

## Statistica Sinica Preprint No: SS-2023-0315

<b>Title</b>	Identifiability and Estimation of Causal Effects with Non-Gaussianity and Auxiliary Covariates
<b>Manuscript ID</b>	SS-2023-0315
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202023.0315
<b>Complete List of Authors</b>	Kang Shuai, Shanshan Luo, Yue Zhang, Feng Xie and Yangbo He
<b>Corresponding Authors</b>	Yangbo He
<b>E-mails</b>	heyb@math.pku.edu.cn
Notice: Accepted version subject to English editing.	

## Identifiability and estimation of causal effects with non-Gaussianity and auxiliary covariates

Kang Shuai<sup>a</sup>, Shanshan Luo<sup>b</sup>, Yue Zhang<sup>a</sup>, Feng Xie<sup>b</sup> and Yangbo He<sup>a\*</sup>

<sup>a</sup>Peking University, <sup>b</sup>Beijing Technology and Business University

*Abstract: Assessing causal effects in the presence of unmeasured confounding is challenging. Although auxiliary variables, such as instrumental variables, are commonly used to identify causal effects, they are often unavailable in practice due to stringent and untestable conditions. To address this issue, previous researches have utilized linear structural equation models to show that the causal effect is identifiable when noise variables of the treatment and outcome are both non-Gaussian. In this paper, we investigate the problem of identifying the causal effect using the auxiliary covariate and non-Gaussianity from the treatment. Our key idea is to characterize the impact of unmeasured confounders using an observed covariate, assuming they are all Gaussian. We demonstrate that the causal effect can be identified using a measured covariate, and then extend the identification results to the multi-treatment setting. We further develop a simple estimation procedure for estimating causal effects and derive a  $\sqrt{n}$ -consistent estimator. Finally, we evaluate the performance of our estimator through simulation studies and apply our method to investigate the effect of the trade on income.*

---

\*correspondence to: heyb@math.pku.edu.cn

*Key words and phrases:* Auxiliary variable, Causal effects, Identification, Multiple treatments, Non-Gaussianity.

## 1. Introduction

Identifying and estimating the causal effect of a treatment on an outcome is crucial in practice, as it provides insight into the effectiveness of a given intervention. However, the existence of unmeasured confounding may be a core issue in observational studies. A common assumption of causal inference using observational data is exchangeability, which requires that one has measured a sufficiently rich set of covariates (Rosenbaum and Rubin, 1983). This is often challenged because investigators usually cannot accurately learn the confounding mechanism from the measured covariates in real scenarios.

Numerous strategies have been proposed for addressing identification issue of causal effects under unmeasured confounding. With a valid instrumental variable (IV) that satisfies relevance, independence, and exclusion restriction assumption, the causal effect can be identified for a binary treatment (Angrist et al., 1996) or a continuous treatment (Kasy, 2014) under certain conditions. Proxy variables can also be used to identify the causal effect, and certain types of proxy variables, known as negative control vari-

---

ables, are leveraged for mitigating confounding bias in observational studies (Lipsitch et al., 2010; Kuroki and Pearl, 2014; Miao et al., 2018; Tchetgen et al., 2020). Among them, Miao et al. (2018) provides sufficient conditions for nonparametric identification of the causal effect using double negative control variables. For an overview of proxy variable based identification strategies, we refer readers to Tchetgen et al. (2020).

However, auxiliary variables like instrumental variables or proxy variables, may be unavailable in real scenarios because of stringent assumptions. To address this, many researchers focus on employing valid instrumental variables from a candidate instrument set (Kang et al., 2016; Guo et al., 2018; Windmeijer et al., 2019) or imposing additional nonlinear model assumptions when only an invalid IV is available (Liu et al., 2022; Sun et al., 2022). But the identification issue without a valid IV cannot be properly tackled even under the linear model setting. In comparison, with linear non-Gaussian model structure, identifying causal order or causal effects is still possible in the absence of auxiliary variables. The work of Shimizu et al. (2006) performs LiNGAM (short for Linear Non-Gaussian Acyclic Model) analysis and shows the causal structure is identifiable using independent component analysis (Stone, 2002). Then the paper from Hoyer et al. (2008) further slightly relax the non-Gaussian assumptions of LiNGAM. Follow-

ing this, Hoyer et al. (2008) extends LiNGAM to linear non-Gaussian settings with hidden variables and shows that the causal order and the causal effect are both identifiable. Following their promising work, many researchers have accomplished work for causal discovery problems (Shimizu et al., 2009, 2011; Entner and Hoyer, 2011; Chen and Chan, 2013; Salehkaleybar et al., 2020; Wang and Drton, 2020). These articles assume non-Gaussianity, where all noise variables are non-Gaussian, except for Chen and Chan (2013), who considers the Gaussian confounder setting. Meanwhile, the use of non-Gaussianity to learn causal effects under hidden endogeneity is a common approach in econometrics. For example, Cattaneo et al. (2012) investigates the optimal inference problem with an IV when error variables are non-Gaussian. The research from Park and Gupta (2012) proposed that if the outcome error is Gaussian and the treatment is non-Gaussian, then the identification can be achieved by modeling the joint distribution of the endogenous regressor and the error term as a Gaussian copula. All of these researches reveal the potential of non-Gaussianity for identifying causal effects.

Regarding multiple treatments, identification issue becomes more difficult. The research of Wang and Blei (2019) proposes a deconfounder method for identifying multiple treatment effects. They establish a factor

model for treatments by assuming treatments are conditionally independent given confounders and adjust confounding bias using the estimated confounder. Following their work, various studies, comments, and discussions have emerged in the literature regarding multiple treatments (D'Amour, 2019; Imai and Jiang, 2019; Miao et al., 2023; Kong et al., 2022). They consider identifying effects of multiple causally independent treatments using a factor model and have highlighted the necessity of using auxiliary variables or non-Gaussianity for identification. For instance, Kong et al. (2022) provides identification of causal effects for a binary outcome using non-Gaussianity from an auxiliary latent variable in the binary choice outcome model. However, identifying the causal effect for a continuous outcome, especially when treatments can be causally correlated, remains an unsolved challenge.

In this paper, we establish the identification results of the causal effect using a measured covariate and the non-Gaussianity from the treatment. This provides a new perspective for identifying the causal effect, which suggests that finding some Gaussian covariate instead of a valid IV in a candidate instrument set may also work. Additionally, we show that our proposed method can be extended to handle scenarios involving multiple treatments, among which the minimum number of required Gaussian co-

---

variates equals the smaller dimension between the dimension of treatments and the dimension of unmeasured confounders. We finally develop a simple estimation procedure for calculating causal effects using our strategy and derive a  $\sqrt{n}$ -consistent estimator. The simulation results show that our estimator can provide favorable performance. We finally apply our method to study the effect of the trade on the income, revealing the similar conclusion as the previous researches using a valid IV.

