GAHMN: A Generative Approach for High-Dimensional Mediation Analysis

Jiaming Zhang[1], Yiqi Lin[2], Rou Zhang[1] Xinyuan Song[2], Hanwen Ning[1 3] *

[1]Department of Statistics, Zhongnan University of Economics and Law, Wuhan, 430073, P.R. China
[2]Department of Statistics, The Chinese University of Hong Kong, Shatin, NT., Hong Kong, China
[3]Innovation and Talent Base for Digital Technology and Finance, Zhongnan University of Economics and Law, P.R. China

Abstract

High-dimensional mediation analysis (HMA) seeks to uncover complex causal mechanisms involving numerous mediators and plays a crucial role in scientific and social sciences. In this work, we introduce the Generative Adversarial High-dimensional Mediation Network (GAHMN), a novel, scalable structured generative framework designed for causal analysis in high-dimensional settings. GAHMN formulates mediation analysis as dual conditional generative blocks, explicitly capturing mediators' dual roles as outcomes influenced by treatments and as predictors affecting outcomes. Each block integrates a high-dimensional partially linear structure with multi-channel convolutional layers, promoting effective parameter sharing and enhanced representation learning. To induce sparsity and accurate mediator selection, GAHMN employs customized min-max optimization problems with ℓ_1 penalties on generator parameters, alongside specially designed optimization algorithms for efficient computation. Unlike existing benchmark methods relying on restrictive parametric assumptions or random-effect specifications, GAHMN flexibly captures heterogeneity, complex distributions, and inter-mediator correlations. With careful design, the complexity of GAHMN is $O(p)$ rather than $O(p^2)$ (p is the number of mediators) in conventional approach. Theoretical results rigorously ensures estimation consistency, convergence rate, and accurate sparse recovery. GAHMN also serves as a structured generative causal modeling framework, extending to causal decomposition, structural equation modeling, and counterfactual policy evaluation. Extensive experiments confirm GAHMN's superior performance and robustness in synthetic and real-world scenarios.

Introduction

Causal mediation analysis (CMA) provides a principled framework for understanding how an input variable X affects an outcome variable Y through intermediate variable, known as mediator M. The goal is to decompose the total effect of X on Y into a direct effect and an indirect effect transmitted via M. Traditional mediation analysis has been well studied in low-dimensional settings, where the number of mediators is small (MacKinnon 2012). However, in many modern scientific domains, such as genomics (Liu et al. 2022), neuroscience (Lindquist 2012), and social science (Liu et al. 2021), researchers are increasingly faced with HMA, where the number of mediators can be substantial. Classical statistical methods encounter significant challenges in HMA due to the curse of dimensionality, computational scalability, and modeling flexibility (Zhang, Hou, and Liu 2021).

Existing benchmark HMA methods extend low-dimensional models by incorporating regularization techniques such as LASSO for variable selection and estimation (Guo et al. 2022; Zhang, Yang, and Yang 2022). While these methods can partially alleviate dimensionality issues, they rely on rigid assumptions, such as linearity, homogeneity, and Gaussianity, that are often violated in practice. In real-world data from fields such as financial econometrics (Carpena and Zia 2020), neuroscience (Nath et al. 2023), and survival analysis (Zhou and Song 2021; Zhang et al. 2021), the underlying relationships among variables are often nonlinear, heterogeneous, and influenced by latent confounders, leading to strong inter-mediator correlations. In low-dimensional scenarios, these issues can be addressed by modeling covariance structures or latent factors explicitly (Celli 2022; Xu, Liu, and Liu 2022). Yet, such strategies become computationally infeasible as the number of mediators grows, often requiring $O(p^2)$ model complexity to accommodate heterogeneity and dependencies, where p is the number of mediators. The rigid assumptions together with quadratic complexity, pose a bottleneck for accurate high-dimensional inference. (Zeng, Shao, and Zhou 2021; Yang et al. 2024).

To overcome these limitations, we propose a new approach to HMA from the perspective of generative learning. We note that GANs (Goodfellow et al. 2014; Gui et al. 2021) have proven effective in modeling complex distributions, particularly in image generation, where data are typically high-dimensional, structured, and highly correlated. Inspired by this, we explore the potential of GANs to effectively model causal mediation mechanisms in high-dimensional settings.

We introduce a novel generative learning model termed GAHMN. GAHMN specifically addresses the structural characteristics of HMA, where mediators simultaneously function as outcomes influenced by treatments and as predictors of the final outcome. GAHMN consists of two condi-

*Corresponding Author. (Email: ninghanwen@gmail.com)

tional GAN (CGAN) modules: the first models mediator distributions conditioned on treatment exposure, and the second models outcomes conditioned on mediators and covariates. Both modules incorporate partially linear structures combined with multi-channel convolutional layers. This design enables clear causal effect decomposition within a counterfactual framework, facilitates effective parameter sharing, and enhances representation learning without compromising modeling flexibility. Unlike existing benchmark methods that rely heavily on restrictive parametric assumptions or random-effect specifications, GAHMN's carefully designed architecture flexibly captures mediator heterogeneity, nonlinear interactions, and inter-mediator correlations. The complexity of GAHMN is $O(p)$ rather than $O(p^2)$, substantially improving estimation accuracy and generalization to large-scale causal inference problems. To encourage sparsity and accurate mediator selection, we propose two customized min-max optimization problems with ℓ_1 penalties imposed on generator parameters, accompanied by specialized optimization algorithms ensuring efficient computation and stable training.

Beyond HMA, GAHMN serves as a structured generative causal modeling framework applicable to broader tasks, including causal decomposition, structural equation modeling, and counterfactual policy evaluation. We rigorously establish statistical properties, including estimation consistency, convergence rate, and sparse recovery. Extensive numerical experiments conducted on both synthetic and real-world datasets validate the superior performance and robustness of our proposed approach.

Benchmark Methods and Limitations

Notations and Benchmark Methods

Causal mediation analysis decomposes the total treatment effect into a direct effect and an indirect effect through mediators (MacKinnon 2012). In the following, let $\boldsymbol{X} \in \mathbb{R}^d$ denotes covariates, $\boldsymbol{M} \in \mathbb{R}^p$ the mediators, $T \in \{t_0, t_1\}$ the binary treatment, and $Y \in \mathbb{R}$ the outcome. For each sample $i = 1, \ldots, n$, we observe $(\boldsymbol{X}_i, \boldsymbol{M}_i, T_i, Y_i)$. We assume $\boldsymbol{M}$ is high-dimensional and sparse, meaning only a few mediators transmit the effect of T to Y. We investigate HMA problems under counterfactual framework (Imai, Keele, and Tingley 2010). When $T_i = t_1$, the factual outcomes $\boldsymbol{M}_i(t_1)$ and $Y_i(t_1, \boldsymbol{M}_i(t_1))$ are observed. The unobserved counterfactuals include $\boldsymbol{M}_i(t_0)$, $Y_i(t_1, \boldsymbol{M}_i(t_0))$, and $Y_i(t_1, \boldsymbol{M}_i^{(-k)}(t_1))$. When $T_i = t_0$, the reverse holds. Our objective is to learn the conditional distributions of $\boldsymbol{M}$ and Y given $(T, \boldsymbol{X})$. The benchmark mediation model can be generally formulated as follows:

