		Linear response function						Non-linear response function					
	Method	2SLS	IVGMM	DML	Ortho	Poly2SLS	KernelIV	2SLS	IVGMM	DML	Ortho	Poly2SLS	KernelIV
Binary	TrueIV	0.35	0.35	1.03	1.15	1.05	1.34	1.18	1.18	1.67	1.15	0.33	0.44
	UAS	2.96	2.96	4.99	3.92	1.96	0.22	1.63	1.63	4.63	3.92	1.96	0.58
	WAS	2.25	2.25	3.61	2.12	**1.06**	**0.21**	2.78	2.78	2.92	2.12	1.05	0.64
	DVAE.CIV	1.43 (0.24)	1.43 (0.24)	1.99 (1.74)	1.64 (2.7)	1.12 (0.88)	1.24 (0.02)	**0.83** (0.67)	**0.83** (0.67)	0.88 (0.42)	1.17 (0.71)	1.8 (2.4)	0.79 (0.01)
	CoCoIV	**0.26** (0.09)	**0.31** (0.07)	**1.59** (1.17)	**1.3** (1.25)	1.66 (3.24)	0.38 (0.31)	1.12 (0.17)	1.06 (0.2)	**0.4** (0.56)	**0.37** (0.47)	**0.47** (0.18)	**0.33** (0.02)
Continuous	TrueIV	2.2	2.27	0.002	0.11	8.41	5.21	3.5	3.61	1.48	1.48	5.05	5.1
	UAS	1.75	1.75	0.52	0.5	1.4	5.85	2.41	2.41	**1.34**	**1.34**	6.28	4.87
	WAS	0.79	0.79	0.25	0.24	0.69	3.51	1.57	1.57	1.52	1.51	4.95	5.18
	AutoIV	4.15 (9.41)	4.15 (9.41)	0.23 (0.47)	0.16 (0.2)	0.7 (0.79)	**3.18** (0.9)	25.25 (63.06)	25.25 (63.06)	1.65 (0.41)	1.67 (0.45)	4.99 (0.94)	4.26 (1.33)
	CoCoIV	**0.07** (0.09)	**0.08** (0.09)	**0.03** (0.03)	**0.02** (0.01)	**0.35** (0.07)	3.51 (0.55)	**1.38** (0.1)	**0.3** (0.15)	1.35 (0.12)	1.37 (0.1)	**4.13** (0.32)	**3.82** (1.16)

Table 13: Experimental results on low-dimensional synthetic datasets.

		Linear response function						Non-linear response function					
	Method	2SLS	IVGMM	DML	Ortho	Poly2SLS	KernelIV	2SLS	IVGMM	DML	Ortho	Poly2SLS	KernelIV
Binary	TrueIV	0.19	0.19	0.57	0.87	0.45	0.41	2.6	2.6	0.71	0.54	0.19	0.24
	UAS	0.52	0.52	N/A	17.17	14.57	0.22	4.06	4.06	7.06	6.78	3.08	2.41
	WAS	0.37	0.37	1.74	1.74	0.85	**0.13**	2.74	2.74	4.35	4.41	2.09	2.1
	DVAE.CIV	1.87 (1.72)	1.87 (1.72)	1.96 (2.05)	3.75 (8.24)	**0.5** (0.13)	0.66 (0.11)	9.31 (28.3)	9.31 (28.3)	**2.48** (4.08)	14.2 (32.48)	**0.75** (0.11)	**0.7** (0.1)
	CoCoIV	**0.07** (0.03)	**0.05** (0.02)	**0.78** (0.8)	**0.54** (0.44)	0.61 (0.49)	0.24 (0.2)	**0.76** (0.06)	**0.7**(0.07)	3.19 (0.5)	**3.11** (0.42)	1.37 (0.19)	1.68 (0.29)
Continuous	TrueIV	0.2	0.2	0.93	0.88	0.46	0.12	2.6	2.6	0.59	0.60	0.19	0.22
	UAS	0.52	0.52	16.92	17.1	14.57	0.22	4.06	4.06	6.79	6.76	3.08	2.41
	WAS	0.37	0.37	1.76	1.72	0.85	**0.13**	2.74	2.74	4.43	4.4	2.09	2.1
	AutoIV	0.62 (0.76)	0.62 (0.76)	1.65 (0.94)	1.71 (0.96)	1.19 (0.5)	0.32 (0.25)	9.84 (32.47)	9.84 (32.47)	2.14 (0.59)	2.11 (0.59)	0.93 (0.32)	0.96 (0.34)
	CoCoIV	**0.05** (0.01)	**0.02** (0.01)	**0.23** (0.21)	**0.1** (0.08)	**0.24** (0.06)	0.26 (0.12)	**0.6** (0.05)	**0.6** (0.05)	**2.0** (0.19)	**1.89** (0.12)	**0.85** (0.07)	**0.82** (0.12)