The remainder sections are organized as follows. Section 2 provides the framework of our setting, including model and identification results in Section 2.1, and the extension to multi-treatment scenario in Section 2.2. In Section 3, we present a simple estimation process and the  $\sqrt{n}$ -consistent estimator. We evaluate the finite-sample performance of the proposed estimator in Section 4. Then, in Section 5, we demonstrate the proposed strategy using the cross-sectional data for the year 2019 from the World Bank. Section 6 concludes the paper and provides several notable future research topics. The appendix presents the proof for Theorem 1. Additional proofs and simulation results can be found in the supplementary material.

---

## 2. Framework

### 2.1 Model and Identification

Firstly, we consider the single treatment case. Denote  $A \in \mathbb{R}^1$  as a treatment,  $Y \in \mathbb{R}^1$  an outcome,  $Z \in \mathbb{R}^1$  an observed covariate,  $U \in \mathbb{R}^t$  the vector of unmeasured confounders. For simplicity, we consider a single measured covariate, but the same conclusion holds by conditioning on other measured covariates implicitly. Without loss of generality, we assume  $U, Z$  both have zero mean and unit variance. We propose the following model

$$A = \gamma Z + \lambda^T U + \varepsilon_A, \quad (2.1a)$$

$$Y = \alpha A + \beta Z + s^T U + \varepsilon_Y, \quad (2.1b)$$

where  $(Z, U)$  are Gaussian with  $\text{cov}(Z, U) = \xi^T$ ,  $\varepsilon_A$  is non-Gaussian with non-zero variance, and  $\varepsilon_A \perp\!\!\!\perp (Z, U)$ . The noise term  $\varepsilon_Y$  satisfies  $E(\varepsilon_Y | A) = 0$ . Note that we do not require any prior knowledge about the distribution of  $\varepsilon_Y$ , so whether  $\varepsilon_Y$  follows Gaussian distribution does not change the following conclusion, which differs from the requirement of Hoyer et al. (2008) or Park and Gupta (2012). Also, the Gaussianity of  $Z$  can be directly tested and if we admit the Gaussianity of  $U$  and the linear



## 2.1 Model and Identification

---

model assumption, then the non-Gaussianity of  $\varepsilon_A$  is equivalent to the non-Gaussianity of treatment  $A$ .

Figure 1 provides a graphical illustration of our model. According to the potential outcome framework (Rubin, 2005),  $Y_a$  is denoted as the outcome for an individual under a potential treatment  $A = a$ . We maintain the causal consistency assumption throughout the paper, implying no interference between units and there is only one version of the potential outcome (Rubin, 1980). The average causal effect of  $A$  on  $Y$  for a unit increase of  $A$  can be defined as follows

$$\text{ACE}_{A \rightarrow Y} = E(Y_{a+1} - Y_a) = \alpha.$$

The main task is to identify the causal parameter  $\alpha$  from observational data of  $(Z, A, Y)$ . We summarize the identification results in the following theorem.

**Theorem 1.** Under model (2.1a)-(2.1b) and assuming  $\text{cov}(Z, A) \neq 0$ , the effect of  $A$  on  $Y$ , namely  $\alpha$ , is identifiable.

The above theorem establishes the identifiability of the causal effect  $\alpha$  when the known source of non-Gaussianity arises from  $\varepsilon_A$ . As previously demonstrated, this conclusion holds regardless of the distribution of  $\varepsilon_Y$ . Un-

## 2.1 Model and Identification

---

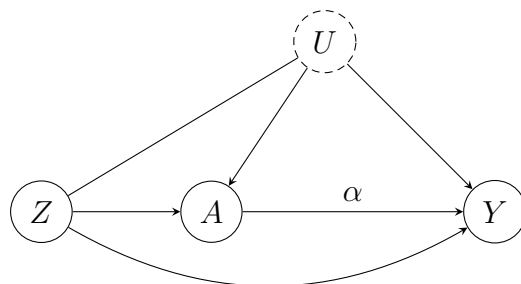


Figure 1: Causal diagram with an observed covariate  $Z$ , a vector of unmeasured confounders  $U$ , a treatment  $A$  and an outcome  $Y$ .

like LiNGAM analysis assuming all noise variables to be non-Gaussian, focusing on the identification of causal order, we only require non-Gaussianity from  $\varepsilon_A$  for identifying a causal effect  $\alpha$ , which also differs from Park and Gupta (2012) assuming outcome error to be Gaussian.

The non-Gaussianity of  $\varepsilon_A$  ensures that  $E(Z | A)$  is a nonlinear function of  $A$ . Furthermore, the Gaussianity of  $(Z, U)$  implies that  $E(U | A)$  is proportional to  $E(Z | A)$ , which enables us to characterize the impact of  $U$  with  $Z$  when conditioning on  $A$ . Intuitively, the Gaussianity helps us separate the effect of  $A$  on  $Y$  from the effect of  $(Z, U)$  on  $Y$ . Moreover, the non-Gaussianity of  $\varepsilon_A$  makes it possible to identify  $\alpha$  through the linear independence of  $A$  and  $E(Z | A)$ . More generally, we can also include the interaction terms and some observed non-Gaussian covariates into consideration. This can be summarized in the following two corollaries.

**Corollary 1.** Replacing the model (2.1b) by  $Y = \alpha A + \beta Z + s^T U + rZA +$

## 2.1 Model and Identification

---

$w^T U A + \varepsilon_Y$  and also assuming (i)  $\text{cov}(Z, A) \neq 0$ , (ii)  $(Z, U)$  are Gaussian, (iii)  $\varepsilon_A$  is non-Gaussian with non-zero variance, and (iv)  $(Z, U, \varepsilon_A, \varepsilon_Y)$  have zero mean, then the effect of  $A$  on  $Y$ , namely  $\alpha$ , is identifiable.

**Corollary 2.** Replacing the models (2.1a)-(2.1b) by the following models with independent observed non-Gaussian covariates  $\tilde{Z} \perp\!\!\!\perp (Z, U)$

$$A = \gamma Z + \tilde{\gamma}^T \tilde{Z} + \lambda^T U + \varepsilon_A, \quad (2.2a)$$

$$Y = \alpha A + \beta Z + \tilde{\beta}^T \tilde{Z} + s^T U + \varepsilon_Y, \quad (2.2b)$$

and also requiring the same distributional assumptions as Corollary 1, then the effect of  $A$  on  $Y$ , namely  $\alpha$ , is identifiable.

Note that we center  $(Z, U, \varepsilon_A, \varepsilon_Y)$  to be of zero mean in order to ensure  $\text{ACE}_{A \rightarrow Y} = E(Y_{a+1} - Y_a) = \alpha$ , so  $\alpha$  still represents the corresponding causal effect. Also, interaction terms and observed non-Gaussian covariates can be simultaneously taken into consideration as long as we combine the two corollaries. Throughout the following parts of the paper, we focus on models (2.1a)-(2.1b) for better illustration. Actually, our identification result provides a new perspective for identifying causal effects in the presence of unmeasured confounding. For example, with a candidate instrument set, Kang et al. (2016) establishes the identification result of average causal ef-

## 2.1 Model and Identification

---

fect with a candidate instrument set by assuming that at least 50% of the instruments are valid. However, our identification results demonstrate that the causal effect  $\alpha$  can be identified using an observed Gaussian covariate under the assumptions that  $\varepsilon_A$  is non-Gaussian and  $U$  is Gaussian. This is sufficient for identification, even when all candidate instruments are invalid. In such a situation, an alternative is to perform normality tests in the candidate instrument set instead of verifying stringent and untestable conditions of instrumental variable. Furthermore, the measured covariate  $Z$  can serve as an invalid negative control variable, eliminating the need to find two valid negative control variables. The detailed proof of Theorem 1 can be found in the Appendix. The following example provides a specific illustration of Theorem 1.

