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

Kang Shuai[a], Shanshan Luo[b], Yue Zhang[a], Feng Xie[b] and Yangbo He[a*]

[a]Peking University, [b]Beijing Technology and Business University

Supplementary Material

This supplementary material includes additional proofs, simulation results for two treatments case, sensitivity analysis and a data illustration for evaluating the effect of the trade on the income.

S1 Proofs

S1.1 Proof of Corollary 1

From the proof of Theorem 1, we obtain the following relations

$$E(Y \mid A) = \alpha A + hE(Z \mid A) + \tilde{h}AE(Z \mid A),$$

where $E(Z \mid A)$ must be nonlinear in A. Thus, $A, E(Z \mid A)$ and $AE(Z \mid A)$ must be linearly independent. Then the target parameter can be computed

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

as follows

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

where $\tilde{g}(A) = \big\{A, E(Z \mid A), AE(Z \mid A)\big\}^{\mathrm{T}}$.

S1.2 Proof of Corollary 2

Firstly, we have the following relations because $(\tilde{Z}, \varepsilon_A) \perp\!\!\!\perp (Z, U)$

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

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

which immediately implies (similar as Theorem 1)

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

Thus, we have

$$E(Y \mid A, \tilde{Z}) = \alpha A + \left\{\beta + \frac{s^{\mathrm{T}}(\gamma\xi + \lambda)}{\gamma + \xi^{\mathrm{T}}\lambda}\right\} E(Z \mid A, \tilde{Z}) + \tilde{\beta}^{\mathrm{T}}\tilde{Z}.$$

So the only requirement for identification is that $E(Z \mid A, \tilde{Z})$ should not be linear in $A, \tilde{Z}$. This is similar to the proof of Theorem 1 as long as we observe that

$$E(Z \mid A, \tilde{Z}) = E(Z \mid \tilde{A}),$$

where $\tilde{A} = A - \tilde{\gamma}^{\mathrm{T}}\tilde{Z}$. Thus, similar arguments as Theorem 1 will imply that $E(Z \mid A, \tilde{Z})$ is nonlinear in $\tilde{A}$ and we complete the proof of Corollary 2.

S1.3 Proof of Corollary 3

From the proof of Theorem 1, the conditional expectation $E(Z \mid A)$ and $E(U \mid A)$ are proportional. Thus, we have

$$E(Y \mid A) = f(A) + \beta E(Z \mid A) + s^{\mathrm{T}} E(U \mid A) = f(A) + \tilde{\beta} E(Z \mid A) = f(A) + \check{\beta} A,$$

where $\tilde{\beta}, \check{\beta}$ are constants. Under first condition of Corollary 1, $\check{\beta}$ can be solved as follows

$$\check{\beta} = \lim_{a \to \infty} \frac{E(Y \mid A = a)}{a},$$

which implies

$$f(A) = E(Y \mid A) - \lim_{a \to \infty} \frac{E(Y \mid A = a)}{a} E(Z \mid A).$$

Under the second condition that $f(A)$ does not include the linear term of A, write the conditional expectation $E(Y \mid A)$ as the following expansion

$$E(Y \mid A) = \sum_{i=1}^{\infty} b_i A^i,$$

which demonstrates the constants b_1 is identifiable. Thus, $f(A) = E(Y \mid A) - b_1 A$ is identifiable. So we have completed the proof.

S1.4 Proof of Theorem 2

Let

$$\begin{pmatrix} Z \\ U \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} I_l & \Sigma \\ \Sigma^{\mathrm{T}} & I_t \end{pmatrix}\right),$$

and the conditional expectation of Z given ε_A and A will be

$$\begin{aligned} E(Z \mid \varepsilon_A, A) &= E(Z \mid \varepsilon_A, \Gamma Z + \Lambda U + \varepsilon_A) \\ &= E(Z \mid \varepsilon_A, \Gamma Z + \Lambda U) \\ &= E(Z \mid \Gamma Z + \Lambda U), \end{aligned}$$

where the last equality holds because $\varepsilon_A \perp\!\!\!\perp (Z, U)$. Similar calculation as proof of Theorem 1 gives

$$E(Z \mid \varepsilon_A, A) = E(Z \mid \Gamma Z + \Lambda U) = (\Gamma^{\mathrm{T}} + \Sigma\Lambda^{\mathrm{T}})\mathrm{cov}(\Gamma Z + \Lambda U)^{-1}(\Gamma Z + \Lambda U).$$

