Causal mediation analysis with multiple mediators and censored outcomes by GAN approach

Hanwen Ning Zhanfeng Li

Zhongnan University of Economics and Law

ninghanwen@163.com

lizf63@aliyun.com

Seminar on Frontier Methods in Econometrics
August 19, 2024

The presentation of our work

1. Background and Introduction
2. Existing benchmarks and Counterfactual framework
3. CGAN and Our motivations
4. Generative Adversarial Censored Mediation Networks (GACMN)
5. Further discussions
6. Convergence results
7. Experiments
8. Conclusion

Background and Introduction

What is CMA?

- Causal mediation analysis (CMA) is a powerful tool for investigating causal mechanism of treatment. CMA aims to understand the causal pathway or process by which the treatment variable or factor influences the outcome variable, and holds significant importance in social and medical sciences. It has numerous applications in various research fields, such as neuroscience, psychology, epidemiology and economics.

Background and Introduction

CMA with multiple mediators and censored output

- The dependent variable is typically assumed to follow an underlying continuous distribution, with observed values either censored at a specific threshold. Utilizing standard regression models without accounting for the censored nature of the data can yield inconsistent parameter estimates and a skewed understanding of statistical relationships.
- As an important subtopic, CMA with censored outcomes has garnered particular attention due to the prevalence of censored observations in diverse domains, including the examination of children's academic achievement in educational research, the analysis of wages or job search durations in labor economics, and the investigation of purchasing behaviors and spending patterns in consumer expenditure analysis.

Background and Introduction

The challenges in the context of large data

- It is worth noting that available mediation models usually impose restrictive parametric settings (e.g., linear) and strong distribution assumptions (e.g., homogenous and normal) to facilitate estimation and hypothesis testing. However, there are no clear theoretical or intuitive justifications for these convenient assumptions in practice.
- As shown in economics and financial econometrics, the data are sampled from individuals, essentially determined by personal characteristics, and exhibit high levels of nonlinearity, complexity and heterogeneity. It is particularly true in the context of large datasets.
- How to develop efficient econometric models using machine learning, especially generative learning approach (VAE, GAN, diffusion models).

Existing benchmarks and Counterfactual framework

Let T, M and $Y^{\star}$ be the treatment, mediator and latent true outcome, respectively.

$$Y = \max(C_t, Y^{\star}) = \begin{cases} Y^{\star}, & \text{if } Y^{\star} > C_t, \\ C_t, & \text{if } Y^{\star} \leq C_t, \end{cases} \tag{1}$$

where C_t is threshold parameter, we are interested in the following mediation system

$$\begin{cases} M &= \beta_1 + aT + \varepsilon_1, \\ Y^{\star} &= \beta_2 + bM + c'T + \varepsilon_2, \end{cases} \tag{2}$$

where $\varepsilon_1 \sim N(0, \sigma_1^2)$ and $\varepsilon_2 \sim N(0, \sigma_2^2)$ are normally distributed error terms. c' represents the relationship between T and Y after controlling the effect of M, which is also called the direct effect of T on Y.

Existing benchmark methods

1. **Nonlinear least square methods** A. J. Fairchild, et.al., R2 effect-size measures for mediation analysis, Behavior research methods 41 (2) (2009) 486-498.
2. **Heckman Two-step method** A. C. Cameron, P. K. Trivedi, Microeconometrics: methods and applications, Cambridge university press, 2005.
3. **Baysian Tobit estimation** L. Wang, Z. Zhang, Estimating and testing mediation effects with censored data, Structural Equation Modeling 18 (1) (2011) 18-34.
4. **Maximum likelihood estimation** P. Carneiro, et.al., Estimating marginal returns to education, American Economic Review 101 (6) (2011) 2754-2781.

Existing benchmark methods

Basically, we have the conditional density of Y can be written as

$$f(Y|M,T) = f^{\star}(Y|M,T)^{d} F^{\star}(C_t|M,T)^{1-d}. \tag{3}$$

The indicator variable d takes a value of 1 when $Y^{\star} > C_t$ and 0 otherwise. The conditional probability density function $f^{\star}(Y|M,T)$ represents the probability density function for $Y^{\star} > C_t$, and $F^{\star}(C_t|M,T)$ denotes the probability that the latent outcome $Y^{\star} \leq C_t$. To bring an effective estimation, strict assumptions on model structure have be to imposed.

Counterfactual framework

To identify the direct and indirect effects through each causal pathway, two important unconfoundedness assumptions need to be stated as follows:
(I) conditional independence of the treatment:
$\{Y^{\star}(t,\boldsymbol{m}), \boldsymbol{M}(t^{*})\} \perp T \mid \boldsymbol{X}$,
(II) conditional independence of the mediator:
$Y^{\star}(t,\boldsymbol{m}) \perp \boldsymbol{M}(t^{*}) \mid T, \boldsymbol{X}$.
Here, the notation $A \perp B \mid C$ indicates the independence of A and B given C.