**Example 1.** Suppose  $Z, U, \varepsilon_Y$  are three independent random variables following standard normal distribution,  $\varepsilon_A \sim U(-1, 1)$  and  $\gamma = \lambda = 1$ , we know that

$$E(Y | A) = \alpha A + \beta E(Z | A) + sE(U | A) = \alpha A + \tilde{\beta}E(Z | A),$$

where  $\tilde{\beta}$  is some constant because  $E(U | A)$  is proportional to  $E(Z | A)$  according to the proof of Theorem 1. We show the contradiction if  $E(Z | A)$

is linear in  $A$ :

$$E(Z | A) = kA \Rightarrow E\{(Z - kA)f(A)\} = 0, \forall f(\cdot),$$

which implies  $k = E\{Zf(A)\}/E\{Af(A)\}$ . So we have

$$f(A) = A \Rightarrow k = \frac{E(ZA)}{E(A^2)} = \frac{3}{7}, \quad f(A) = A^3 \Rightarrow k = \frac{E(ZA^3)}{E(A^4)} = \frac{35}{81},$$

where  $k$  takes different values for different  $f(\cdot)$ . This is a contradiction, implying that  $E(Z | A)$  cannot be linear in  $A$ . Hence,  $A$  and  $E(Z | A)$  must be linearly independent, which implies that causal parameter  $\alpha$  will be identifiable.

The above example illustrates the implication of non-Gaussianity from  $\epsilon_A$ . Such kind of non-Gaussianity allows us to distinguish  $A$  from  $E(Z | A)$ . The similar idea can also be naturally extended to the nonlinear model setting, as illustrated in the following corollary.

**Corollary 3.** *Replace the outcome model in (2.1b) by  $Y = f(A) + \beta Z + s^T U + \epsilon_Y$ , then  $f(\cdot)$  is identifiable if  $(\epsilon_A, Z, U)$  is Gaussian and either one of the following conditions holds*

- (i)  $\lim_{a \rightarrow \infty} f(a)/|a| = 0$ ;

## 2.1 Model and Identification

---

(ii)  $f(A)$  does not include the linear term of  $A$ .

When we replace model (2.1b) by  $Y = f(A) + \beta Z + s^T U + \epsilon_Y$ , the main task becomes separating  $f(A)$  from  $E(Z | A)$ . A natural choice for simplifying the discussion is by assuming that  $\epsilon_A, Z$  and  $U$  are Gaussian. In this case,  $E(Z | A)$  is linear in  $A$  and hence, we only need to separate  $f(A)$  from  $A$ . The first condition in the above corollary is the so-called sublinear growth condition derived from the recent nonlinear causal discovery research (Li et al., 2023). This implies that causal effects are still identifiable under Gaussian graphical modeling framework with nonlinear causation. Intuitively, the philosophy of separating  $f(A)$  from  $A$  is consistent with that in Theorem 1 by separating  $A$  from  $E(Z | A)$ , as demonstrated in the following example.

**Example 2.** Assume  $Z, U, \epsilon_Y, \epsilon_A$  are four independent random variables following standard normal distribution in model (2.1a)-(2.1b), we replace the linear outcome model by the following nonlinear additive model

$$Y = f(A) + \beta Z + sU + \epsilon_Y.$$

Taking conditional expectation with respect to  $A$  gives

$$E(Y | A) = f(A) + \beta E(Z | A) + sE(U | A) = f(A) + \psi A,$$

where  $\psi$  is some constant. The main task is to identify  $f(\cdot)$ , which requires an additional assumption on  $f(\cdot)$ , e.g., the sublinear growth assumption from (Li et al., 2023). This assumption means  $\lim_{a \rightarrow \infty} f(a)/|a| = 0$  and  $\psi$  can be intuitively regarded as the slope in outcome model when we see the tail of  $A$ . Once  $\psi$  is identified,  $f(\cdot)$  can be immediately identified and estimated nonparametrically. If the Taylor expansion of  $f(A)$  does not include the linear term of  $A$ , it is still possible to identify  $f(\cdot)$ .

This example implies that generalizing to a nonlinear outcome model can also be achieved. However, for the sake of simplicity, we restrict our attention to the linear setting. Next, we extend our identification results to the multi-treatment setting.

## 2.2 Extensions

We now extend our identification results from the single treatment case to the multiple treatments setting. Here, we omit the discussion with interaction terms and some observed non-Gaussian covariates because the

similar conclusions, like those in Corollaries 1 and 2, can be easily derived for multi-treatment cases. Let  $A \in \mathbb{R}^p$ ,  $Z \in \mathbb{R}^l$  and  $U \in \mathbb{R}^t$ , we consider the following linear model

$$A = \Gamma Z + \Lambda U + \varepsilon_A, \quad (2.3a)$$

$$Y = \alpha^T A + \beta^T Z + s^T U + \varepsilon_Y, \quad (2.3b)$$

where  $\Gamma \in \mathbb{R}^{p \times l}$ ,  $\Lambda \in \mathbb{R}^{p \times t}$ , and  $(Z, U)$  are jointly Gaussian with  $\text{var}(Z) = I_l$ ,  $\text{var}(U) = I_t$ , and  $\text{cov}(Z, U) = \Sigma$ . We summarize our identification results for multiple treatments in the following theorem.

**Theorem 2.** Under model (2.3a)-(2.3b), the direct effect of each  $A_j$  on  $Y$ , namely  $\alpha_j$ , is identifiable if the following conditions hold

- (i)  $A$  and  $E(Z | A)$  are linearly independent;
- (ii)  $\text{rank}(\Gamma + \Lambda \Sigma^T, \Gamma \Sigma + \Lambda) = \text{rank}(\Gamma + \Lambda \Sigma^T)$ ,

where  $\text{rank}(\Sigma_1, \Sigma_2)$  represents the rank of the matrix consisting of all columns of  $\Sigma_1$  and  $\Sigma_2$ .

The proof of Theorem 2 can be found in the supplementary material. This theorem provides sufficient conditions for identifying causal effects of multiple treatments simultaneously. In practice, the conditions outlined by this theorem are relatively mild.