$$\begin{cases} \boldsymbol{M} = \boldsymbol{\beta}_1 + \boldsymbol{a}T + g_1(\boldsymbol{X}) + \boldsymbol{\epsilon}_1, \\ Y = \beta_2 + c'T + \boldsymbol{b}^\top \boldsymbol{M} + g_2(\boldsymbol{X}) + \epsilon_2. \end{cases} \tag{1}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are unknown linear or nonlinear functions. The indirect effect is given by $\boldsymbol{a} \odot \boldsymbol{b}$, and parameters can be estimated via OLS, MLE, or SEM-based methods (Valente et al. 2020; Gunzler et al. 2013; Yuan and Qu 2024; Hu et al. 2024). In high-dimensional settings, it is often assumed that only a small subset of mediators are active. Penalized regression methods such as LASSO, SCAD, and Elastic Net are used to identify relevant mediators (Guo et al. 2022; Wei et al. 2025; Jacobucci, Brandmaier, and Kievit 2019). These allow flexible modeling of confounding while retaining interpretability. A two-step estimation is typically used to recover active mediators and their corresponding effects (Cai et al. 2022; Wang and Huang 2024). Further reviews of existing HMA methods and counterfacual framework are presented in Appendix A.

Limitations

Despite their effectiveness, the existing HMA methods face several major limitations:

Restrictive random assumptions: The random terms $\boldsymbol{\epsilon_1}$ and ϵ_2 are assumed to be normally distributed and homoscedastic to facilitate estimation and inference. However, mediation analysis data from individuals (Hayes 2017; Preacher and Hayes 2008; Xie, Zou, and Qi 2018) exhibit significant complexity and heterogeneity, making these assumptions overly restrictive. The covariates $\boldsymbol{X}$ are treated additively and independently of the noise terms, ignoring their interactions. This limits their ability to capture heterogeneity and individual-level variation.

Inadequate modeling of mediator dependencies: The mediators are often driven by underlying common factors, resulting in non-negligible correlations. Assuming conditional independence among mediators oversimplifies their complex interactions and may lead to biased estimates. Traditional approaches, such as modeling the covariance matrix of $\boldsymbol{\epsilon_1}$, are effective in low-dimensional settings but become infeasible in high-dimensional contexts due to parameter explosion ($O(p^2)$).

Limitations of likelihood methods: Likelihood-based methods (Yuan and MacKinnon 2009; Sun and Song 2024) provide asymptotically efficient estimators under correct model specification (e.g., achieving the Cramér-Rao lower bound) but are highly sensitive to misspecification and require strong distributional assumptions. These challenges necessitate more flexible and robust alternatives.

Within the counterfactual framework, estimating unobserved potential outcomes requires modeling the conditional distribution of outcomes given treatment, mediators, and covariates. CGANs offer a flexible and assumption-light alternative by learning the conditional distribution via adversarial training. These advantages motivate us to use CGANs for HMA problems, providing a more robust and flexible approach under the counterfactual perspective.

Methodology

Corresponding to (1), GAHMN consists of two distinct components: a mediation block and an outcome block, both constructed using CGANs.

The Mediator Block of GAHMN

Corresponding to the first equation of (1), the generator of the mediator block is given by

$$\hat{\boldsymbol{M}} = G_M(\boldsymbol{Z}_M, T, \boldsymbol{X}; \theta_{G_M}), \tag{2}$$

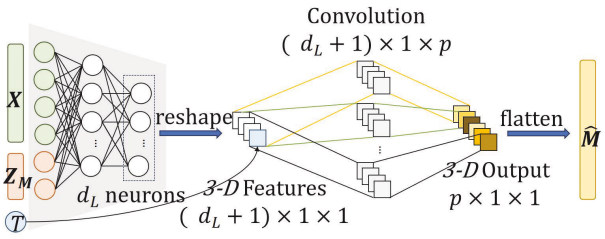

Figure 1: The Network Structure of G_M.

where $\hat{\boldsymbol{M}}$ represents the generated p-dimensional mediator, $\boldsymbol{Z}_M \sim N(\boldsymbol{0}, \boldsymbol{I}_{d_M})$ is a d_M-dimensional i.i.d. noise. θ_{G_M} represents model parameters. The network structure is illustrated in Figure 1. The input layer of G_M consists of $(\boldsymbol{Z}_M, T, \boldsymbol{X})$, where $\boldsymbol{Z}_M$ and $\boldsymbol{X}$ constitute a fully-connected neural network (FNN). The FNN maps $\boldsymbol{Z}_M$ and $\boldsymbol{X}$ into a latent feature representation by a hidden layer with d_L neurons. We reshape the extracted latent features into a tensor of size $(1, 1, d_L)$, then, T is appended as the $(d_L + 1)$th channel, augmenting the tensor to dimension $(1, 1, d_L + 1)$. A 1×1 convolutional kernel is applied to this tensor. This operation produces an output tensor with p channels and each channel involves $d_L + 1$ weight parameters. Specifically, the $(d_L + 1)$th weight parameter of each channel, associated with T, forms a high-dimensional partially linear structure that quantifies the direct effect of the treatment on each corresponding mediator, providing a clear interpretation within the learned representation. The discriminator in the mediator block, denoted as $D_M(\boldsymbol{M}, T, \boldsymbol{X})$ is implemented by an FNN with parameters θ_{D_M}. The output layer of D_M consists of p neurons and activated by sigmoid function.

Minmax Loss Function In HMA, the influence of of T on $\boldsymbol{M}$ is sparse. The expected change in $\boldsymbol{M}$ when T changes from t_0 to t_1 is $\mathbb{E}(\boldsymbol{M}(t_1) - \boldsymbol{M}(t_0))$, and only a few elements of this difference are expected to deviate significantly from zero. By G_M, for any given individual i, we can generate its factual and counterfactual under both treatments, i.e., $G_M(\boldsymbol{Z}_M, t_1, \boldsymbol{X}_i; \theta_{G_M})$ and $G_M(\boldsymbol{Z}_M, t_0, \boldsymbol{X}_i; \theta_{G_M})$. The minmax optimization for the mediation block is given by

$$\begin{aligned}&\min_{\theta_{G_M}} \max_{\theta_{D_M}} \mathbb{E}_{M \sim P_{data}}[\log D_M(\boldsymbol{M}, T, \boldsymbol{X})] \\ &+\mathbb{E}_{\boldsymbol{Z}_M}[\log(1 - D_M(G_M(\boldsymbol{Z}_M, T, \boldsymbol{X}), T, \boldsymbol{X}))] \\ &+\lambda_1 \mathbb{E}_{\boldsymbol{Z}_M} ||G_M(\boldsymbol{Z}_M, t_1, \boldsymbol{X}) - G_M(\boldsymbol{Z}_M, t_0, \boldsymbol{X})||_1, \quad (3)\end{aligned}$$

where λ_1 is a hyperparameter use for balancing. The first two terms drive the model to learn the conditional distribution associated with $\boldsymbol{M}$, T and $\boldsymbol{X}$. The third term is introduced to identify the sparsity associated with T.