Counterfactual framework

Four potential outcomes: $Y_i^\star(t_0, \boldsymbol{M}_i(t_0))$, $Y_i^\star(t_1, \boldsymbol{M}_i(t_1))$, $Y_i^\star(t_0, \boldsymbol{M}_i(t_1))$, and $Y_i^\star(t_1, \boldsymbol{M}_i(t_0))$. For the ith individual, the direct effect, indirect effect, and indirect effect through the pth mediator M^p:

$$\begin{aligned}
\triangle_{i,T\to Y^\star} &= Y_i^\star(t_1, \boldsymbol{M}_i(t_0)) - Y_i^\star(t_0, \boldsymbol{M}_i(t_0)), && (4)\\
\triangle_{i,T\to \boldsymbol{M}\to Y^\star} &= Y_i^\star(t_1, \boldsymbol{M}_i(t_1)) - Y_i^\star(t_1, \boldsymbol{M}_i(t_0)), && (5)\\
\triangle_{i,T\to M^p\to Y^\star} &= Y^\star(t_1, \boldsymbol{M}_i(t_1)) - Y^\star(t_1, \boldsymbol{M}_i^{(-p)}(t_1)), && (6)
\end{aligned}$$

where $\boldsymbol{M}_i = (M_i^1, M_i^2, \dots, M_i^P)$, $\boldsymbol{M}_i^{(-p)}(t_1)$ refers to the replacement of the pth component of $\boldsymbol{M}_i(t_1)$ with the corresponding component from $\boldsymbol{M}_i(t_0)$, while keeping the other $P-1$ components unchanged.

$$\boldsymbol{M}_i^{(-p)}(t_1) = [M_i^1(t_1), \dots, M_i^{p-1}(t_1), M_i^p(t_0), M_i^{p+1}(t_1) \dots, M_i^P(t_1)]. \qquad (7)$$

Based on individual treatment effects (ITEs) presented in (4), (5) and (6), average treatment effects (ATEs) can be easily obtained by taking the expectation of the ITEs.

Conditional GAN

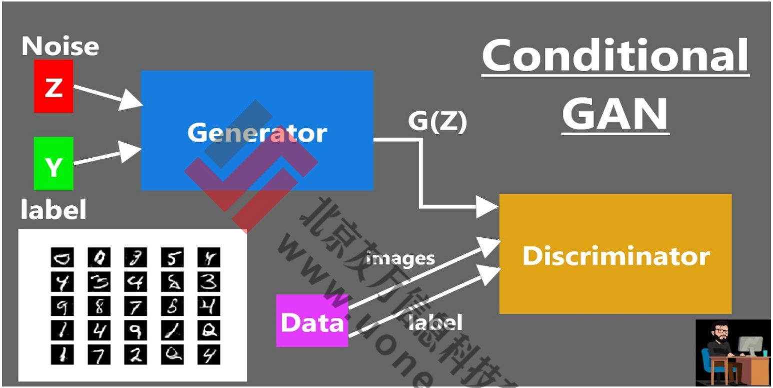


Figure: Basic architecture of CGAN. CGANs are well-suited for modeling complex conditional distributions in image tasks, such as neural style transfer (NST), image manipulation, data augmentation and image restoration.

Conditional GAN

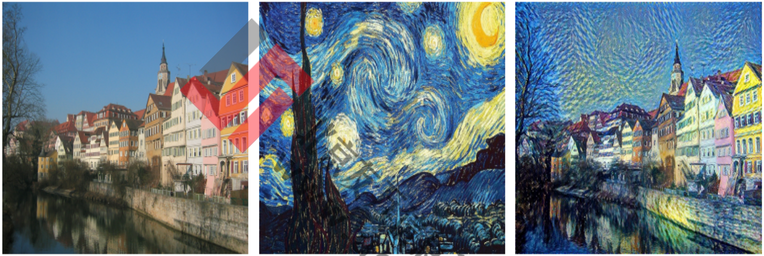


Figure: Application of CGAN in NST: generating a Van Gogh-style morden street view

Conditional GAN

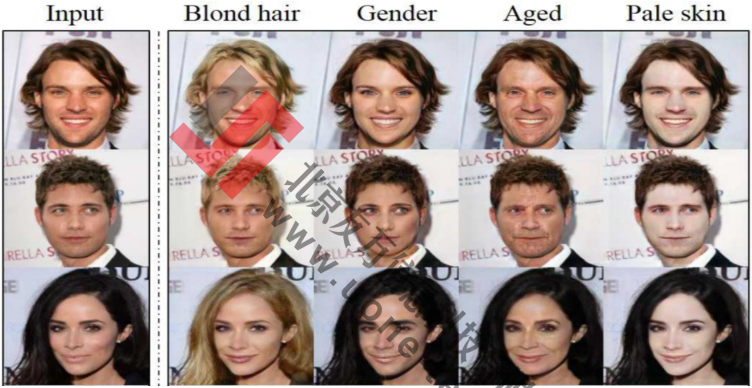


Figure: Application of CGAN in image-to-image translation

Conditional GAN

CGAN can be trained by

$$\min_{G} \max_{D} \mathbb{E}_{\boldsymbol{y} \sim p_{data}}[\log D(\boldsymbol{y}|\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_Z}[\log(1 - D(G(\boldsymbol{z}|\boldsymbol{x})))]. \tag{8}$$

where ***y*** represents real data, ***x*** is conditioning information.

Motivation

- Both of CMA and NST essentially estimate the conditional distributions of the variables of interest.
- If the treatment, counterfactual, and covariants are considered as the style label, desired generated data, and other characteristics of the image, respectively, the counterfactual estimation problem in CMA can be reinterpreted as an NST.
- If CGAN approach can employed for CMA, the advantages of GAN such as flexibility to describe nonlinearity and high capacity to approximate complex distributions, are expected to be introduced for more promising results.

GACMN

Analogous to the benchmark model (2), our proposed GACMN consists of two blocks, a mediator block and an outcome block. Each block is designed as a CGAN. In the following, we give the architectures of the two blocks.