- Condition (i) actually requires  $E(Z | A)$  to be nonlinear in  $A$  in the single treatment case with  $Z \in \mathbb{R}^1$ , which is equivalent to non-Gaussianity of  $\varepsilon_A$ . Moreover, Condition (i) can be tested using observational data of  $(Z, A)$ .
- Condition (ii) is sufficient to guarantee that  $E(U | A)$  can be expressed as the linear combination of  $E(Z | A)$ , which is equivalent to requiring  $\text{cov}(Z, A) \neq 0$  in the single treatment case. More generally, condition (ii) implies that the column space of  $\Gamma\Sigma + \Lambda$  is subsumed in the column space of  $\Gamma + \Lambda\Sigma^T$ , which demonstrates that the equation  $\text{cov}(U, A) = \Phi \cdot \text{cov}(Z, A)$  has a solution for  $\Phi$ . Condition (ii) holds if  $\text{cov}(A, Z) = \Gamma + \Lambda\Sigma^T$  is of full row rank, which can be verified using observational data of  $(Z, A)$ .

Therefore, by finding a proper vector of Gaussian covariates  $Z$  that satisfies both the linear independence condition (i) and the rank condition (ii), we can identify the multiple treatment effects  $\alpha$ .

How many observed Gaussian covariates do we need for identification of  $\alpha$ ? We regard  $Z$  as some randomized variable for convenience, implying that  $Z$  and  $U$  are independent with  $\Sigma = 0$ . In this case, condition (ii) reduces to  $\text{rank}(\Gamma, \Lambda) = \text{rank}(\Gamma)$ . Consider the special case  $\text{rank}(\Lambda) = \min(p, t)$ , then the dimension of  $Z$  should be at least  $\min(p, t)$ . This demonstrates that the

minimum number of required Gaussian covariates equals the smaller dimension between the dimension of treatments and the dimension of unmeasured confounders. Especially in the single treatment case, one observed Gaussian covariate is sufficient for identifying  $\alpha$ . Under the scalar unmeasured confounder setting with  $t = 1$ , one-dimensional  $Z$  will also be enough for identification when  $\Gamma$  is proportional to  $\Lambda$ .

Unlike the multiple treatments setting considered in Wang and Blei (2019), we do not assume that treatments are conditionally independent given confounders. Instead, we utilize a linear model with distributional assumptions that permits causal relationships among treatments, even if the causal structure of  $A$  is unknown. In the context of multiple causally independent treatments, Kong et al. (2022) provides identification results for binary outcome by assuming a linear Gaussian model for treatments under scalar unmeasured confounder setting. They require an additional auxiliary latent variable to be non-Gaussian in the binary choice outcome model. Our identification results serve as a complement of Kong et al. (2022) for identifying effects of multiple treatments with a continuous outcome, while not restricting the treatment model to have a factor structure and not assuming the dimension of unmeasured confounder to be one. The following example gives an illustration of Theorem 2 when effects of two

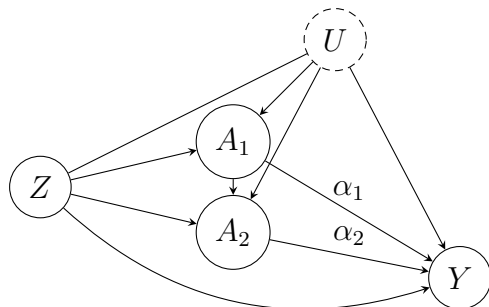


Figure 2: Causal diagram with observed covariates  $Z = (Z_1, Z_2)$ , unmeasured confounders  $U$ , two treatments  $(A_1, A_2)$  and an outcome  $Y$ .

causally correlated treatments are simultaneously of interest.

**Example 3.** Suppose we have two treatments  $A_1, A_2$ , and four independent random variables  $U, \varepsilon_Y$ , and  $Z = (Z_1, Z_2)$ , all following standard normal distributions. Let  $\varepsilon_{A_1} \sim U(-1, 1)$  and  $\varepsilon_{A_2} \sim U(-2, 2)$ . The corresponding model is as follows (Figure 2 gives a graphical illustration)

$$A_1 = Z_1 + U + \varepsilon_{A_1},$$

$$A_2 = A_1 + Z_2 + U + \varepsilon_{A_2},$$

$$Y = \alpha_1 A_1 + \alpha_2 A_2 + \beta_1 Z_1 + \beta_2 Z_2 + sU + \varepsilon_Y,$$

where the covariance matrix  $\text{cov}(A, Z)$  is of full row rank with  $\text{cov}(A_1, Z_1) = \text{cov}(A_2, Z_1) = \text{cov}(A_2, Z_2) = 1$  and  $\text{cov}(A_1, Z_2) = 0$ , so condition (ii) of Theorem 2 is satisfied. To verify condition (i), we need to prove if there

exists  $x_1, x_2, x_3, x_4$  subject to

$$x_1 A_1 + x_2 A_2 + x_3 E(Z_1 | A_1, A_2) + x_4 E(Z_2 | A_1, A_2) = 0,$$

then  $x_1 = x_2 = x_3 = x_4 = 0$ . This immediately implies

$$E\{(x_1 A_1 + x_2 A_2 + x_3 Z_1 + x_4 Z_2) \cdot g(A_1, A_2)\} = 0, \quad \forall g(\cdot).$$

We take  $g(A_1, A_2) \in \{A_1, A_2, A_1^3, A_2^3\}$  and obtain the following linear equations

$$\begin{pmatrix} 7/3 & 10/3 & 1 & 0 \\ 10/3 & 31/3 & 1 & 1 \\ 121/5 & 131/5 & 7 & 0 \\ 516/5 & 1197/5 & 31 & 31 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 0,$$

which implies that the unique solution is  $x_1 = x_2 = x_3 = x_4 = 0$  because the determinant of coefficient matrix is  $-636$ , and both conditions of Theorem 2 are satisfied. Thus, the effects of  $A = (A_1, A_2)^T$  on  $Y$ , namely  $\alpha = (\alpha_1, \alpha_2)$ , are identifiable.

The above example demonstrates the implications of our identification results for two causally correlated treatments. Unlike existing methods, our approach does not require the treatment model to have a factor structure

where treatments are conditionally independent given confounders. Moreover, our results can be viewed as an extension of Kong et al. (2022) from binary outcome to continuous outcome using non-Gaussianity. The measured covariates  $Z$  can be treated as invalid instruments or proxy variables, and their distributional assumptions can be verified empirically. While extending our results to a nonlinear outcome model for multiple treatments is feasible, we omit it here for simplicity.

### 3. Estimation and Asymptotics

Before proceeding with the estimation procedure, it is essential to check all the required conditions using the observational data of  $(Z, A, Y)$ . For instance, we may use the Anderson-Darling test (Anderson and Darling, 1952) to assess the Gaussianity of  $Z$  and the non-Gaussianity of  $A$ . Additionally, depending on the number of treatments, we may need to verify the conditions in either Theorem 1 or Theorem 2. In the single treatment case, verifying  $\text{cov}(A, Z) \neq 0$  suffices. The only untestable assumption is the Gaussianity of the unmeasured confounders  $U$ .

After obtaining all the testable analysis results, we can construct a two-step estimator. In the first step, we estimate the conditional expectation  $E(Z | A)$  using either a parametric working model or standard nonparamet-

---

ric regression methods. Although parametric methods such as maximum likelihood are suitable for estimation, they are prone to model misspecification. Alternatively, nonparametric methods or machine learning techniques, such as random forest and multi-layer perceptron, have gained wide popularity because of their robustness to model misspecification (Bickel et al., 1993; Chernozhukov et al., 2022).