Similar arguments show

$$E(U \mid \varepsilon_A, A) = E(U \mid \Gamma Z + \Lambda U) = (\Lambda^{\mathrm{T}} + \Sigma^{\mathrm{T}}\Gamma^{\mathrm{T}})\mathrm{cov}(\Gamma Z + \Lambda U)^{-1}(\Gamma Z + \Lambda U).$$

Combined from all above, we have

$$\begin{aligned} E(Z \mid A) &= (\Gamma^{\mathrm{T}} + \Sigma\Lambda^{\mathrm{T}})\mathrm{cov}(\Gamma Z + \Lambda U)^{-1}E_{\varepsilon_A|A}(\Gamma Z + \Lambda U), \\ E(U \mid A) &= (\Lambda^{\mathrm{T}} + \Sigma^{\mathrm{T}}\Gamma^{\mathrm{T}})\mathrm{cov}(\Gamma Z + \Lambda U)^{-1}E_{\varepsilon_A|A}(\Gamma Z + \Lambda U). \end{aligned}$$

So if the second condition in Theorem 2 holds, we know the equation $\Lambda^{\mathrm{T}} + \Sigma^{\mathrm{T}}\Gamma^{\mathrm{T}} = \Phi \cdot (\Gamma^{\mathrm{T}} + \Sigma\Lambda^{\mathrm{T}})$ has solution for Φ. This also means $E(U \mid A) =$

$\Phi \cdot E(Z \mid A)$ has solution for Φ. Now we can obtain

$$\begin{aligned} E(Y \mid A) &= \alpha^{\mathrm{T}} A + \beta^{\mathrm{T}} E(Z \mid A) + s^{\mathrm{T}} E(U \mid A) \\ &= \alpha^{\mathrm{T}} A + (\beta^{\mathrm{T}} + s^{\mathrm{T}} \Phi) E(Z \mid A), \end{aligned}$$

where $E(Z \mid A)$ and A are linearly independent from the first condition of Theorem 2. Let

$$h = \beta + \Phi^{\mathrm{T}} s, \quad g(A) = \left\{A^{\mathrm{T}}, E(Z \mid A)^{\mathrm{T}}\right\}^{\mathrm{T}},$$

then

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

which implies α is identifiable.

S1.5 Proof of Theorem 3

To prove estimator $\hat{\alpha}$ is root-n consistent, it is enough to show each term in our expression is root-n consistent. The $n^{-1/4}$-consistency of $\hat{E}(Z \mid A)$

implies

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} A_i\hat{E}(Z \mid A_i)^{\mathrm{T}} &= \frac{1}{n}\sum_{i=1}^{n} A_i E(Z \mid A_i)^{\mathrm{T}} + o_p(n^{-1/2}),\\
\frac{1}{n}\sum_{i=1}^{n} Y_i\hat{E}(Z \mid A_i) &= \frac{1}{n}\sum_{i=1}^{n} Y_i E(Z \mid A_i) + o_p(n^{-1/2}),\\
\frac{1}{n}\sum_{i=1}^{n} \hat{E}(Z \mid A_i)\hat{E}(Z \mid A_i)^{\mathrm{T}} &= \frac{1}{n}\sum_{i=1}^{n} E(Z \mid A_i)E(Z \mid A_i)^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{E}(Z \mid A_i) - E(Z \mid A_i)\right\}\\
&\quad \cdot \left\{\hat{E}(Z \mid A_i) - E(Z \mid A_i)\right\}^{\mathrm{T}} + o_p(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n} E(Z \mid A_i)E(Z \mid A_i)^{\mathrm{T}} + \left\{o_p(n^{-1/4})\right\}^2 + o_p(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n} E(Z \mid A_i)E(Z \mid A_i)^{\mathrm{T}} + o_p(n^{-1/2}),
\end{aligned}
$$