The mediator block of GACMN. Corresponding to the first equation of (2), the mediator block is formulated by

$$\hat{\boldsymbol{M}} = G_M(\boldsymbol{Z}_M, T, \boldsymbol{X}; \theta_{G_M}), \tag{9}$$

where $\hat{\boldsymbol{M}}$ is the generated P-dimensional mediator, $\boldsymbol{Z}_M$ is a d_M-dimensional i.i.d. noise following a standard normal distribution, denoted as $\boldsymbol{Z}_M \sim N((0,1)^{d_M})$, θ_{G_M} represents the network parameters.

GACMN

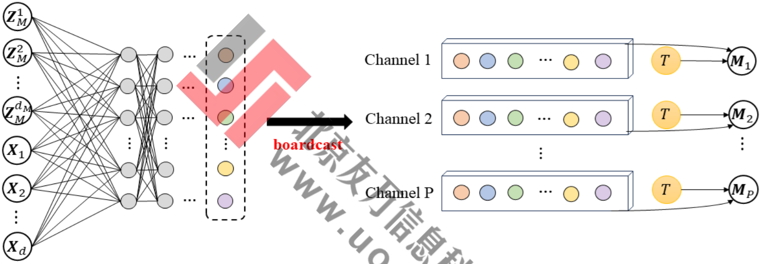


Figure: The network structure of G_M

GACMN

$D_M(\boldsymbol{M}, T, \boldsymbol{X}; \theta_{D_M})$ is set as a fully-connected FNN with parameters θ_{D_M}, and its output layer employs a sigmoid function. G_M and D_M together form the Conditional GAN of the mediator block, and the minimax optimization is given by

$$\begin{aligned}\min_{\theta_{G_M}} \max_{\theta_{D_M}} \quad & \mathbb{E}_{\boldsymbol{M} \sim P_{data}}[\log D_M(\boldsymbol{M}, T, \boldsymbol{X}; \theta_{D_M})] \\ & + \mathbb{E}_{\boldsymbol{Z}_M \sim N((0,1)^{d_M})}[\log(1 - D_M(G_M(\boldsymbol{Z}_M, T, \boldsymbol{X}; \theta_{G_M}), T, \boldsymbol{X}; \theta_{D_M}))].\end{aligned}$$

GACMN

The outcome block of GACMN. Corresponding to the second equation of (2), the outcome block is first formulated by

$$\hat{Y}^{\star} = G_Y(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M}; \theta_{G_Y}), \tag{10}$$

where $\hat{Y}^{\star}$ is the generated latent outcome, $\boldsymbol{Z}_Y$ is a d_Y-dimensional i.i.d. noise following a standard normal distribution, denoted as $\boldsymbol{Z}_Y \sim N((0,1)^{d_Y})$. θ_{G_Y} represents network parameters of G_Y.

GACMN

For inspecting the characteristics of the outcome variable, we propose using ReLU activation function to enable the network to achieve the censored output mechanism. ReLU is widely used in deep learning, defined as $RuLU(x) = \max(0, x)$. It is a piecewise linear function that directly outputs the input if it is positive, otherwise zero. Noticed that for the left-censored outcome variable Y,

$$Y = \max(0, Y^{\star}) = ReLU(Y^{\star}), \tag{11}$$

which indicates that the censored outcome mechanism can be perfectly realized by ReLU activation. Correspondingly, the outcome block is reasonably formulated by

$$\hat{Y} = \max(0, \hat{Y}^{\star}) = ReLU(\hat{Y}^{\star}) = ReLU(G_Y(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M}; \theta_{G_Y})). \tag{12}$$

GACMN

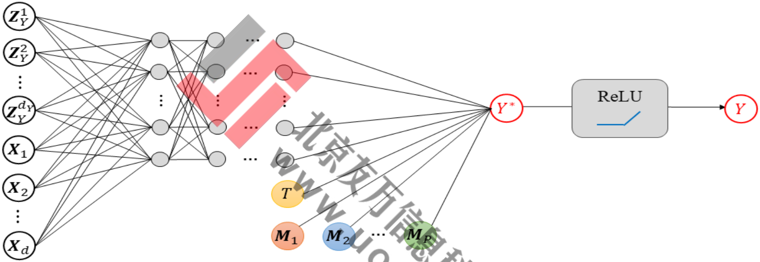


Figure: The network structure of G_Y

GACMN

$D_Y(Y, \boldsymbol{M}, T, \boldsymbol{X}; \theta_{D_Y})$ serves as the discriminator for the outcome block. D_Y is constructed as a multiple-layer fully-connected feedforward neural network with parameters θ_{D_M} and sigmoid function is employed for its output layer. The outcome block is trained by

$$
\begin{aligned}
&\min_{\theta_{G_Y}} \max_{\theta_{D_Y}} \mathbb{E}_{Y\sim P_{data}}[\log D_Y(Y, \boldsymbol{M}, T, \boldsymbol{X}; \theta_{D_Y})] \\
&+\mathbb{E}_{\boldsymbol{Z}_Y\sim N((0,1)^{d_Y})}[\log(1 - D_Y(ReLU(G_Y), \boldsymbol{M}, T, \boldsymbol{X}; \theta_{D_Y}))] \\
&+\alpha \cdot \mathbb{E}_{Y\sim P_{data}, \boldsymbol{Z}_Y\sim N((0,1)^{d_Y})}|Y - ReLU(G_Y(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M}; \theta_{G_Y}))|,
\end{aligned}
$$

A supervised loss is incorporated into the optimization process to boost training performance, motivated by J. Yoon, J. Jordon, M. Schaar, Gain: Missing data imputation using generative adversarial nets, in: International conference on machine learning, PMLR, 2018, pp. 5689-5698