After obtaining the estimator  $\hat{E}(Z | A)$  for  $E(Z | A)$ , the second step involves directly regressing  $Y$  on  $A$  and  $\hat{E}(Z | A)$  to construct a consistent estimator  $\hat{\alpha}$  for  $\alpha$ . The estimation performance of  $\hat{\alpha}$  heavily relies on the precision of the conditional expectation estimator. The final estimator of  $\alpha$  is given by the formula

$$\hat{\alpha} = \hat{X}_{11} \left( \frac{1}{n} \sum_{i=1}^n A_i Y_i \right) + \hat{X}_{12} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{E}(Z | A_i) Y_i \right\},$$

where

$$\hat{X} = \begin{pmatrix} \hat{X}_{11} & \hat{X}_{12} \\ \hat{X}_{12}^T & \hat{X}_{22} \end{pmatrix} = \left\{ \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} A_i \\ \hat{E}(Z | A_i) \end{pmatrix} \begin{pmatrix} A_i \\ \hat{E}(Z | A_i) \end{pmatrix}^T \right\}^{-1}.$$

The estimation process is summarized in the following procedure called EUNC (short for Estimation Using Non-Gaussianity and Covariates).

**Input:**  $n$  i.i.d. data  $\{Z_i, A_i, Y_i\}_{i=1}^n$

**Output:** Estimator  $\hat{\alpha}$  for effects of  $A$  on  $Y$

```
1: procedure EUNC(input)
2:   Center the data of  $(Z, A, Y)$  to be of zero mean.
3:   Test  $H_0$  :
       $Z$  is Gaussian,  $A$  is non-Gaussian and  $\text{cov}(A, Z)$  is of full row rank.
4:   if  $H_0$  is true then
5:     Use nonparametric methods to obtain estimator  $\hat{E}(Z | A)$ .
6:     if  $A, \hat{E}(Z | A)$  are linearly correlated
7:       return FAIL
8:     end if
9:     Regress  $Y$  on  $A, \hat{E}(Z | A)$ 
10:  else return FAIL
11:  end if
12:  return the coefficient in front of  $A$  as  $\hat{\alpha}$ 
13: end procedure
```

The EUNC procedure can also accommodate the situations with additional interaction terms like Corollary 1 or some observed non-Gaussian covariates like Corollary 2. If interaction terms  $AU^T$  or  $AZ^T$  is present, we can include  $ZE(Z | A)^T$  in the EUNC procedure. If some observed non-Gaussian

---

covariates  $\tilde{Z}$  is present, we should regress  $Y$  on  $A, \tilde{Z}, \hat{E}(Z | A, \tilde{Z})$  rather than  $A, \hat{E}(Z | A)$  at the final step of the EUNC procedure. The estimator obtained from the above procedure is  $\sqrt{n}$  consistent, which is guaranteed when the convergence rate of  $\hat{E}(Z | A)$  is at least  $n^{-1/4}$ . This convergence rate can be achieved by a correctly specified parametric model (Hansen, 1982) or by using existing nonparametric techniques (Chernozhukov et al., 2022). We summarize this in the following theorem, and the proof is given in the supplementary material.

**Theorem 3.** Suppose we obtain a uniformly consistent estimator  $\hat{E}(Z | A)$  satisfying

$$\sup_{A \in \mathcal{A}} |\hat{E}(Z | A) - E(Z | A)| = o_p(n^{-1/4}),$$

where  $\mathcal{A}$  is the support of  $A$  and  $n$  is the sample size, then  $\hat{\alpha}$  is root- $n$  consistent for  $\alpha$  in the following sense

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \Sigma_\alpha),$$



---

where

$$\begin{aligned}\Sigma_\alpha &= \text{var}\left\{X_{11}AY + X_{12}E(Z | A)Y + \frac{\partial\alpha(\mu)}{\partial\mu}\xi\right\}, \\ \xi &= \text{vec}\{AA^T, E(Z | A)A^T, E(Z | A)E(Z | A)^T\}, \quad \mu = E(\xi), \\ \alpha(\mu) &= X_{11}E(AY) + X_{12}E\{E(Z | A)Y\},\end{aligned}$$

and

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{pmatrix} = E \left\{ \begin{pmatrix} A \\ E(Z | A) \end{pmatrix} \begin{pmatrix} A \\ E(Z | A) \end{pmatrix}^T \right\}^{-1}.$$

Here, the notation  $\text{vec}(\cdot)$  represents the vectorization of the corresponding matrix. The asymptotical variance can be estimated by the sample variance with all expectation terms replaced by their sample means. In our simulation and application parts, we will implement the bootstrap procedure in order not to introduce complex variance expressions.

#### 4. Simulation Study

In this section, we conduct several experiments to evaluate the performance of the proposed estimator. We compare our approach with two stage least square method (2SLS).

---

We generate seven simulated datasets based on the model (2.1a)-(2.1b). The non-Gaussian noise  $\varepsilon_A$  is drawn from exponential distribution with rate 0.1 while  $Z, U$  and  $\varepsilon_Y$  are all marginally following a standard normal distribution. The value of  $\alpha$  will be fixed at 1, the values of  $\lambda, s$  will be fixed at 0.5, while the values of  $\gamma, \beta, \xi$  will vary. If the corresponding edge in Figure 1 is present, we will set the corresponding parameter to a nonzero number, i.e.,  $\gamma = \beta = 1, \xi = 0.5$ ; otherwise, it will be set to 0.

The purpose of these settings is to compare the performance of our proposed estimator with the commonly used 2SLS estimator. Cases 1 and 2 evaluate the performance of our estimator in the valid instrumental variable (IV) setting, with Case 1 representing a strong IV scenario with  $(\gamma, \beta, \xi) = (1, 0, 0)$ , and Case 2 representing a weak IV scenario with  $(\gamma, \beta, \xi) = (0.01, 0, 0)$ . Cases 3-7 are designed to compare the two estimators in the presence of invalid IVs due to direct effects ( $Z \rightarrow Y$ ) or unmeasured confounding. For more information on all simulated scenarios, please refer to Table 1.

We consider sample sizes of  $n \in \{100, 300, 500\}$  and perform 300 Monte Carlo simulations for each scenario. We employ the gradient boosting method to estimate  $E(Z | A)$  (Friedman, 2002). The simulation results are summarized in Table 1.

- For Cases 1-2, covariate  $Z$  can be viewed as a valid instrumental variable (IV). In Case 1, both methods exhibit relatively small bias, but our method has smaller standard errors and more significant coverage probability. However, in the weak IV setting (Case 2), the 2SLS estimator becomes more unstable with increased variances and conservative coverage probability, as reported in previous literature (Stock et al., 2002).
- Cases 3-5 allow for the exclusion restriction or independence assumption of the IV to be violated. In such cases, our EUNC procedure still yields a consistent estimator, while 2SLS results in a relatively large bias and poor coverage probability. Thus, we can conclude that employing a possibly invalid IV using the 2SLS method will provide worse estimation results.
- In Cases 6 and 7, covariate  $Z$  can be regarded as a non-differential and an outcome-inducing confounding proxy (Miao et al., 2018). Though  $A$  and  $Z$  are conditionally independent, the requirement  $\text{cov}(Z, A) \neq 0$  in Theorem 1 still holds due to the presence of latent confounders  $U$ . The EUNC method exhibits remarkable robustness against deviations from invalid IV estimation and achieves superior performance across all evaluation metrics.