so the matrix estimator $\hat{X}$ is $\sqrt{n}$-consistent for X, where

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

and

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

which implies

$$
\begin{aligned}
\sqrt{n}(\hat{\alpha}-\alpha) =& \hat{X}_{11}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i Y_i\right) + \hat{X}_{12}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{E}(Z \mid A_i) Y_i\right\} - \sqrt{n} X_{11} E(AY) \\
& - \sqrt{n} X_{12} E\{E(Z \mid A) Y\} \\
=& \hat{X}_{12}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \big\{E(Z \mid A_i) Y_i - E\{E(Z \mid A) Y\}\big\} + \sqrt{n}(\hat{X}_{12}-X_{12}) E\{E(Z \mid A) Y\} \\
& + \hat{X}_{11}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{A_i Y_i - E(AY)\} + \sqrt{n}(\hat{X}_{11}-X_{11}) E(AY) + o_p(1) \\
=& X_{12}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \big\{E(Z \mid A_i) Y_i - E\{E(Z \mid A) Y\}\big\} + \sqrt{n}(\hat{X}_{12}-X_{12}) E\{E(Z \mid A) Y\} \\
& + X_{11}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{A_i Y_i - E(AY)\} + \sqrt{n}(\hat{X}_{11}-X_{11}) E(AY) + o_p(1), \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \big\{X_{11} A_i Y_i + X_{12} E(Z \mid A_i) Y_i - \alpha\big\} \\
& + \sqrt{n}(\hat{X}_{11}-X_{11}) E(AY) + \sqrt{n}(\hat{X}_{12}-X_{12}) E\{E(Z \mid A) Y\} + o_p(1).
\end{aligned}
$$

Let $\xi = \text{vec}\{AA^{\mathrm{T}}, E(Z \mid A)A^{\mathrm{T}}, E(Z \mid A)E(Z \mid A)^{\mathrm{T}}\}$, $\mu = E(\xi)$, $\tilde{\xi} = n^{-1}\sum_{i=1}^{n}\xi_i$, then the delta method demonstrates

$$
\begin{aligned}
\sqrt{n}(\hat{X}_{11}-X_{11}) E(AY) &= \frac{\partial}{\partial \mu}\{X_{11} E(AY)\} \cdot \sqrt{n}(\tilde{\xi}-\mu), \\
\sqrt{n}(\hat{X}_{12}-X_{12}) E\{E(Z \mid A) Y\} &= \frac{\partial}{\partial \mu}\big\{X_{12} E\{E(Z \mid A) Y\}\big\} \cdot \sqrt{n}(\tilde{\xi}-\mu).
\end{aligned}
$$

Thus, we have

$$
\sqrt{n}(\hat{X}_{11}-X_{11}) E(AY) + \sqrt{n}(\hat{X}_{12}-X_{12}) E\{E(Z \mid A) Y\} = \frac{\partial \alpha(\mu)}{\partial \mu} \cdot \sqrt{n}(\tilde{\xi}-\mu),
$$

where $\alpha(\mu) = X_{11}E(AY) + X_{12}E\{E(Z \mid A)Y\}$. Finally, we obtain the following linear expansion

$$\sqrt{n}(\hat{\alpha}-\alpha) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{X_{11}A_iY_i+X_{12}E(Z \mid A_i)Y_i-\alpha+\frac{\partial\alpha(\mu)}{\partial\mu}\xi_i-\mu\}+o_p(1) \xrightarrow{d} N(0,\Sigma_\alpha),$$

where

$$\Sigma_\alpha = \text{var}\left\{X_{11}AY + X_{12}E(Z \mid A)Y + \frac{\partial\alpha(\mu)}{\partial\mu}\xi\right\},$$

which demonstrates that $\hat{\alpha}$ is root-n consistent for α.

S2 Simulation

In this section, we will present the simulation results for two treatments case and the sensitivity analysis for the Gaussianity of unmeasured confounders.

S2.1 Two treatments case

In this section, we conduct extensive experiments based on the model for multiple treatments in terms of nine scenarios with two causally correlated treatments, as shown in Figure 1(b) of our paper. Here we use the EUNC algorithm to calculate causal effects of $A = (A_1, A_2)^{\mathrm{T}}$ on Y, namely α. The vector of two-dimensional observed covariate $Z = (Z_1, Z_2)$ follow multivariate normal distribution with unit variance. The latent confouder $U \in \mathbb{R}^1$.