The Outcome Block of GAHMN

Corresponding to the second equation of (1), the generator of the outcome block is given by

$$\hat{Y} = G_Y(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M}; \theta_{G_Y}), \quad (4)$$

where $\hat{Y}$ is the generated outcome, $\boldsymbol{Z}_Y \sim N(\boldsymbol{0}, \boldsymbol{I}_{d_Y})$ is a d_Y-dimensional i.i.d. noise. θ_{G_Y} represents the network parameters. The network structure is shown in Figure 2. The

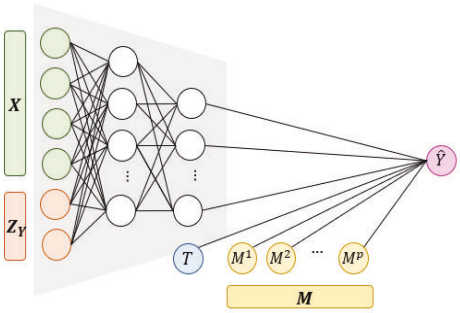

Figure 2: The Network Structure of G_Y.

input layer of G_Y consists of $(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M})$, where $\boldsymbol{Z}_Y$ and $\boldsymbol{X}$ constitute a multiple-layer FNN, while T and $\boldsymbol{M}$ are in the last hidden layer, forming a partially linear structure. The discriminator in the mediator block, denoted as $D_Y(Y, \boldsymbol{M}, T, \boldsymbol{X}; \theta_{D_Y})$ is implemented as an FNN with parameters θ_{D_Y}.

Minmax Loss Function In the both blocks, we encounter challenges related to high dimensionality. The key difference is that in the mediator block, $\boldsymbol{M}$ appears as a high-dimensional output variable, whereas in the outcome block, $\boldsymbol{M}$ serves as an input to the model. $\boldsymbol{M}$ is continuous with $\boldsymbol{M}_i$ varying substantially across individuals. As the dimensionality of $\boldsymbol{M}$ increases, the observed set $\{\boldsymbol{M}_i\}_{i=1}^n$ becomes more diverse, and more gaps between different samples associated with $\boldsymbol{M}_i$s arise, there may be few or even no samples corresponding to certain conditions. This can lead to poor generalization when the generator encounters a new condition that is not observed in the training set. It is important to note that the generator $\boldsymbol{G}_Y$ is used to describe the conditional distribution $P(Y|\boldsymbol{T}, \boldsymbol{X}, \boldsymbol{M})$. For continuous-valued conditioning variables, we want a minor perturbation to the condition to only slightly disturb the conditional distribution, meaning that the distribution obtained by $\boldsymbol{G}_Y$ should shift smoothly as we change $\boldsymbol{M}$ gradually.

To improve the generalization and identify the mediators that truly influence the outcome, we introduce the following partially L_1 regularization

$$L_{RY}(G) = \mathbb{E}_{\substack{\boldsymbol{Z}_Y \sim N(\boldsymbol{0}, \boldsymbol{I}_{d_Y}) \\ T, \boldsymbol{M}, \boldsymbol{X} \sim P_{\text{data}}}} \| \bigtriangledown_{\boldsymbol{M}} G_Y(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M}) \|_1, \quad (5)$$

where $\nabla_{\boldsymbol{M}} G_Y$ denotes the gradient vector of the generator G_Y with respect to the mediator input $\boldsymbol{M}$. This regularization term penalizes the ℓ_1-norm of the mediator gradients, encouraging sparsity in the functional dependency of the outcome on mediators, and thus enhancing interpretability and variable selection.

By integrating this regularization term, we have the following minmax optimization for the outcome block

$$\begin{aligned}&\min_{\theta_{G_Y}} \max_{\theta_{D_Y}} \mathbb{E}_{Y \sim P_{data}}[\log D_Y(Y, \boldsymbol{M}, T, \boldsymbol{X})] \\ &+\mathbb{E}_{\boldsymbol{Z}_Y}[\log(1 - D_Y(G_Y(\boldsymbol{Z}_Y, T, \boldsymbol{X}, \boldsymbol{M}), \boldsymbol{M}, T, \boldsymbol{X}))] \\ &+\lambda_2 L_{RY}(G). \quad (6)\end{aligned}$$

where λ_2 is a hyperparameter. An appropriate λ_2 can bring both accurate predictions and identification of sparsity.

Training of GAHMN

We adopt a stepwise estimation approach to train the two blocks. Each iteration of the optimization of (3) is in three steps. First, we update θ_{D_M} by maximizing the first two terms using the Adam. Second, with θ_{D_M} fixed, we update θ_{G_M} by minimizing the same two terms. Third, we further refine θ_{G_M} by minimizing $\mathbb{E}_{\boldsymbol{Z}_M}\|G_M(\boldsymbol{Z}_M, t_1, \boldsymbol{X}) - G_M(\boldsymbol{Z}_M, t_0, \boldsymbol{X})\|_1$ using the proximal gradient method (PROX) (Parikh, Boyd et al. 2014). The updated parameters are passed to the next iteration. Similarly, the optimization of (6) also follows a three-step process in each iteration. The first two steps involve standard GAN training: we alternately update θ_{D_Y} and θ_{G_Y} via min-max optimization using Adam. In the third step, we update θ_{G_Y} by minimizing $L_{RY}(G)$ using the PROX. The update schemes of GAHMN are further specified in Algorithm 1 in Appendix B.