We can see that the EUNC procedure always provides favorable estimation results even when all edges in Figure 1 are present. More simulation results including two treatments case and sensitivity analysis for the Gaussianity of  $U$  are provided in the supplementary material.

Table 1: Comparison results of EUNC procedure and 2SLS. Here  $\checkmark$  implies the corresponding edge exists and  $\times$  means the absence of the edge. SD represents the standard deviation and 95% CP is the coverage proportion of the 95% asymptotic confidence intervals. Results are averaged over 300 repeated experiments. The bias and SD have been multiplied by 1000.

	$Z \rightarrow A$	$Z \rightarrow Y$	$U \rightarrow Z$	Sample size	Bias		SD		95% CP	
					EUNC	2SLS	EUNC	2SLS	EUNC	2SLS
Case 1	$\checkmark$	$\times$	$\times$	100	0.8	35.2	15.2	4396.1	96.7%	97.7%
				300	0.5	28.6	7.7	2799.4	95.7%	99.0%
				500	0.4	29.0	5.8	1449.2	96.0%	98.3%
Case 2	$\checkmark$ (weak)	$\times$	$\times$	100	3.1	57.7	15.4	6169.1	97.3%	92.7%
				300	2.3	18.5	8.2	18612.8	98.3%	94.3%
				500	2.3	150.8	6.4	8043.3	95.7%	93.7%
Case 3	$\checkmark$	$\checkmark$	$\times$	100	6.4	2394.6	20.7	44166.6	94.0%	95.0%
				300	3.1	1420.3	9.9	29372.0	97.3%	93.7%
				500	1.5	1442.4	6.9	23295.0	95.3%	94.7%
Case 4	$\checkmark$	$\times$	$\checkmark$	100	0.7	175.6	14.9	7749.6	97.7%	95.3%
				300	0.7	282.4	7.5	6789.6	94.3%	93.7%
				500	1.2	77.3	5.7	6341.5	94.7%	87.3%
Case 5	$\checkmark$	$\checkmark$	$\checkmark$	100	9.5	835.8	22.4	185903.3	96.7%	91.7%
				300	3.6	1148.0	10.0	14379.8	96.7%	86.7%
				500	2.4	1210.0	7.1	81481.0	97.0%	81.7%
Case 6	$\times$	$\times$	$\checkmark$	100	1.7	1978.0	15.1	8744.9	97.0%	87.7%
				300	0.2	476.9	8.4	19060.6	97.7%	89.7%
				500	0.7	1017.0	6.5	22641.7	96.3%	91.0%
Case 7	$\times$	$\checkmark$	$\checkmark$	100	1.8	713.7	25.4	56002.6	97.0%	89.0%
				300	1.6	82528.3	14.6	111138.4	97.3%	89.7%
				500	0.9	1605.3	10.6	88177.2	96.3%	89.7%

---

## 5. Application

In this section, we apply our method to a real-world dataset investigating the impact of the trade on income. We prioritize the search for Gaussian covariates, rather than focusing on identifying valid instrumental variables from a candidate set like previous literatures (Frankel and Romer, 1999; Kukla-Gryz, 2009; Lin et al., 2022; Fan and Wu, 2022). One of the main challenges in estimating the effect of trade on income is the potential endogeneity due to unmeasured common confounders for trade and income. Based on the study of Lin et al. (2022), we consider a modified structural model between income and trade share as follows:

$$\tilde{Y}_i = b + \alpha A_i + \beta \tilde{Z}_i + \varepsilon_i, \quad (5.1)$$

where  $\tilde{Y}_i = \log(Y_i)$ ,  $\tilde{Z}_i = \log(Z_i)$ , and  $Y_i$  is income per capita in country  $i$ ,  $A_i$  is the share of international trade to GDP,  $Z_i$  is an observed covariate, and  $\varepsilon_i$  is an endogenous error term.

We collect cross-sectional data for the year 2019 from the World Bank and use 154 records for analysis after removing missing values. The trade share  $A$  is a non-Gaussian treatment because it fails to pass the Anderson-Darling test. We then perform the Anderson-Darling test to select an ob-

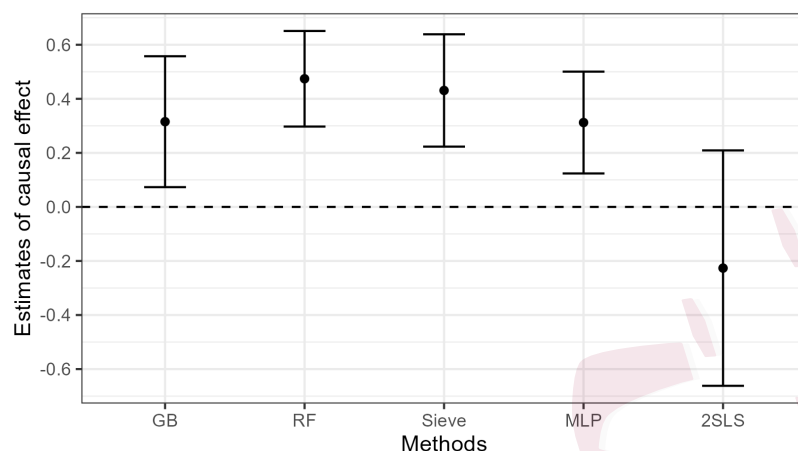


Figure 3: Confidence intervals of causal effect estimates for our method and 2SLS. Here GB, RF, Sieve and MLP represent Gradient Boosting, Random Forest, sieve estimation and Multi-Layer Perceptron. These methods are respectively used for calculating conditional expectation in our method.

served Gaussian covariate  $\tilde{Z}$ , and the log value of total annual freshwater withdrawals (in billion cubic meters) is finally regarded as a Gaussian covariate with a  $p$ -value of 0.75.

The results are summarized in Figure 3. We employ four nonparametric methods for calculating  $E(\tilde{Z} | \tilde{A})$  and use 2SLS estimator as a comparison. The 95% confidence intervals are obtained by nonparametric bootstrap method. Our method yields significantly positive effect estimates regardless of the method used for estimating the conditional expectation. This corroborates with previous findings using instrumental variables (Frankel and Romer, 1999; Kukla-Gryz, 2009; Lin et al., 2022; Fan and Wu, 2022). In contrast, when considering the Gaussian covariate as a valid IV in the

2SLS method, there is no statistically significant evidence suggesting that the trade share affects income. These results indicate that our method can yield similar conclusions by employing a Gaussian covariate instead of a valid IV, which can be particularly useful when only potentially invalid instrumental variables are available. Practitioners can also include some interaction terms or observed non-Gaussian covariates into analysis to enhance credibility, as demonstrated by Corollary 1 and Corollary 2.