The non-Gaussian noise terms ε_{A_1} and ε_{A_2} are respectively sampled from exponential distribution with rate 0.1 while Z_1, Z_2, U and ε_Y are all sampled from a standard normal distribution, where $Z_1 \perp\!\!\!\perp Z_2$. We set the values $\alpha = (1,1)^{\mathrm{T}}, \Lambda = (0.5, 0.5)^{\mathrm{T}}$ and $s = (0.5, 0.5)^{\mathrm{T}}$. The elements of vectors $\Gamma = (\gamma_1, \gamma_2)^{\mathrm{T}}, \beta = (\beta_1, \beta_2)^{\mathrm{T}}$ and $(\xi_1, \xi_2) = \{\mathrm{cov}(Z_1, U), \mathrm{cov}(Z_2, U)\}$ will vary. If the edge in Figure 2 is present, we will set the corresponding parameter to a nonzero number ($\gamma_i = 1, \beta_i = 1, \xi_i = 0.5, i = 1, 2$); except in weak IV settings, we set $\gamma_i = 0.01$ ($i = 1, 2$). Otherwise, it would be zero.

We consider these settings here to compare the estimation performance of our proposed estimator with the 2SLS estimator. Case 1-3 demonstrate the efficiency of our method in the valid IV setting, where case 2 and 3 respectively represent situations with one and two weak IVs. Cases 4-9 are designed for evaluating the performance of the two estimators with a possibly invalid IV, where the exclusion restriction or independence assumption may be violated. The details of these setting are presented in Table 1-2.

We consider sample sizes $n \in \{100, 300, 500\}$, and perform 300 repeated experiments for each scenario. The conditional expectation $E(Z \mid A)$ is estimated using the gradient boosting method. We report the main estimation results for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in Table 1 and 2, respectively.

As illustrated, the EUNC algorithm is considerably more effective than the 2SLS estimator and achieves superior performance across all evaluation metrics in all scenarios.

S2.2 Sensitivity analysis

In this section, we evaluate the robustness of our estimator with respect to the non-Gaussian assumption of treatment noise variable, and the possible violation of Gaussian assumption for latent concounders by imposing an additional non-Gaussian confounder $W \sim t_\nu$ confounding the treatment and outcome. Figure 1 gives an illustration of our setting in the presence of W. Here $\varepsilon_A \sim t_\nu$ is the treatment noise variable generated from Student's t_ν-distribution. As the degree of freedom ν increases, ε_A or W is more approaching the standard normal distribution. We display the results for $\nu \in [5, 30]$ with step size of 0.5, and the absolute estimation bias of causal effects are shown in Figure 2, 3, 4 and 5. Note that in Figure 4 and Figure 5 with $W \sim t_\nu$ confounding A-Y relation, ε_A still follows exponential distribution with the rate of 0.1. The sample size n is fixed at 300, and all scenario are based on 300 repeated experiments. We also fix α, β, γ all at 1 and λ, s at 0.5, and Z, U and ε_Y all marginally follow a standard normal distribution. It is important to note that Figure 2 and 4 present the results under the assumption that Z is unconfounded by U, while Figure 3 and 5 allow Z to be confounded by U with $\xi = \text{cov}(Z, U) = 0.5$.

Table 1: Comparison results for $\hat{\alpha}_1$ of EUNC procedure and 2SLS in two-treatment scenarios. Here ✓ implies the edge exists and × means the absence of the edge. SD represents standard deviation and 95% CP is the coverage proportion of the 95% asymptotic confidence intervals. Simulations results are averaged over 300 repeated experiments.

	$Z_1 \to A_1$	$Z_2 \to A_2$	$Z_1 \to Y$	$Z_2 \to Y$	$U \to Z$	Sample size	Bias		SD		95%CP	
							EUNC	2SLS	EUNC	2SLS	EUNC	2SLS
Case 1	✓	✓	×	×	×	100	1.1	38.5	23.6	6226.2	89.3%	94.0%
						300	2.0	19.1	12.2	4416.3	94.0%	97.3%
						500	2.6	48.3	9.0	2938.8	94.0%	98.0%
Case 2	✓ (weak)	✓	×	×	×	100	1.5	771.0	23.6	10026.9	90.3%	94.0%
						300	0.4	159.4	12.8	8670.1	92.0%	90.7%
						500	0.3	629.7	9.6	8029.0	92.0%	94.0%
Case 3	✓ (weak)	✓ (weak)	×	×	×	100	1.7	659.9	23.6	53275.0	93.7%	90.3%
						300	0.4	973.3	12.9	7115.5	91.0%	91.7%
						500	1.2	9557.8	9.8	9251.7	91.3%	92.0%
Case 4	✓	✓	✓	×	×	100	6.6	248.4	31.1	32625.6	92.7%	91.7%
						300	6.6	1109.5	15.3	40624.7	92.3%	97.3%
						500	6.3	1257.8	10.9	34004.9	93.7%	97.0%
Case 5	✓	✓	✓	✓	×	100	5.8	993.9	38.1	39949.3	90.0%	92.3%
						300	5.0	219.3	19.1	64458.8	93.0%	96.3%
						500	6.0	1284.0	13.7	59922.9	88.7%	96.7%
Case 6	✓ (weak)	✓	✓	✓	×	100	2.0	7252.4	39.0	90346.6	89.3%	90.3%
						300	1.8	1398.2	21.9	132812.6	92.0%	89.0%
						500	0.1	64809.9	16.5	149433.3	91.0%	89.7%
Case 7	✓ (weak)	✓ (weak)	✓	✓	×	100	3.5	2220.6	39.4	312877.9	89.3%	87.7%
						300	0.2	14827.5	22.7	90396.9	92.0%	88.7%
						500	0.8	516347.5	17.6	152163.3	88.3%	88.0%
Case 8	×	×	×	×	✓	100	2.1	28.0	23.5	25008.4	91.7%	88.0%
						300	1.4	24391.3	12.7	17748.2	93.3%	90.0%
						500	2.3	31.5	9.2	34062.2	94.0%	86.7%
Case 9	×	×	✓	✓	✓	100	4.1	203.5	44.7	90593.2	91.3%	87.0%
						300	5.7	128267.0	24.8	84866.2	90.7%	91.7%
						500	6.4	698.0	17.4	179913.0	91.7%	87.3%