GACMN

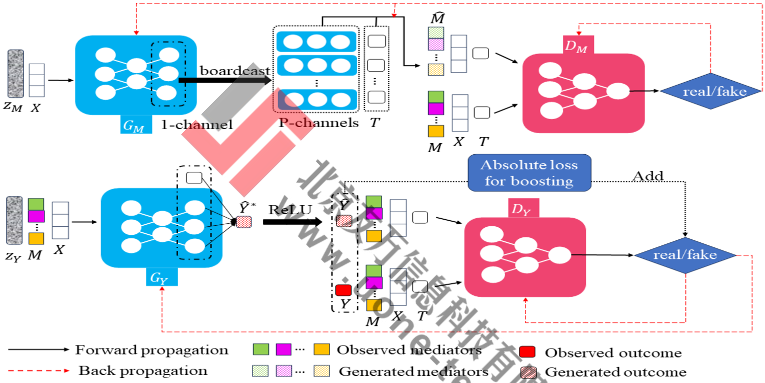


Figure: Schematics that depict the network structure of GACMN

GACMN

The direct effect. According to (4), the direct effects for the *i*th individual can be generated by

$$\triangle_{i,T\to Y^{\star}}(j) = \hat{Y}_i^{\star}(1, \hat{\boldsymbol{M}}_i(0), \boldsymbol{Z}^1_{Y_i}(j)) - \hat{Y}_i^{\star}(0, \hat{\boldsymbol{M}}_i(0), \boldsymbol{Z}^2_{Y_i}(j)),$$

where $\boldsymbol{Z}^1_{Y_i}(j)$ and $\boldsymbol{Z}^2_{Y_i}(j)$ ($j = 1, 2, \ldots N_g$) are independently drawn from the normal distribution $N((0,1)^{d_Y})$. Accordingly, the direct effect can be estimated as

$$\triangle_{T\to Y^{\star}} = \frac{1}{N_g}\frac{1}{n}\sum_{j=1}^{N_g}\sum_{i=1}^{n}\left(\hat{Y}_i^{\star}(1, \hat{\boldsymbol{M}}_i(0), \boldsymbol{Z}^1_{Y_i}(j)) - \hat{Y}_i^{\star}(0, \hat{\boldsymbol{M}}_i(0), \boldsymbol{Z}^2_{Y_i}(j))\right).$$

GACMN

The total indirect effect. According to (5), with $\hat{\boldsymbol{M}}_i(1)$, the total indirect effects for the ith individual can be generated by

$$\triangle_{i,T\to\boldsymbol{M}\to Y^\star}(j) = \hat{Y}_i^\star(1,\hat{\boldsymbol{M}}_i(1),\boldsymbol{Z}_{Y_i}^3(j)) - \hat{Y}_i^\star(1,\hat{\boldsymbol{M}}_i(0),\boldsymbol{Z}_{Y_i}^4(j)),$$

where $\boldsymbol{Z}_{Y_i}^3(j)$ and $\boldsymbol{Z}_{Y_i}^4(j)$ $(j = 1,2,\ldots N_g)$ are independently drawn from the normal distribution $N((0,1)^{d_Y})$. Then, the total indirect effect can be calculated as

$$\triangle_{T\to\boldsymbol{M}\to Y^\star} = \frac{1}{N_g}\frac{1}{n}\sum_{j=1}^{N_g}\sum_{i=1}^{n}\left(\hat{Y}_i^\star(1,\hat{\boldsymbol{M}}_i(1),\boldsymbol{Z}_{Y_i}^3(j)) - \hat{Y}_i^\star(1,\hat{\boldsymbol{M}}_i(0),\boldsymbol{Z}_{Y_i}^4(j))\right)$$

GACMN

The indirect effects through a given mediator. For $\forall p$ $(p = 1, 2, \ldots, P)$, with $\hat{\boldsymbol{M}}_i(1)$ and $\hat{\boldsymbol{M}}_i^{(-p)}(1)$, according to (6), the individual indirect effects implemented through the pth mediator M^p can be generated by

$$\triangle_{i,T\to M^p\to Y^\star}(j) = \hat{Y}_i^\star(1, \hat{\boldsymbol{M}}_i(1), \boldsymbol{Z}_{Y_i}^{p(1)}(j)) - \hat{Y}_i^\star(1, \hat{\boldsymbol{M}}_i^{(-p)}(1), \boldsymbol{Z}_{Y_i}^{p(2)}(j)),$$

where $\boldsymbol{Z}_{Y_i}^{p(1)}(j)$ and $\boldsymbol{Z}_{Y_i}^{p(2)}(j)$ $(j = 1, 2, \ldots N_g)$ are two group of noises independently drawn from $N((0,1)^{d_Y})$. Then, we have

$$\triangle_{T\to M^p\to Y^\star}$$
$$= \frac{1}{N_g}\frac{1}{n}\sum_{j=1}^{N_g}\sum_{i=1}^{n}\left(\hat{Y}_i^\star(1, \hat{\boldsymbol{M}}_i(1), \boldsymbol{Z}_{Y_i}^{p(1)}(j)) - \hat{Y}_i^\star(1, \hat{\boldsymbol{M}}_i^{(-p)}(1), \boldsymbol{Z}_{Y_i}^{p(2)}(j))\right).$$

Further discussions

1. **Modeling non-normality and heterogeneity.**
2. **The linear and nonlinear structures of GACMN.**
3. **Dealing with multiple mediators flexibly.**
4. **Handling censored outcome effectively.**
5. **Modeling different censoring situations.**

