.; and Welling, M. 2017. Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems*, 6446–6456.

Miao, W.; Geng, Z.; and Tchetgen Tchetgen, E. J. 2018. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4): 987–993.

Paszke, A.; Gross, S.; et al. 2019. PyTorch: an imperative style, high-performance deep learning library. In *International Conference on Neural Information Processing Systems*, 8026–8037.

Pearl, J. 2009. *Causality*. Cambridge university press.

Pearl, J.; and Mackenzie, D. 2018. *The Book of Why: the New Science of Cause and Effect*. Basic Books.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12: 2825–2830.

Rissanen, S.; and Marttinen, P. 2021. A critical look at the consistency of causal estimation with deep latent variable models. In *International Conference on Neural Information Processing Systems*, 4207–4217.

Sachs, K.; Perez, O.; et al. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721): 523–529.

Schölkopf, B. 2022. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, 765–804.

Schölkopf, B.; Locatello, F.; Bauer, S.; Ke, N. R.; Kalchbrenner, N.; Goyal, A.; and Bengio, Y. 2021. Toward causal representation learning. *Proceedings of the IEEE*, 109(5): 612–634.

Silva, R.; and Shimizu, S. 2017. Learning instrumental variables with structural and non-gaussianity assumptions. *Journal of Machine Learning Research*, 18(120): 1–49.

Singh, R.; Sahani, M.; and Gretton, A. 2019. Kernel instrumental variable regression. In *International Conference on Neural Information Processing Systems*, 4593–4605.

Sjolander, A.; and Martinussen, T. 2019. Instrumental variable estimation with the R package ivtools. *Epidemiologic Methods*, 8(1).

Sohn, K.; Yan, X.; and Lee, H. 2015. Learning structured output representation using deep conditional generative models. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*, 3483–3491.

Spirtes, P.; Glymour, C. N.; Scheines, R.; and Heckerman, D. 2000. *Causation, prediction, and search*. MIT press.

Syrgkanis, V.; Lei, V.; et al. 2019. Machine learning estimation of heterogeneous treatment effects with instruments. In *International Conference on Neural Information Processing Systems*, 15193–15202.

Van der Zander, B.; Liśkiewicz, M.; and Textor, J. 2015. Efficiently finding conditional instruments for causal inference. 3243–3249.

Verbeek, M. 2008. *A guide to modern econometrics*. John Wiley & Sons.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Yao, L.; Chu, Z.; Li, S.; Li, Y.; Gao, J.; and Zhang, A. 2021. A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(5): 1–46.

Yuan, J.; Wu, A.; et al. 2022. Auto IV: Counterfactual Prediction via Automatic Instrumental Variable Decomposition. *ACM Transactions on Knowledge Discovery from Data*, 16(4): 1–20.

Zhang, W.; Liu, L.; and Li, J. 2021. Treatment Effect Estimation with Disentangled Latent Factors. In *The AAAI Conference on Artificial Intelligence*, 10923–10930.