Estimating Direct and Indirect Effects

The well trained GAHMN is used to generate counterfactuals for mediation analysis. For a given individual i ($i = 1, 2, \ldots, n$), we sample $\boldsymbol{Z}_{M_i}(j)$s from $N(\boldsymbol{0}, \boldsymbol{I}_{d_M})$. $j = 1, 2, \ldots, N_g$. The generated $\boldsymbol{M}_i$s under different treatments are:

$$\hat{\boldsymbol{M}}_i(t, \boldsymbol{Z}_{M_i}(j)) = G_M(\boldsymbol{Z}_{M_i}(j), t, \boldsymbol{X}_i; \hat{\theta}_{G_M}), \quad (7)$$

where $t = t_0, t_1$. By averaging these generated mediator values, we can compute the filtered factual for $\boldsymbol{M}_i$.

$$\hat{\boldsymbol{M}}_i(t) = \frac{1}{N_g}\sum_{j=1}^{N_g} \hat{\boldsymbol{M}}_i(t_0, \boldsymbol{Z}_{M_i}(j)). \quad (8)$$

Then, the counterfactual associated with $\boldsymbol{M}_i^{(-k)}$ is given by

$$\hat{\boldsymbol{M}}_i^{(-k)}(t_1) = \hat{\boldsymbol{M}}_i(t_1) \odot \boldsymbol{T}_k + \hat{\boldsymbol{M}}_i(t_0) \odot (\boldsymbol{1} - \boldsymbol{T}_k), \quad (9)$$

where $\boldsymbol{T}_k$ is a p-dimensional vector with 0 at the k-th position and 1 elsewhere. $\boldsymbol{1}$ is a p-dimensional vector of all ones. $\odot$ denotes element-wise product. In the following, $\boldsymbol{Z}^1_{Y_i}(j)$, $\boldsymbol{Z}^2_{Y_i}(j)$ and $\boldsymbol{Z}^3_{Y_i}(j)$ ($j = 1, 2, \ldots N_g$) are independently sampled from $N(\boldsymbol{0}, \boldsymbol{I}_{d_Y})$. All the treatment/mediation effects are calculated under counterfactual framework.

The Direct Effect Let $\hat{Y}_i(t, \hat{\boldsymbol{M}}_i(t_0), \boldsymbol{Z}^1_{Y_i}(j))$s ($t = t_0, t_1$) denote the generated values of Y_i under varying treatments. The generated values are computed by

$$\hat{Y}_i(t, \hat{\boldsymbol{M}}_i(t_0), \boldsymbol{Z}^1_{Y_i}(j)) = G_Y(\boldsymbol{Z}^1_{Y_i}(j), \hat{\boldsymbol{M}}_i(t_0), t, \boldsymbol{X}_i; \hat{\theta}_{G_Y}).$$

Then, the direct effect can be estimated as

$$\triangle_{T\to Y} = \frac{1}{N_g}\frac{1}{n}\sum_{j=1}^{N_g}\sum_{i=1}^{n}\Big(\hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t_0), \boldsymbol{Z}^1_{Y_i}(j)) - \hat{Y}_i(t_0, \hat{\boldsymbol{M}}_i(t_0), \boldsymbol{Z}^1_{Y_i}(j))\Big). \quad (10)$$

The Total Indirect Effect The generated outcomes $\hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t_1), \boldsymbol{Z}^2_{Y_i}(j))$s and $\hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t_0), \boldsymbol{Z}^2_{Y_i}(j))$s, are computed as follows: $\hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t), \boldsymbol{Z}^2_{Y_i}(j)) = G_Y(\boldsymbol{Z}^2_{Y_i}(j), \hat{\boldsymbol{M}}_i(t), t_1, \boldsymbol{X}_i)$. where $t = t_0, t_1$. Then, the total indirect effect can be calculated as

$$\triangle_{T\to \boldsymbol{M}\to Y} = \frac{1}{N_g}\frac{1}{n}\sum_{j=1}^{N_g}\sum_{i=1}^{n}\Big(\hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t_1), \boldsymbol{Z}^2_{Y_i}(j)) - \hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t_0), \boldsymbol{Z}^2_{Y_i}(j))\Big). \quad (11)$$

The Indirect Effects Through a Given Mediator For $\forall k$ ($k = 1, 2, \ldots, p$), with $\hat{\boldsymbol{M}}_i(t_1)$, $\hat{\boldsymbol{M}}_i^{(-k)}(t_1)$ and the generated outcomes by $\hat{G}_Y$, the indirect effect through the kth mediator can be estimated as

$$\triangle_{T\to M^k\to Y} = \frac{1}{N_g}\frac{1}{n}\Big(\hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i(t_1), \boldsymbol{Z}^3_{Y_i}(j)) - \hat{Y}_i(t_1, \hat{\boldsymbol{M}}_i^{(-k)}(t_1), \boldsymbol{Z}^3_{Y_i}(j))\Big). \quad (12)$$

Confidence Intervals (CI) The CIs of the treatment effects can be established by the quantiles of mimic results.

The block diagram of GAHMN is summarized and presented in Figure 3 .

Discusssions

GAHMN demonstrates several promising properties and notable advantages.

Modeling mediator inter-correlations and heterogeneity. In contrast to the benchmark methods relying on restrictive distributional assumptions, GAHMN utilize $\boldsymbol{Z}_M$ and $\boldsymbol{Z}_Y$ to represent latent noise components, which are transformed through deep networks to approximate complex and potentially correlated distributions of $\boldsymbol{\epsilon}_1$ and ϵ_2. This eliminates the need to explicitly estimate a $p \times p$ covariance matrix for inter-mediator correlations, reducing model complexity. Furthermore, the architecture places $\boldsymbol{Z}_M$, $\boldsymbol{Z}_Y$, and $\boldsymbol{X}$ jointly in the input layer, allowing interactions between covariates and latent noise, and enabling the model to naturally capture individual-level heterogeneity.

Scalable mediation modeling. In conventional frameworks, each mediator is treated through an independent subequation, leading to excessive complexity when p is large. GAHMN addresses this by employing a shared neural network with a 1×1 convolution layer, allowing all mediators to share input layers while maintaining unique output channels. This design permits parameter sharing, improves scalability, and captures mediator-specific treatment effects using a single CGAN, rather than p separate models.

Improving generalization. For a given treatment $T \in t_0^\star, t_1^\star$, we can only observe samples of the form $(Y, \boldsymbol{X}, \boldsymbol{M}(t_0^\star), t_0^\star)$ or $(Y, \boldsymbol{X}, \boldsymbol{M}(t_1^\star), t_1^\star)$, while the counterfactuals $(Y, \boldsymbol{X}, \boldsymbol{M}(t_1^\star), t_0^\star)$ and $(Y, \boldsymbol{X}, \boldsymbol{M}(t_0^\star), t_1^\star)$ remain unobserved. As a continuous neural network, G_Y must extrapolate to these unseen counterfactuals using its learned mappings over $(\boldsymbol{X}, \boldsymbol{M}, T)$. However, the distances