Convergence results

Identification is the first order problem in field of censored regression. For the outcome block, we follow the non-parametrics identification framework of S. Chen, G. B. Dahl, S. Khan, Nonparametric identification and estimation of a censored location scale regression model, JASA 100 (469) (2005) 212-221. for censored location-scale models:

$$
\begin{aligned}
Y_i^\star &= \mu(T_i, X_i) + \sigma_0(T_i, X_i)\epsilon_i, && (13)\\
Y_i &= \max(Y_i^\star, 0), && (14)
\end{aligned}
$$

where $\mu(\cdot)$ and $\sigma_0(\cdot)$ are the unknown location and scale functions and Y_i is observed right censored object while $Y_i^\star$ is latent object. $\mu(\cdot)$ is the our target for identification and estimation given zero median/mean of ϵ_i.

Convergence results

We begin by listing the key assumptions required for identification:

- **I1**: $P_{X,T}\big((X_i, T_i) : \mu(X_i, T_i) \geq 0\big) > 0$, where $P_{X,T}(\cdot)$ denotes the probability measure for joint random variables (X, T).
- **I2**: ϵ_i has median 0 and independent with (X_i, T_i).
- **I3**: The scale function $\sigma_0(\cdot)$ is continuous and strictly positive and bounded on every bounded subset of $\mathcal{X}$.
- **I4**: The location function $\mu(\cdot)$ is continuous and $|\mu(\cdot)| < \infty$ on every bounded subset of $\mathcal{X}$.

Convergence results

For a deep neutral network that has depth $\mathcal{H}$, width $\mathcal{W}$, and whole size S, we have following assumptions.

- **A1**: Target conditional generator $G(\theta_g^*)$ is continuous and its l_∞ norm is upper bounded.
- **A2**: For the optimal discriminator, $\frac{p_{X,Y,T}}{p_{X,RELU(G),Y}+p_{X,Y,T}}$ is lower and upper bounded in the support.
- **A3**:$\partial t_0 \mathbb{E}\left(Y^\star \mid X, T = t_0\right) = \partial t_0 \mu(X, t_0)$ exists and is finite (for T $\in \left\{t_1, t_0\right\}, \partial t_0 = t_1 - t_0$).
- **A4**: The l_∞ norm generator G within its support is upper bounded by constant B.
- **A5**: As sample size n goes to infinity, $\mathcal{H}\mathcal{W} \to 0$ and $\frac{BS\mathcal{H}\log(S)\log n}{n} \to 0$.

Let $\mathbb{D}_{TV}$ be the total variation defined as

$$\mathbb{D}_{TV}\left(p_{(X,Y,T)}, p_{(X,ReLU(G(\theta_g)),T)}\right) = \frac{1}{2}\left\|p_{(X,Y,T)} - p_{X,ReLU(G(\theta_g)),T}\right\|_1 .$$

Convergence results

Theorem 1. Assume **I1-I4** hold, $\mu(\cdot)$ for all $(X, T) \in \mathcal{X} \times \mathcal{T}$ is identifiable, and min-max CGANs problems can identify $\mu(X, T)$ as *median* $G(Z, X, T; \theta_g^*)$ in population.

Theorem 2. Suppose **A1-A5** and the assumptions in Theorem 1 hold, we have

$$\mathbb{E}_{X,T,Y,Z} D_{TV}(p_{(X,Y,T)}, p_{(X,ReLU(G(\hat{\theta}_g)),T)}) \xrightarrow{p} 0, \tag{15}$$

$$\mathbb{P}_{(X,ReLU(G(\hat{\theta}_g)),T)} \xrightarrow{d} \mathbb{P}_{(X,Y,T)}. \tag{16}$$

Hence we have further

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \left[\text{Median}(\hat{G}(Z_j, X_i, T = t_1)) - \text{Median}(\hat{G}(Z_j, X_i, T = t_0)) \right]$$

$$\xrightarrow{p} \quad \partial t_0 \mathbb{E}(Y^{\star} | X, T = t_0). \tag{17}$$

Experiments for artificial data

Consider the following data generation process

$$\begin{cases} M_i^1 = 0.5 + 0.2X_{1,i} + 0.8X_{2,i} + 0.5X_{1,i}X_{2,i} + 0.2X_{3,i} + 0.3X_{4,i} \\ \qquad + 0.1X_{5,i} + 0.4X_{6,i} + 2T_i + \varepsilon_{M^1}(i), \\ M_i^2 = 0.5 + 0.4X_{1,i} + 0.2\left(X_{1,i} + 1\right)^2 + 0.1\exp(X_{2,i}) + 0.2X_{5,i} \\ \qquad + 0.2X_{6,i} + \ 0.3X_{7,i} + 0.3X_{8,i} + 0.5T_i + \varepsilon_{M^2}(i), \\ M_i^3 = -0.5 + 0.3X_{1,i} + 0.2X_{2,i} + 0.1X_{1,i}^3 + 0.6X_{1,i}X_{2,i} + 0.3X_{7,i} \\ \qquad + 0.1X_{8,i} + \ 0.1X_{9,i} + 0.3X_{10,i} + T_i + \varepsilon_{M^3}(i), \\ Y_i^\star = 2.25 - 2\sqrt{X_{1,i} + 5} + \sin(1.5X_{2,i}) + 0.5M_i^1 + M_i^2 + 0.25X_{3,i} \\ \quad + 0.3X_{5,i} + \ 0.2X_{8,i} + 0.2X_{10,i} + 0.75T_i + \varepsilon_Y(i), \\ Y_i = \max\left(0, Y_i^\star\right) = ReLU(Y_i^\star), \quad i = 1, 2, \ldots, 25000. \end{cases}$$

For $\forall i$, $T_i \sim \mathcal{B}(0.5)$, $X_{1,i}, X_{2,i}, \cdots, X_{10,i}$ are independently drawn from $\mathcal{N}(0,1)$, where $\mathcal{B}(0.5)$ is the Bernoulli distribution, implying a equal probabilities for treated and untreated samples. The random terms are quite complex.