Table 2: Comparison results for $\hat{\alpha}_2$ of EUNC procedure and 2SLS in two-treatment scenarios. Here ✓ implies the edge exists and × means the absence of the edge. SD represents standard deviation and 95% CP is the coverage proportion of the 95% asymptotic confidence intervals. Simulations results are averaged over 300 repeated experiments.

	$Z_1 \to A_1$	$Z_2 \to A_2$	$Z_1 \to Y$	$Z_2 \to Y$	$U \to Z$	Sample size	Bias		SD		95%CP	
							EUNC	2SLS	EUNC	2SLS	EUNC	2SLS
Case 1	✓	✓	×	×	×	100	3.1	14.7	17.3	3918.8	90.0%	93.3%
						300	1.5	13.4	9.0	3680.8	93.3%	96.3%
						500	1.8	16.7	6.7	2301.9	91.0%	99.0%
Case 2	✓ (weak)	✓	×	×	×	100	3.2	664.0	17.8	7747.0	86.3%	93.3%
						300	1.5	27.3	9.8	4827.6	87.7%	91.7%
						500	2.0	21.4	7.4	3112.0	89.0%	94.7%
Case 3	✓ (weak)	✓ (weak)	×	×	×	100	3.3	192.7	17.3	34050.8	91.0%	91.0%
						300	1.9	995.4	9.2	4914.1	90.0%	92.3%
						500	3.5	3942.6	7.0	7922.2	88.0%	90.7%
Case 4	✓	✓	✓	×	×	100	1.9	101.6	22.8	20445.8	89.3%	89.0%
						300	0.8	22.9	11.3	33023.8	92.0%	96.0%
						500	1.6	199.4	8.1	22557.8	91.0%	97.3%
Case 5	✓	✓	✓	✓	×	100	2.1	1002.3	28.0	28381.1	89.3%	92.0%
						300	4.3	1075.1	14.2	59410.9	90.7%	96.3%
						500	1.6	2687.3	10.4	43382.8	88.0%	97.3%
Case 6	✓ (weak)	✓	✓	✓	×	100	0.3	8857.1	29.0	89392.1	88.3%	91.7%
						300	4.5	123.6	16.6	73801.8	89.3%	89.7%
						500	1.0	9095.5	12.8	51014.7	90.3%	92.7%
Case 7	✓ (weak)	✓ (weak)	✓	✓	×	100	4.9	1150.6	28.9	189587.8	89.7%	88.3%
						300	1.1	14439.3	16.4	70957.2	89.3%	88.7%
						500	3.7	215947.8	12.8	128235.0	90.0%	87.3%
Case 8	×	×	×	×	✓	100	0.4	85.9	17.4	20488.9	91.3%	89.0%
						300	0.8	8829.0	9.2	12436.5	92.7%	91.3%
						500	0.3	57.1	6.8	20171.9	91.7%	85.0%
Case 9	×	×	✓	✓	✓	100	4.8	266.3	33.3	61598.3	91.3%	89.0%
						300	4.5	46584.4	18.2	61424.9	90.7%	91.0%
						500	2.7	767.0	12.9	106884.5	93.3%	85.3%