## **6. Conclusions and Future work**

Our study focuses on the problem of identifying causal effects using auxiliary covariates and non-Gaussianity from the treatment. We present sufficient conditions for identifying effects of both single and multiple treatments using observed Gaussian covariates, which can be partially verified with observational data. Our results reveal that for a single treatment, identifying the causal effect can be facilitated by finding a Gaussian covariate with other non-Gaussian covariates being conditioned, rather than a valid instrumental variable. Additionally, our identification results for multiple treatments indicate that the minimum number of required Gaussian covariates equals the smaller dimension between the dimension of treatments and the dimension of latent confounders. Treatments can also be causally

correlated without a factor structure.

There are several notable aspects of the approach developed in this paper. First, as illustrated in our second example, we can extend from a linear setting to a nonlinear outcome model. Second, we can perform sensitivity analysis to evaluate the robustness of Gaussian assumptions of latent confounders. Third, simulation studies show that our method has the potential to handle the weak instrumental variable problem. Thus, comparing theoretical properties between the proposed estimator and the 2SLS estimator would be an interesting problem. Fourth, efficiently searching for Gaussian covariates in a large candidate instrument set remains challenging but of great practical and theoretical interest. We leave the study of these issues as a future research topic as it is beyond the scope of this paper.

### **Supplementary Materials**

Supplementary Material available online includes additional technical proofs and simulation results including two treatments case and sensitivity analysis of the Gaussianity of unmeasured confounders.



---

## Appendix

### Proof of Theorem 1

Let

$$\begin{pmatrix} Z \\ U \end{pmatrix} \sim N \left( \mathbf{0}, \begin{pmatrix} 1 & \xi^T \\ \xi & I_t \end{pmatrix} \right), \quad \xi \in \mathbb{R}^t.$$

The conditional expectation of  $Z$  given  $\varepsilon_A$  and  $A$  will be

$$\begin{aligned} E(Z \mid \varepsilon_A, A) &= E(Z \mid \varepsilon_A, \gamma Z + \lambda^T U + \varepsilon_A) \\ &= E(Z \mid \varepsilon_A, \gamma Z + \lambda^T U) \\ &= E(Z \mid \gamma Z + \lambda^T U), \end{aligned}$$

where the last equality holds because  $\varepsilon_A \perp\!\!\!\perp (Z, U)$ . It is easy to observe

$$\begin{pmatrix} Z \\ \gamma Z + \lambda^T U \end{pmatrix} \sim N \left( \mathbf{0}, \begin{pmatrix} 1 & \gamma + \xi^T \lambda \\ \gamma + \lambda^T \xi & \gamma^2 + \lambda^T \lambda + 2\gamma \xi^T \lambda \end{pmatrix} \right),$$

which implies

$$E(Z \mid \varepsilon_A, A) = E(Z \mid \gamma Z + \lambda^T U) = \frac{\gamma + \xi^T \lambda}{\gamma^2 + \lambda^T \lambda + 2\gamma \xi^T \lambda} (\gamma Z + \lambda^T U).$$

---

Similar arguments show

$$E(U | \varepsilon_A, A) = E(U | \gamma Z + \lambda^T U) = \frac{\gamma \xi + \lambda}{\gamma^2 + \lambda^T \lambda + 2\gamma \xi^T \lambda} (\gamma Z + \lambda^T U).$$

When  $\text{cov}(A, Z) = \gamma + \xi^T \lambda \neq 0$ , we have

$$E(U | \varepsilon_A, A) = \frac{\gamma \xi + \lambda}{\gamma + \xi^T \lambda} E(Z | \varepsilon_A, A),$$

which means

$$E(U | A) = \frac{\gamma \xi + \lambda}{\gamma + \xi^T \lambda} E(Z | A).$$

So the conditional expectation of  $Y$  given  $A$  will be

$$\begin{aligned} E(Y | A) &= \alpha A + \beta E(Z | A) + s^T E(U | A) \\ &= \alpha A + \left\{ \beta + \frac{s^T (\gamma \xi + \lambda)}{\gamma + \xi^T \lambda} \right\} E(Z | A). \end{aligned} \tag{6.1}$$

Now we demonstrate that  $E(Z | A)$  is nonlinear in  $A$  when  $\varepsilon_A$  is non-Gaussian with non-zero variance below. To show  $E(Z | A)$  is actually nonlinear in  $A$ , we prove by contradiction. Suppose  $E(Z | A) = cA$  for some constant  $c \in \mathbb{R}^1$ , then we must have

$$E(U | A) = \frac{\gamma \xi + \lambda}{\gamma + \xi^T \lambda} E(Z | A) = \frac{c(\gamma \xi + \lambda)}{\gamma + \xi^T \lambda} A,$$

---

which implies

$$E(\zeta \mid \zeta + \varepsilon_A) = \gamma E(Z \mid A) + \lambda^T E(U \mid A) = \tilde{c}(\zeta + \varepsilon_A), \quad (6.2)$$

where

$$A = \zeta + \varepsilon_A, \quad \zeta = \gamma Z + \lambda^T U, \quad \tilde{c} = c \left\{ \gamma + \lambda^T \frac{(\gamma \xi + \lambda)}{\gamma + \xi^T \lambda} \right\}.$$

Then the equation (6.2) demonstrates

$$E\{(1 - \tilde{c})\zeta - \tilde{c}\varepsilon_A \mid \zeta + \varepsilon_A\} = 0. \quad (6.3)$$

Multiply equation (6.3) by  $\exp\{it(\zeta + \varepsilon_A)\}$  and take expectation:

$$E[(1 - \tilde{c})\zeta \exp\{it(\zeta + \varepsilon_A)\}] - E[\tilde{c}\varepsilon_A \exp\{it(\zeta + \varepsilon_A)\}] = 0.$$

The independence of  $\zeta$  and  $\varepsilon_A$  implies

$$E\{(1 - \tilde{c})\zeta \exp(it\zeta)\} E\{\exp(it\varepsilon_A)\} - E\{\tilde{c}\varepsilon_A \exp(it\varepsilon_A)\} E\{\exp(it\zeta)\} = 0.$$

---

Without loss of generality, we assume  $\text{var}(\zeta) = \text{var}(\varepsilon_A) = 1$  below for convenience. The characteristic function of the standard normal distribution implies

$$E\{\exp(it\zeta)\} = \exp\left(-\frac{t^2}{2}\right), \quad E\{(1-\tilde{c})\zeta \exp(it\zeta)\} = (1-\tilde{c})it \exp\left(-\frac{t^2}{2}\right).$$

Let  $\phi(t) = E\{\exp(it\varepsilon_A)\}$  and  $\phi'(t) = iE\{\varepsilon_A \exp(it\varepsilon_A)\}$ , then we have

$$(1-\tilde{c})it \exp\left(-\frac{t^2}{2}\right) \cdot \phi(t) + \tilde{c} \exp\left(-\frac{t^2}{2}\right) \cdot i\phi'(t) = 0,$$

which means

$$(1-\tilde{c})t\phi(t) + \tilde{c}\phi'(t) = 0 \tag{6.4}$$

When  $\tilde{c} \in \{0, 1\}$ ,  $\phi(t)$  is a constant and  $\varepsilon_A$  is deterministic, which is infeasible because we have assumed the variance of  $\varepsilon_A$  is non-zero. Thus, for all