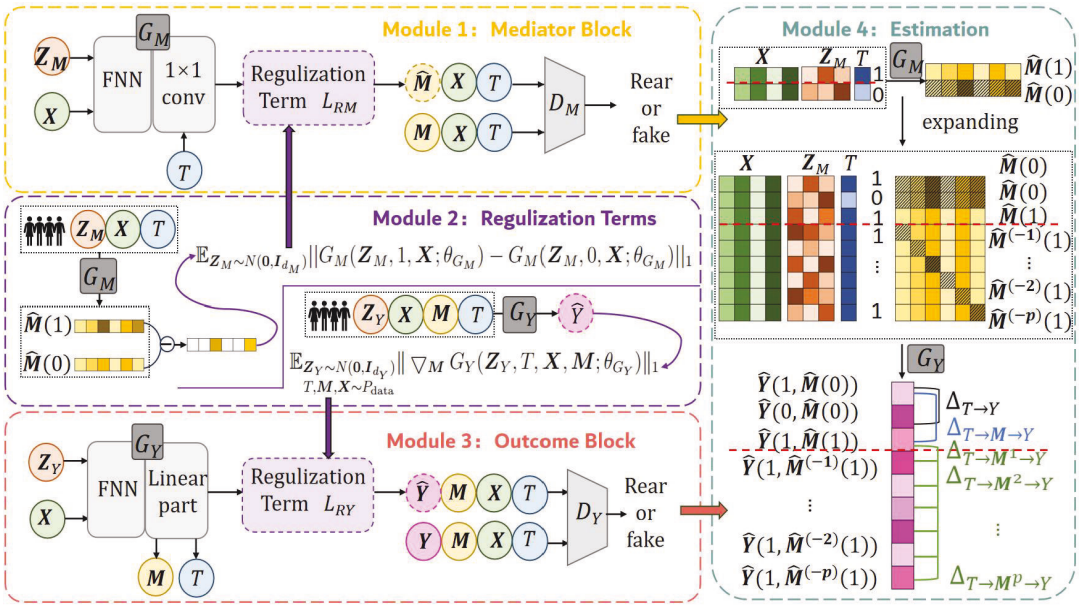

Figure 3: Block Diagram of GAHMN. The Mediator Block and Outcome Block are GANs for generating high-dimensional mediators and outcomes, respectively. The Regularization Block selects relevant mediators via sparsity penalties. The Estimation Block computes mediation effects by generating counterfactual outcomes $\hat{Y}$ using the learned generators G_M and G_Y.

between observed and counterfactual configurations—such as $||(\boldsymbol{M}(t_1^\star), t_0^\star) - (\boldsymbol{M}(t_0^\star), t_0^\star)||$—can be large, especially when $\boldsymbol{M}$ is high-dimensional. This makes extrapolation along the $(\boldsymbol{M}, T)$ axis both difficult and error-prone. To address this, we employ two strategies. First, inspired by Occam's Razor, we design a simpler network architecture by placing $\boldsymbol{M}$ and T as linear components in the shallow layers, while embedding $\boldsymbol{Z}_Y$ and $\boldsymbol{X}$ in deeper layers. This reduces model complexity along the extrapolation path. Second, we impose a regularization term on the partial derivatives of G_Y with respect to $\boldsymbol{M}$, promoting smoothness along the $\boldsymbol{M}$ dimension. These strategies collectively enhance the extrapolation capability and robustness of the outcome generator.

Handling sparsity. Not all mediators contribute to the outcome, nor are all affected by the treatment. To identify the effective ones, GAHMN incorporates sparsity-inducing ℓ_1 regularizations on the gradient of G_M and G_Y. In the partially linear structures, this reduces to a LASSO-like penalty on generators, effectively promoting sparsity. The model can perform automatic variable selection, enhancing interpretability by pinpointing key mediators that transmit treatment effects and yielding meaningful mediation pathways.

Theoretical View

Since both blocks are formulated by a general high-dimensional partially linear framework, we propose a unified approach to establish their convergence results. Let $\boldsymbol{V}$ and $\boldsymbol{X}$ denote the conditioning variables, Y denote the outcome variable. The generating processes for both blocks can be expressed in the following partially linear form.

$$Y_i = \boldsymbol{\beta}^T \boldsymbol{V}_i + g(\boldsymbol{X}_i, \boldsymbol{Z}_i), \tag{13}$$

where $\boldsymbol{Z}_i$ is out-source noise, and g is an unknown nonlinear term. For fitting the process according to our proposed approach, we have a generating function as follows

$$\hat{Y}_i = \hat{G}_{\beta, G_F}(\boldsymbol{V_i}, \boldsymbol{X_i}, \boldsymbol{Z_i}) = \hat{\boldsymbol{\beta}}^T \boldsymbol{V}_i + \hat{G}_F(\boldsymbol{X}_i, \boldsymbol{Z}_i), \tag{14}$$

where $\hat{G}_F$ is an flexible engough FNN. For simplicity, Y is set as one-dimensional.

To be concrete, the estimated $\hat{G}_{\beta, G_F}$ is obtained through

$$\begin{aligned}(\hat{G}_{\boldsymbol{\beta}, G_F}, \hat{D}) &= \min_{G \in \mathcal{G}} \max_{D \in \mathcal{D}} \mathcal{L}_n(G, D) \\ &\equiv \min_{G \in \mathcal{G}} \max_{D \in \mathcal{D}} \lambda \|\boldsymbol{\beta}\|_1 + \frac{1}{n} \sum_{i=1}^{n} \left[\log D\left(Y_i, V_i, X_i\right)\right] \\ &\quad + \frac{1}{n} \sum_{i=1}^{n} \left[\log\left(1 - D\left(G\left(V_i, X_i, Z_i\right), V_i, X_i\right)\right)\right],\end{aligned}$$

where $\mathcal{G}, \mathcal{D}$ are the functional spaces spanned by FNN and $\lambda > 0$ controls penalty strength, bringing sparsity among $\hat{\boldsymbol{\beta}}$.

For (13) and (14), we have following assumptions.

- **(A1)**: The target generative model G^* exists, and its linear form is $G^*(v, x, z) = \boldsymbol{\beta}^{*\top} v + G_F^*(x, z)$, where $\|\beta^*\|_1 < \infty$, and G_F^* is continuous and its L_∞ norm is upper bounded by constant B.
- **(A2)**: $\frac{P_{\boldsymbol{V}, \boldsymbol{X}, Y}}{P_{\boldsymbol{V}, \boldsymbol{X}, Y} + P_{\boldsymbol{V}, \boldsymbol{X}, \hat{Y}}}$ is lower and upper bounded.
- **(A3)**: G_F is a ReLU-based FNN with the whole model size S, depth H, width W. As sample size n goes to infinity, $HW \to \infty$ and $BSH \log S \log n / n \to 0$.
- **(A4)**: D is implemented using a ReLU-based FNN, satisfying the same parameter constraints of G_F.
- **(A5)**: $Y_i | \boldsymbol{V}_i, \boldsymbol{X}_i$ is sub-gaussian.

- **(A6)**: $\boldsymbol{V}$ satisfy Restricted Eigenvalue (RE) condition over S with pararmeters (κ, α): $\frac{1}{n}\|\boldsymbol{V}\boldsymbol{\Delta}\|_2^2 \geq \kappa\|\boldsymbol{\Delta}\|_2^2$ for all $\boldsymbol{\Delta} \in \mathbb{C}_\alpha(S) := \{\boldsymbol{\Delta} \in \mathbb{R}^p | \|\boldsymbol{\Delta}_{S^c}\|_1 \leq \alpha\|\boldsymbol{\Delta}_S\|_1\}$.
- **(A7)**: $\boldsymbol{\beta}^*$ is supoported on a subset $S \subseteq \{1, 2, \ldots, p\}$ with $|S| = s \ll p$.