Experiments for artificial data

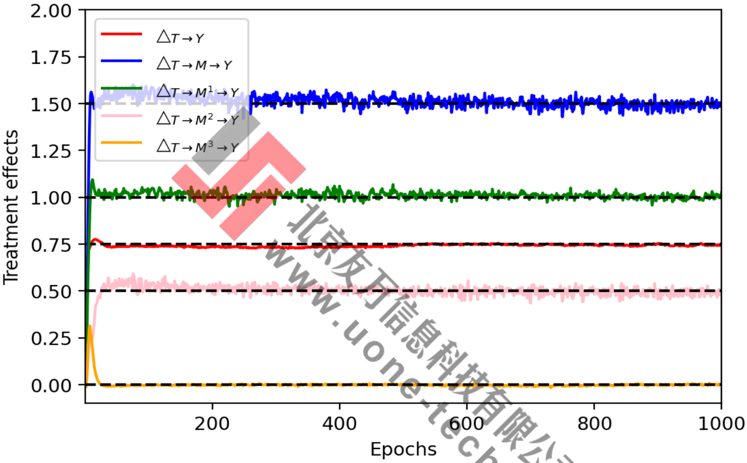
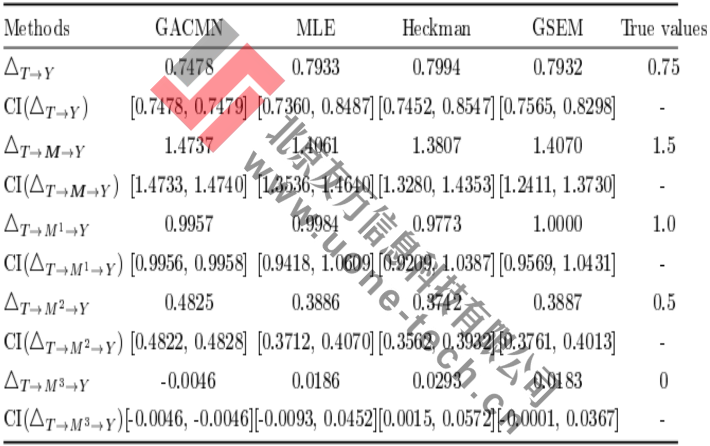


Figure: The estimated direct and indirect effects on testing set for each epoch.

Experiments for artificial data

Table 1: Average treatment effects by GAMN-M and traditional methods



Methods	GACMN	MLE	Heckman	GSEM	True values
$\Delta_{T\to Y}$	0.7478	0.7933	0.7994	0.7932	0.75
$\text{CI}(\Delta_{T\to Y})$	[0.7478, 0.7479]	[0.7360, 0.8487]	[0.7452, 0.8547]	[0.7565, 0.8298]	-
$\Delta_{T\to \boldsymbol{M}\to Y}$	1.4737	1.4061	1.3807	1.4070	1.5
$\text{CI}(\Delta_{T\to \boldsymbol{M}\to Y})$	[1.4733, 1.4740]	[1.3536, 1.4640]	[1.3280, 1.4353]	[1.2411, 1.3730]	-
$\Delta_{T\to M^1\to Y}$	0.9957	0.9984	0.9773	1.0000	1.0
$\text{CI}(\Delta_{T\to M^1\to Y})$	[0.9956, 0.9958]	[0.9418, 1.0609]	[0.9209, 1.0387]	[0.9569, 1.0431]	-
$\Delta_{T\to M^2\to Y}$	0.4825	0.3886	0.3742	0.3887	0.5
$\text{CI}(\Delta_{T\to M^2\to Y})$	[0.4822, 0.4828]	[0.3712, 0.4070]	[0.3562, 0.3932]	[0.3761, 0.4013]	-
$\Delta_{T\to M^3\to Y}$	-0.0046	0.0186	0.0293	0.0183	0
$\text{CI}(\Delta_{T\to M^3\to Y})$	[-0.0046, -0.0046]	[-0.0093, 0.0452]	[0.0015, 0.0572]	[-0.0001, 0.0367]	-

Experiments for artificial data

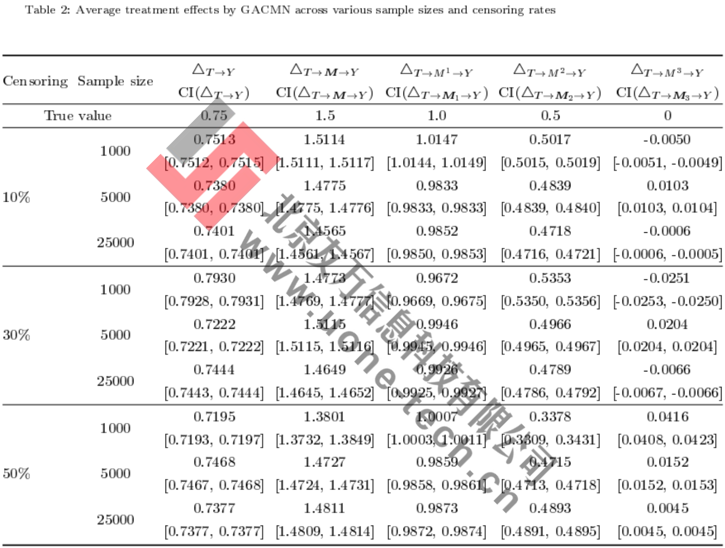


Table 2: Average treatment effects by GACMN across various sample sizes and censoring rates