$\tilde{c} \notin \{0, 1\}$ , we have

$$\phi'(t) + (\tilde{c}^{-1} - 1)t\phi(t) = 0, \text{ and } \phi(0) = 1. \tag{6.5}$$

---

Solving (6.5) gives the characteristic function of  $\varepsilon_A$

$$\phi(t) = \exp\left(-\frac{\tilde{c}^{-1} - 1}{2}t^2\right),$$

which implies  $\varepsilon_A$  must be Gaussian. This is a contradiction and hence,  $E(Z | A)$  must be nonlinear in  $A$ . Now let

$$h = \beta + \frac{s^T(\gamma\xi + \lambda)}{\gamma + \xi^T\lambda}, \quad g(A) = \{A, E(Z | A)\}^T.$$

From equation (6.1), we have

$$\begin{pmatrix} \alpha \\ h \end{pmatrix} = E\{g(A)g(A)^T\}^{-1}E\{g(A)Y\},$$

which means  $\alpha$  is identifiable.

### Acknowledgments

The research work is supported by National Key R&D Program of China (2022ZD0160300). The authors thank the referees and an editor for helpful comments.

## References

- Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *The Annals of Mathematical Statistics* 23(2), 193–212.
- Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434), 444–455.
- Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Berlin: Springer.
- Cattaneo, M. D., R. K. Crump, and M. Jansson (2012). Optimal inference for instrumental variables regression with non-Gaussian errors. *Journal of Econometrics* 167(1), 1–15.
- Chen, Z. and L. Chan (2013). Causality in linear non-Gaussian acyclic models in the presence of latent Gaussian confounders. *Neural Computation* 25(6), 1605–1641.
- Chernozhukov, V., W. K. Newey, and R. Singh (2022). Automatic debiased machine learning of causal and structural effects. *Econometrica* 90(3), 967–1027.
- D’Amour, A. (2019). Comment: Reflections on the deconfounder. *Journal of the American Statistical Association* 114(528), 1597–1601.
- Entner, D. and P. O. Hoyer (2011). Discovering unconfounded causal relationships using linear non-Gaussian models. In *JSAI International Symposium on Artificial Intelligence*, pp. 181–195. Springer.
- Fan, Q. and Y. Wu (2022). Endogenous treatment effect estimation with a large and mixed set of instruments and control variables. *The Review of Economics and Statistics*, 1–45.
- Frankel, J. A. and D. Romer (1999). Does trade cause growth? *American Economic Review* 89(3), 379–399.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38(4), 367–378.
- Guo, Z., H. Kang, T. Tony Cai, and D. S. Small (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.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 50(4), 1029–1054.

## REFERENCES

---

- Hoyer, P. O., A. Hyvarinen, R. Scheines, P. L. Spirtes, J. Ramsey, G. Lacerda, and S. Shimizu (2008). Causal discovery of linear acyclic models with arbitrary distributions. *Conference on Uncertainty in Artificial Intelligence*.
- Hoyer, P. O., S. Shimizu, A. J. Kerminen, and M. Palviainen (2008). Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *International Journal of Approximate Reasoning* 49(2), 362–378.
- Imai, K. and Z. Jiang (2019). Comment: The challenges of multiple causes. *Journal of the American Statistical Association* 114(528), 1605–1610.
- Kang, H., A. Zhang, T. T. Cai, and D. S. Small (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American Statistical Association* 111(513), 132–144.
- Kasy, M. (2014). Instrumental variables with unrestricted heterogeneity and continuous treatment. *The Review of Economic Studies* 81(4), 1614–1636.
- Kong, D., S. Yang, and L. Wang (2022). Identifiability of causal effects with multiple causes and a binary outcome. *Biometrika* 109(1), 265–272.
- Kukla-Gryz, A. (2009). Economic growth, international trade and air pollution: A decomposition analysis. *Ecological Economics* 68(5), 1329–1339.
- Kuroki, M. and J. Pearl (2014). Measurement bias and effect restoration in causal inference. *Biometrika* 101(2), 423–437.
- Li, C., X. Shen, and W. Pan (2023). Nonlinear causal discovery with confounders. *Journal of the American Statistical Association*, 1–10.
- Lin, Y., F. Windmeijer, X. Song, and Q. Fan (2022). On the instrumental variable estimation with many weak and invalid instruments. *arXiv preprint arXiv:2207.03035*.
- Lipsitch, M., E. T. Tchetgen, and T. Cohen (2010). Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)* 21(3), 383.
- Liu, Z., T. Ye, B. Sun, M. Schooling, and E. T. Tchetgen (2022). Mendelian randomization mixed-scale treatment effect robust identification and estimation for causal inference. *Biometrics* 79(3).
- Miao, W., Z. Geng, and E. J. Tchetgen Tchetgen (2018). Identifying causal effects with proxy

## REFERENCES

---

- variables of an unmeasured confounder. *Biometrika* 105(4), 987–993.
- Miao, W., W. Hu, E. L. Ogburn, and X.-H. Zhou (2023). Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association* 118(543), 1953–1967.
- Park, S. and S. Gupta (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science* 31(4), 567–586.
- Rosenbaum, P. R. and D. B. Rubin (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)* 45(2), 212–218.
- Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association* 75(371), 591–593.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association* 100(469), 322–331.
- Salehkaleybar, S., A. Ghassami, N. Kiyavash, and K. Zhang (2020). Learning linear non-Gaussian causal models in the presence of latent variables. *The Journal of Machine Learning Research* 21(1), 1436–1459.
- Shimizu, S., P. O. Hoyer, and A. Hyvärinen (2009). Estimation of linear non-Gaussian acyclic models for latent factors. *Neurocomputing* 72(7-9), 2024–2027.
- Shimizu, S., P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research* 7(10).
- Shimizu, S., T. Inazumi, Y. Sogawa, A. Hyvarinen, Y. Kawahara, T. Washio, P. O. Hoyer, K. Bollen, and P. Hoyer (2011). Directlingam: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research-JMLR* 12(2011), 1225–1248.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518–529.
- Stone, J. V. (2002). Independent component analysis: an introduction. *Trends in Cognitive Sciences* 6(2), 59–64.
- Sun, B., Y. Cui, and E. T. Tchetgen (2022). Selective machine learning of the average treat-



## REFERENCES

---

- ment effect with an invalid instrumental variable. *Journal of Machine Learning Research* 23(204), 1–40.
- Tchetgen, E. J. T., A. Ying, Y. Cui, X. Shi, and W. Miao (2020). An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*.
- Wang, Y. and D. M. Blei (2019). The blessings of multiple causes. *Journal of the American Statistical Association* 114(528), 1574–1596.
- Wang, Y. S. and M. Drton (2020). High-dimensional causal discovery under non-Gaussianity. *Biometrika* 107(1), 41–59.
- Windmeijer, F., H. Farbmacher, N. Davies, and G. Davey Smith (2019). On the use of the Lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association* 114(527), 1339–1350.