The results in Figure 4 and 5 both demonstrate that when an additional non-Gaussian variable W confounds the treatment-outcome relation, the estimation bias will not expand with nearly all values below 0.004, regardless of the degree of freedom ν. The results in Figure 2 and 3 show that with the noise variable $\varepsilon_A \sim t_\nu$, the estimation bias will increase a little when degree of freedom ν increases. However, when ν is less than 15, the bias remains under 0.1, which illustrates that the strength of non-Gaussianity from variable ε_A should be significant for better performance. The non-Gaussianity from ε_A plays a more vital role while the possible violation of Gaussian assumption of U seems not to be a big problem.

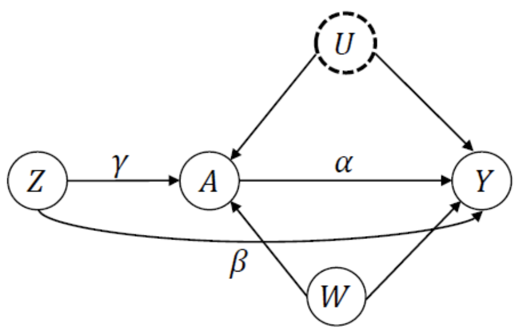

Figure 1: Causal diagram with an observed covariate Z, a treatment A, a latent confounder U, an outcome Y and an additional random variable W confounding A-Y relation.

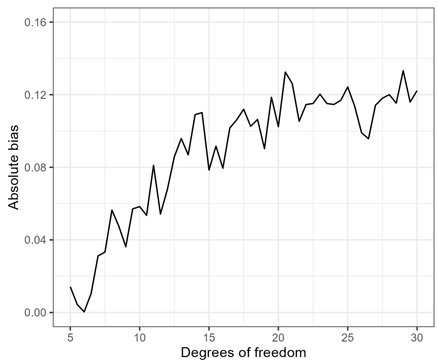

Figure 2: Estimation bias of causal effects α with varying degrees of freedom ν. Here Z is unconfounded by U and $\varepsilon_A \sim t_\nu$.

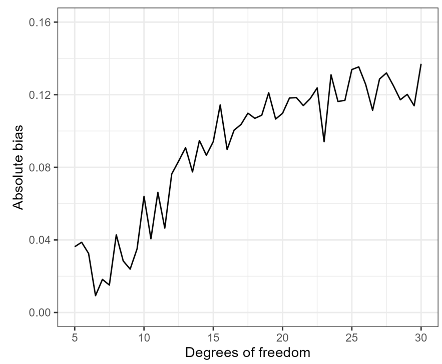

Figure 3: Estimation bias of causal effects α with varying degrees of freedom ν. Here Z is confounded by U and $\varepsilon_A \sim t_\nu$.

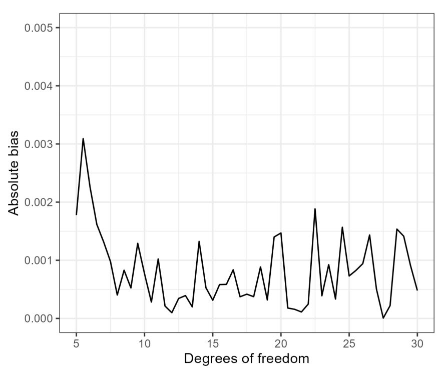

Figure 4: Estimation bias of causal effects α with an additional non-Gaussian $W \sim t_\nu$ confounding $A - Y$ relation, varying degrees of freedom ν. Here Z is unconfounded by U.

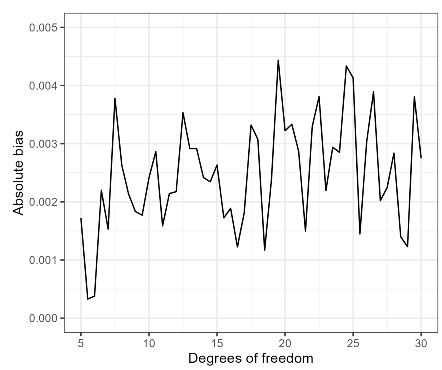

Figure 5: Estimation bias of causal effects α with an additional non-Gaussian $W \sim t_\nu$ confounding $A - Y$ relation, varying degrees of freedom ν. Here Z is confounded by U.