Let $P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}}$ and $P_{\boldsymbol{V},\boldsymbol{X},Y}$ be the densities of $(\boldsymbol{V}, \boldsymbol{X}, \hat{Y})$ and $(\boldsymbol{V}, \boldsymbol{X}, Y)$, respectively. Before presenting the convergence results, We first present the following two lemmas.

Lemma 1. Let D_{TV} and D_{JS} denote the total variance and JS divergence, respectively. We have

$$D_{TV}^2(P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}}, P_{\boldsymbol{V},\boldsymbol{X},Y}) \lesssim \mathcal{D}_{JS}(P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}} || P_{\boldsymbol{V},\boldsymbol{X},Y}).$$

Lemma 2. Let P and Q be sub-Gaussian with means μ_P and μ_Q, and variances σ_P^2 and σ_Q^2, respectively. $\mathcal{M} = \frac{\sigma_P^2+\sigma_Q^2}{2}$. The JS divergence follows:

$$D_{JS}(P\|Q) \leq \frac{(\mu_P - \mu_Q)^2}{8\sigma_{\mathcal{M}}^2},$$

where $\sigma_{\mathcal{M}}^2 = \frac{\sigma_P^2+\sigma_Q^2}{2}$ is the variance of $\mathcal{M} = \frac{P+Q}{2}$.

Theorem 1. Under (A1)-(A7), if $\lambda = O\Big(\frac{\sqrt{\log p}}{n}\Big)$, then

$$P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}} \to P_{\boldsymbol{V},\boldsymbol{X},Y}.$$

If the linear parameters converge, the mediation effect estimates will also converge. At population level, we attempt to obtain an optimal generator $G^\star$ and its associated $\boldsymbol{\beta}^\star$ that mimimizes $\mathcal{D}_{JS}(P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}} || P_{\boldsymbol{V},\boldsymbol{X},Y})$.

Corollary 1. Let $G^\star$ be the minimimizer of the JS-divergence $\mathcal{D}_{JS}(P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}} || P_{\boldsymbol{V},\boldsymbol{X},Y})$.

$$G^\star \in \arg\min_G \mathcal{D}_{JS}(P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}} || P_{\boldsymbol{V},\boldsymbol{X},Y}),$$

if and only if $P_{\boldsymbol{V},\boldsymbol{X},\hat{Y}} = P_{\boldsymbol{V},\boldsymbol{X},Y}$, which yields $\boldsymbol{\beta}^\star = \boldsymbol{\beta}$.

These theoretical results demonstrate the convergence of our new method. We defer all the proofs to Appendix C.

Synthetic Experiment

The data-generating process is defined as:

$$\begin{cases} M_i^k = a_k T_i + g_k(X_{1,i}, X_{2,i}) + \epsilon_{k,i}, \\ Y_i = 2 - 0.5T_i + \sum_{k=1}^{p} b_k M_i^k \\ \qquad - 2\sqrt{|X_{1,i} + 5|} + \sin(1.5X_{2,i}) + \varepsilon_i, \end{cases} \tag{15}$$

where $k = 1, 2, \cdots, 500$ ($p = 500$) and $i = 1, 2, \cdots, 1000$. For $\forall i$, $T_i \in \{0, 1\}$ with equal probability. $X_{1,i}$ and $X_{2,i}$ are independently drawn from $N(0, 1)$. Let $\boldsymbol{\epsilon}_i = [\epsilon_{1,i}, \epsilon_{2,i}, \cdots, \epsilon_{p,i}]$, and $\boldsymbol{\epsilon_i} \sim N(0, \Sigma^*)$. Σ^* combines autoregressive correlation and heteroskedasticity. Specifically, its (i_0, j_0)-th element is given by $\rho^{|i_0-j_0|} * (0.1 + 0.3|X_{1,i}| + 0.3T_i)$. ε_is follow $N(0, 1)$. $g_k(X_{1,i}, X_{2,i})$ is defined as

$$\begin{aligned} g_k(X_{1,i}, X_{2,i}) \quad = \quad & \delta_{k,1}X_{1,i} + \delta_{k,2}X_{2,i} + \delta_{k,3}X_{1,i}^2 \\ & + \delta_{k,4}X_{2,i}^2 + \delta_{k,5}X_{1,i}X_{2,i}, \end{aligned}$$

	Method	DE	CI	TIE	CI
A	GAHMN	**<u>0.502</u>**	[0.485, 0.517]	**<u>2.720</u>**	[2.565, 2.824]
	HIMA	0.478	[0.153, 0.785]	2.702	[2.417, 2.988]
	PLSEM	0.492	[0.469, 0.515]	2.680	[2.646, 2.714]
	Guo (2023)	0.618	[0.378, 0.828]	2.554	[2.247, 2.857]
B	GAHMN	**<u>0.501</u>**	[0.494, 0.521]	**<u>3.501</u>**	[3.356, 3.624]
	HIMA	0.466	[0.204, 0.764]	2.714	[2.431, 2.985]
	PLSEM	0.491	[0.463, 0.527]	3.459	[3.425, 3.492]
	Guo (2023)	0.533	[0.356, 0.649]	3.455	[3.212, 3.706]

Table 1: Comparison results under different settings. The true values of the direct and total indirect effects are (0.5, 2.8) under Setting A and (0.5, 3.6) under Setting B. The best results are highlighted in bold and underlined.

where $\delta_{k,1}$, $\delta_{k,2}$, $\delta_{k,3}$, $\delta_{k,4}$ and $\delta_{k,5}$ are independently drawn from $U[-1, 1]$. a_ks and b_ks, are configured according to the following two settings:

A. $\boldsymbol{a}_{1:p} = [1.6, 1.2, 0.8, 0.4, 1.6, 0, 0, 0, 0, \ldots, 0]$,
$\boldsymbol{b}_{1:p} = [0, 0, 1, 1, 1, 1, 1, 0, 0, \ldots, 0]$.

B. $\boldsymbol{a}_{1:p} = [1, 1, 1, 1, 1, 0, 0, 0, 0, \ldots, 0]$,
$\boldsymbol{b}_{1:p} = [0, 0, 1.6, 1.2, 0.8, 0.4, 1.6, 0, 0, \ldots, 0]$.

In both settings, the first five elements of $\boldsymbol{a}_{1:p}$ are nonzero, and the 3rd to 7th elements of $\boldsymbol{b}_{1:p}$ are nonezero, which means only the indirect pathways through M^3, M^4 and M^5 contribute to the mediation effects. Experiments under more settings are presented in Appendix D.1.

Baselines We compare our method with three baseline ones. The first is HIMA (Zhang et al. 2016), which assumes a linear structural model and identifies mediators via joint significance testing. The second is Guo et al. (2023), which estimates the total indirect effect using a partially penalized framework, focusing on the aggregate mediation effect rather than individual pathways. The third is PLSEM by Cai et al. (2022), which formulates the mediation problem under a partially linear structural equation model, employing spline regression to capture nonlinearity and adaptive LASSO for variable selection.