Censoring	Sample size	$\triangle_{T\to Y}$	$\triangle_{T\to M\to Y}$	$\triangle_{T\to M^1\to Y}$	$\triangle_{T\to M^2\to Y}$	$\triangle_{T\to M^3\to Y}$
		$\mathrm{CI}(\triangle_{T\to Y})$	$\mathrm{CI}(\triangle_{T\to \boldsymbol{M}\to Y})$	$\mathrm{CI}(\triangle_{T\to \boldsymbol{M}_1\to Y})$	$\mathrm{CI}(\triangle_{T\to \boldsymbol{M}_2\to Y})$	$\mathrm{CI}(\triangle_{T\to \boldsymbol{M}_3\to Y})$
True value		0.75	1.5	1.0	0.5	0
10%	1000	0.7513	1.5114	1.0147	0.5017	-0.0050
		[0.7512, 0.7515]	[1.5111, 1.5117]	[1.0144, 1.0149]	[0.5015, 0.5019]	[-0.0051, -0.0049]
	5000	0.7380	1.4775	0.9833	0.4839	0.0103
		[0.7380, 0.7380]	[1.4775, 1.4776]	[0.9833, 0.9833]	[0.4839, 0.4840]	[0.0103, 0.0104]
	25000	0.7401	1.4565	0.9852	0.4718	-0.0006
		[0.7401, 0.7401]	[1.4561, 1.4567]	[0.9850, 0.9853]	[0.4716, 0.4721]	[-0.0006, -0.0005]
30%	1000	0.7930	1.4773	0.9672	0.5353	-0.0251
		[0.7928, 0.7931]	[1.4769, 1.4777]	[0.9669, 0.9675]	[0.5350, 0.5356]	[-0.0253, -0.0250]
	5000	0.7222	1.5115	0.9946	0.4966	0.0204
		[0.7221, 0.7222]	[1.5115, 1.5116]	[0.9945, 0.9946]	[0.4965, 0.4967]	[0.0204, 0.0204]
	25000	0.7444	1.4649	0.9926	0.4789	-0.0066
		[0.7443, 0.7444]	[1.4645, 1.4652]	[0.9925, 0.9927]	[0.4786, 0.4792]	[-0.0067, -0.0066]
50%	1000	0.7195	1.3801	1.0007	0.3378	0.0416
		[0.7193, 0.7197]	[1.3732, 1.3849]	[1.0003, 1.0011]	[0.3309, 0.3431]	[0.0408, 0.0423]
	5000	0.7468	1.4727	0.9859	0.4715	0.0152
		[0.7467, 0.7468]	[1.4724, 1.4731]	[0.9858, 0.9861]	[0.4713, 0.4718]	[0.0152, 0.0153]
	25000	0.7377	1.4811	0.9873	0.4893	0.0045
		[0.7377, 0.7377]	[1.4809, 1.4814]	[0.9872, 0.9874]	[0.4891, 0.4895]	[0.0045, 0.0045]

Experiments for real-world data

The China Household Finance Survey (CHFS) is a comprehensive and nationally representative survey conducted by Southwestern University of Finance and Economics and the Research Institute of Economics and Management every two years. It is designed to collect detailed and comprehensive information on the financial assets, income, expenditures, and demographic characteristics of Chinese households, aiming to provide a comprehensive understanding of their financial conditions and behaviors. In 2019, data was collected from more than 30,000 households.

Experiments for real-world data

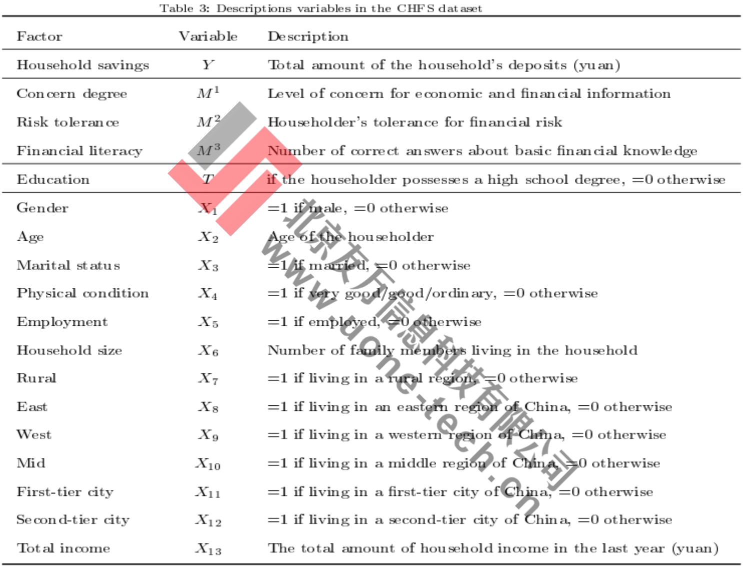
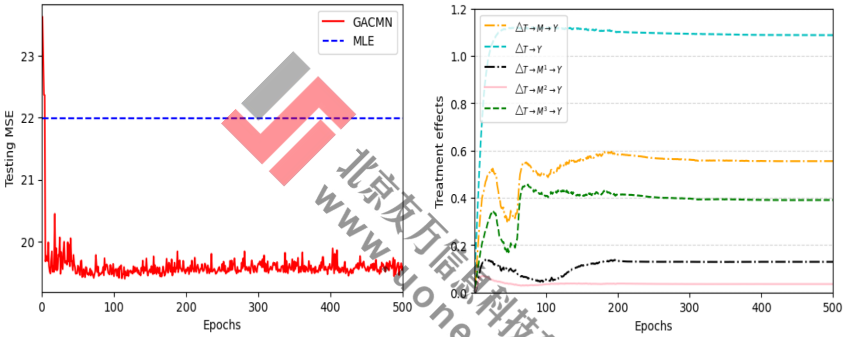