Experimental Settings Among the 1,000 generated samples, 800 are randomly selected for training while the remaining 200 are reserved for treatment effect estimation. We set $N_g = 1000$ to compute direct and indirect treatment effects via equations (10)–(12). The 95% CIs are constructed using the empirical 2.5th and 97.5th percentiles across the N_g repetitions. Additional training details, including network architectures and optimization configurations, are provided in Appendix D.2.

Experiment Results Table 1 presents the estimated direct and total indirect treatment effects (DE and TIE) with 95% CI. Table 2 presents the estimated indirect effects by relevant mediators (M^1 to M^7). We highlight the following key findings: First, GAHMN consistently achieves the most accurate point estimates across all the settings. Second, regarding CIs, GAHMN provides tighter intervals that contain the ground truth. Third, GAHMN can effectively identify the

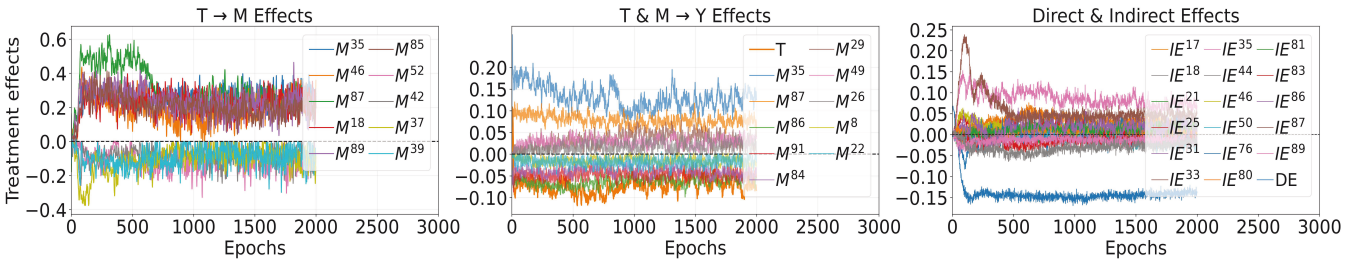

Figure 4: Learning results on real data. The left panel shows the top 10 mediators most affected by treatment T (via G_M), and the middle panel shows the top 10 mediators with the strongest influence on outcome Y (via G_Y). The right panel reports the mediators selected by GAHMN and their estimated direct (DE) and indirect effects ($IE^p = \triangle_{T\to M^p\to Y}$).

	Setting A	Setting B
$\triangle_{T\to M^1\to Y}$	$0.043_{(0)}$ $[0.029, 0.055]$	$0.034_{(0)}$ $[0.030, 0.038]$
$\triangle_{T\to M^2\to Y}$	$0.005_{(0)}$ $[-0.086, 0.043]$	$0.003_{(0)}$ $[-0.007, 0.010]$
$\triangle_{T\to M^3\to Y}$	$0.768_{(0.8)}$ $[0.710, 0.806]$	$1.565_{(1.6)}$ $[1.492, 1.633]$
$\triangle_{T\to M^4\to Y}$	$0.352_{(0.4)}$ $[0.294, 0.394]$	$1.193_{(1.2)}$ $[1.142, 1.246]$
$\triangle_{T\to M^5\to Y}$	$1.628_{(1.6)}$ $[1.564, 1.691]$	$0.769_{(0.8)}$ $[0.730, 0.805]$
$\triangle_{T\to M^6\to Y}$	$-0.040_{(0)}$ $[-0.072, -0.009]$	$-0.012_{(0)}$ $[-0.032, 0.006]$
$\triangle_{T\to M^7\to Y}$	$-0.035_{(0)}$ $[-0.074, 0.002]$	$-0.049_{(0)}$ $[-0.121, 0.019]$

Table 2: Indirect effects and their 95% CI by GAHMN under Settings A and B (true values shown in parentheses).

sparsity of mediation structures. Numerical results on more settings are presented in Appendix D.3.

Empirical Experiment

Macroeconomic shocks exert heterogeneous impacts across industry sectors, reflected in firm-level financial indicators (Guo et al. 2023). We aim to examine whether financial statement metrics mediate the relationship between industry classification and stock returns during the U.S.–China trade conflict. Prior studies (Itakura 2020; Fajgelbaum and Khandelwal 2022) document uneven effects across sectors, with Electronics & ICT, Electrical, and Agriculture particularly affected by demand shocks and tariffs. Using firm-level financial data, we identify key mediators that explain how industry classification influences stock performance.

We collect data from 2,601 A-share listed companies excluding traditional manufacturing from the CSMAR database. The outcome variable (Y) is quarterly stock return from Q1 2018 to Q4 2020. The treatment variable (T) indicates high-tech manufacturing sector membership ($T = 1$ for high-tech, $T = 0$ otherwise), including computer equipment, software, internet services, instrumentation, and pharmaceutical manufacturing. Potential mediators ($\boldsymbol{M}$) comprise 94 financial statement indicators (e.g., debt-to-asset ratio, current ratio). Two covariates ($\boldsymbol{X}$), market capitalization and ownership concentration, are included. The training details are presented in Appendix E.1.

Figures 4 illustrates how G_M and G_Y identify treatment-relevant and outcome-relevant mediators. We notice that estimates converge after approximately 300 epochs, with 17 mediators exhibiting non-zero indirect effects. The estimated effects of these mediators and comparisons in terms of predicting error are provided in Appendix E.2 and E.3. Table 3 summarizes the treatment effect estimates. GAHMN delivers accurate and stable total effect estimates with narrow confidence intervals. Indirect effects are small and statistically insignificant, avoiding overstatement of mediation pathways. In contrast, HIMA and Guo(2023) produce inflated indirect effect estimates due to linear modeling assumptions, while PLSEM exhibits higher variability. GAHMN selects fewer mediators, enhancing sparsity and interpretability in high-dimensional settings.

Method	DE	CI	TIE	CI	# M.
GAHMN	-0.14	[-0.15, -0.13]	0.04	[-0.11, 0.17]	17
HIMA	-0.11	[-0.17, -0.05]	0.43	[0.39, 0.47]	29
PLSEM	-0.13	[-0.19, -0.07]	0.30	[-0.38, 0.98]	32
Guo (2023)	-0.11	[-0.17, -0.06]	0.44	[0.40, 0.48]	N/A

Table 3: Estimated treatment effects for real data. "# M." indicates the number of mediators selected by each method.

Conclusion

In this work, we propose GAHMN, a novel GAN-based HMA method. GAHMN leverages the expressive capacity of generative models to capture complex nonlinearities, heterogeneity, and mediator correlations. By embedding partially linear structures, multi-channel convolutions, and targeted regularization into its architecture, GAHMN integrates domain knowledge and structural priors while maintaining interpretability and scalability. We provide solid theoretical results to prove the convergence of our new method. Extensive numerical experiments are presented to illustrate the superior performance of GAHMN. Our proposed method also offers a promising foundation for applying generative modeling to a broader class of high-dimensional machine learning/statistical problems.

Acknowledgments

This work was supported in part by Humanities and Social Science Fund of Ministry of Education, China under project 23YJC910011, Innovation and Talent Base for Digital Technology and Finance (B21038), General Research Fund grant 14303622 from the Research Grant Council of the Hong Kong. Fundamental Research Funds for the Central Universities under Grant 2722022BQ044, and 2722023EJ002.

References

Cai, X.; Zhu, Y.; Huang, Y.; and Ghosh, D. 2022. High-dimensional causal mediation analysis based on partial linear structural equation models. *Computational Statistics & Data Analysis*, 174: 107501.

Carpena, F.; and Zia, B. 2020. The causal mechanism of financial education: Evidence from mediation analysis. *Journal of Economic Behavior & Organization*, 177: 143–184.

Celli, V. 2022. Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys*, 36(1): 214–234.

Fajgelbaum, P. D.; and Khandelwal, A. K. 2022. The economic impacts of the US–China trade war. *Annual Review of Economics*, 14(1): 205–228.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gui, J.; Sun, Z.; Wen, Y.; Tao, D.; and Ye, J. 2021. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, 35(4): 3313–3332.

Gunzler, D.; Chen, T.; Wu, P.; and Zhang, H. 2013. Introduction to mediation analysis with structural equation modeling. *Shanghai archives of psychiatry*, 25(6): 390.

Guo, X.; Li, R.; Liu, J.; and Zeng, M. 2022. High-dimensional mediation analysis for selecting DNA methylation loci mediating childhood trauma and cortisol stress reactivity. *Journal of the American Statistical Association*, 117(539): 1110–1121.

Guo, X.; Li, R.; Liu, J.; and Zeng, M. 2023. Statistical inference for linear mediation models with high-dimensional mediators and application to studying stock reaction to COVID-19 pandemic. *Journal of Econometrics*, 235(1): 166–179.

Hayes, A. F. 2017. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hu, W.; Chen, S.; Cai, J.; Yang, Y.; Yan, H.; and Chen, F. 2024. High-dimensional mediation analysis for continuous outcome with confounders using overlap weighting method in observational epigenetic study. *BMC Medical Research Methodology*, 24(1): 125.

Imai, K.; Keele, L.; and Tingley, D. 2010. A general approach to causal mediation analysis. *Psychological methods*, 15(4): 309.

Itakura, K. 2020. Evaluating the impact of the US–China trade war. *Asian Economic Policy Review*, 15(1): 77–93.

Jacobucci, R.; Brandmaier, A. M.; and Kievit, R. A. 2019. A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. *Advances in methods and practices in psychological science*, 2(1): 55–76.

Lindquist, M. A. 2012. Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, 107(500): 1297–1309.

Liu, H.; Jin, I. H.; Zhang, Z.; and Yuan, Y. 2021. Social network mediation analysis: A latent space approach. *Psychometrika*, 86(1): 272–298.

Liu, Z.; Shen, J.; Barfield, R.; Schwartz, J.; Baccarelli, A. A.; and Lin, X. 2022. Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *Journal of the American Statistical Association*, 117(537): 67–81.

MacKinnon, D. 2012. *Introduction to statistical mediation analysis*. Routledge.

Nath, T.; Caffo, B.; Wager, T.; and Lindquist, M. A. 2023. A machine learning based approach towards high-dimensional mediation analysis. *NeuroImage*, 268: 119843.

Parikh, N.; Boyd, S.; et al. 2014. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3): 127–239.

Preacher, K. J.; and Hayes, A. F. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, 40(3): 879–891.

Sun, R.; and Song, X. 2024. Heterogeneous Mediation Analysis for Cox Proportional Hazards Model With Multiple Mediators. *Statistics in Medicine*, 43(29): 5497–5512.

Valente, M. J.; Rijnhart, J. J.; Smyth, H. L.; Muniz, F. B.; and MacKinnon, D. P. 2020. Causal mediation programs in R, M plus, SAS, SPSS, and Stata. *Structural equation modeling: a multidisciplinary journal*, 27(6): 975–984.

Wang, S.; and Huang, Y. 2024. DP2LM: leveraging deep learning approach for estimation and hypothesis testing on mediation effects with high-dimensional mediators and complex confounders. *Biostatistics*, 25(3): 818–832.

Wei, K.; Liu, Y.; Huang, C.; Lin, R.; Yu, Y.; and Qin, G. 2025. Debiased machine learning for ultra-high dimensional mediation analysis. *Bioinformatics*, 41(6): btaf282.

Xie, X.; Zou, H.; and Qi, G. 2018. Knowledge absorptive capacity and innovation performance in high-tech companies: A multi-mediating analysis. *Journal of business research*, 88: 289–297.

Xu, S.; Liu, L.; and Liu, Z. 2022. DeepMed: Semiparametric causal mediation analysis with debiased deep learning. *Advances in Neural Information Processing Systems*, 35: 28238–28251.

Yang, H.; Liu, Z.; Wang, R.; Lai, E.-Y.; Schwartz, J.; Baccarelli, A. A.; Huang, Y.-T.; and Lin, X. 2024. Causal Mediation Analysis for Integrating Exposure, Genomic, and Phenotype Data. *Annual Review of Statistics and Its Application*, 12.

Yuan, Y.; and MacKinnon, D. P. 2009. Bayesian mediation analysis. *Psychological methods*, 14(4): 301.

Yuan, Y.; and Qu, A. 2024. De-confounding causal inference using latent multiple-mediator pathways. *Journal of the American Statistical Association*, 119(547): 2051–2065.

Zeng, P.; Shao, Z.; and Zhou, X. 2021. Statistical methods for mediation analysis in the era of high-throughput genomics: current successes and future challenges. *Computational and structural biotechnology journal*, 19: 3209–3224.

Zhang, H.; Hou, L.; and Liu, L. 2021. A review of high-dimensional mediation analyses in DNA methylation studies. *Epigenome-wide association studies: methods and protocols*, 123–135.

Zhang, H.; Zheng, Y.; Hou, L.; Zheng, C.; and Liu, L. 2021. Mediation analysis for survival data with high-dimensional mediators. *Bioinformatics*, 37(21): 3815–3821.

Zhang, H.; Zheng, Y.; Zhang, Z.; Gao, T.; Joyce, B.; Yoon, G.; Zhang, W.; Schwartz, J.; Just, A.; Colicino, E.; et al. 2016. Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics*, 32(20): 3150–3154.

Zhang, Q.; Yang, Z.; and Yang, J. 2022. MedDiC: high dimensional mediation analysis via difference in coefficients. *bioRxiv*, 2022–09.

Zhou, X.; and Song, X. 2021. Mediation analysis for mixture Cox proportional hazards cure models. *Statistical Methods in Medical Research*, 30(6): 1554–1572.