Table 3: Descriptions variables in the CHFS dataset

Factor	Variable	Description
Household savings	Y	Total amount of the household's deposits (yuan)
Concern degree	M^1	Level of concern for economic and financial information
Risk tolerance	M^2	Householder's tolerance for financial risk
Financial literacy	M^3	Number of correct answers about basic financial knowledge
Education	T	if the householder possesses a high school degree, =0 otherwise
Gender	X_1	=1 if male, =0 otherwise
Age	X_2	Age of the householder
Marital status	X_3	=1 if married, =0 otherwise
Physical condition	X_4	=1 if very good/good/ordinary, =0 otherwise
Employment	X_5	=1 if employed, =0 otherwise
Household size	X_6	Number of family members living in the household
Rural	X_7	=1 if living in a rural region, =0 otherwise
East	X_8	=1 if living in an eastern region of China, =0 otherwise
West	X_9	=1 if living in a western region of China, =0 otherwise
Mid	X_{10}	=1 if living in a middle region of China, =0 otherwise
First-tier city	X_{11}	=1 if living in a first-tier city of China, =0 otherwise
Second-tier city	X_{12}	=1 if living in a second-tier city of China, =0 otherwise
Total income	X_{13}	The total amount of household income in the last year (yuan)

Experiments for real-world data



(a)

(b)

Figure: The leaning performance on testing set for each epoch.

Experiments for real-world data

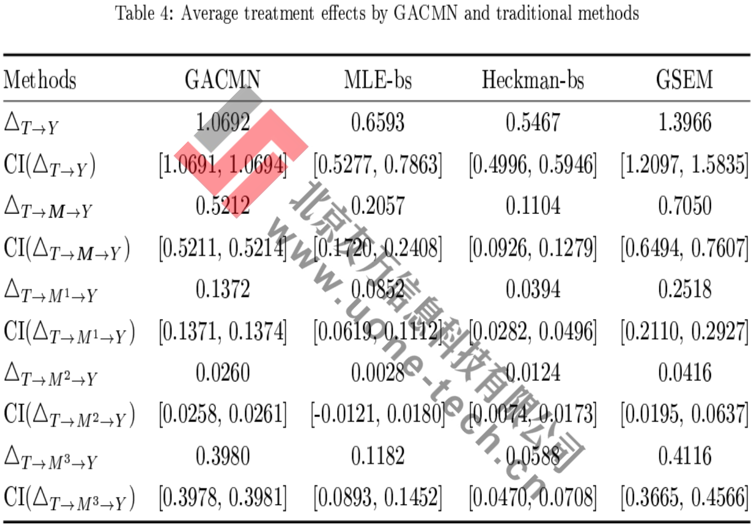


Table 4: Average treatment effects by GACMN and traditional methods

Methods	GACMN	MLE-bs	Heckman-bs	GSEM
$\Delta_{T\to Y}$	1.0692	0.6593	0.5467	1.3966
$\mathrm{CI}(\Delta_{T\to Y})$	[1.0691, 1.0694]	[0.5277, 0.7863]	[0.4996, 0.5946]	[1.2097, 1.5835]
$\Delta_{T\to M\to Y}$	0.5212	0.2057	0.1104	0.7050
$\mathrm{CI}(\Delta_{T\to M\to Y})$	[0.5211, 0.5214]	[0.1720, 0.2408]	[0.0926, 0.1279]	[0.6494, 0.7607]
$\Delta_{T\to M^1\to Y}$	0.1372	0.0852	0.0394	0.2518
$\mathrm{CI}(\Delta_{T\to M^1\to Y})$	[0.1371, 0.1374]	[0.0619, 0.1112]	[0.0282, 0.0496]	[0.2110, 0.2927]
$\Delta_{T\to M^2\to Y}$	0.0260	0.0028	0.0124	0.0416
$\mathrm{CI}(\Delta_{T\to M^2\to Y})$	[0.0258, 0.0261]	[-0.0121, 0.0180]	[0.0074, 0.0173]	[0.0195, 0.0637]
$\Delta_{T\to M^3\to Y}$	0.3980	0.1182	0.0588	0.4116
$\mathrm{CI}(\Delta_{T\to M^3\to Y})$	[0.3978, 0.3981]	[0.0893, 0.1452]	[0.0470, 0.0708]	[0.3665, 0.4566]

Conclusions

- This paper introduces a novel GAN-based mediation model.
- The innovation lies in a creative reinterpretation of the conventional CMA problem, framing it as an NST problem.
- Several key efforts have been made, including network architecture, ReLU activation for managing censored output, partially linear network structure to augment interpretability, a novel min-max optimization scheme.
- Solid theoretical results.
- This study also marks a substantial stride toward the development of effective methodologies for deploying generative learning methods for various mediation problems, and these compelling topics shall be focused in our future studies.

The End