Table 14: Experimental results on high-dimensional synthetic datasets.

Appendix D. General Responses on Experiments

Experimental Results with True IVs Here, we provide the results of experiments of synthetic datasets along with that of true IVs, colored with blue. Table 13, 14 also demonstrate that the estimators with IV representation from our model have better estimates than the estimators fed with true IVs in several cases, such as in low-dimensional datasets with 2SLS, IVGMM and KernelIV and in high-dimensioanl datasets with 2SLS and IVGMM etc. The trends would imply that in those cases, our model managed to generate better condensed IV representation than true IVs. However, in high-dimensional datasets with non-linear response functions, estimators using true IVs exhibit lower MAEs compared to other models, although our model outperforms others in most cases. Therefore, we believe that future research aimed at developing IV representations that can achieve performance comparable to true IVs, even in more challenging and complex scenarios, would be a promising direction.

ATE comparison As described in Appendix B.1, we construct the linear response function as $g(X) = 3X$ and the non-linear response function as $g(X) = \exp(0.5X)$. Accordingly, the ground truth ATE in these cases is 3 and $\exp(0.5) - 1 \approx 0.65$, respectively.

Since the main table focuses on MAEs, we provide a visualization here to illustrate how each model estimates ATEs under binary treatment.[6]

Fig. 5 shows that our model estimates ATEs more accurately than other models in the cases of 2SLS, IVGMM, and KernelIV, regardless of dataset dimensions. Furthermore, it produces fewer extreme values compared to other neural network-based model D.CIV. In contrast, for DML and

6. For continuous treatment, as mentioned in Sec. 5.1, we measure the MAE between the estimated and true values of the response function. However, since most estimators do not offer a way to calculate APCE, results for continuous treatment are omitted.

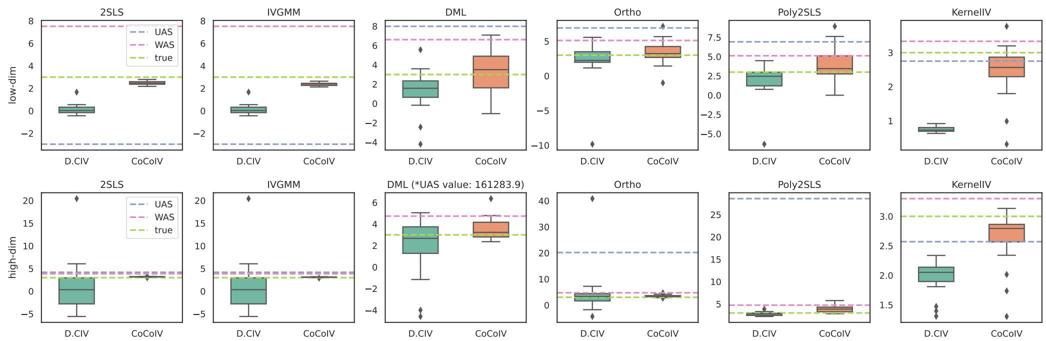

Figure 5: Experiment results of ATE on linear response function. Green line refers to true ATE 3. Rows imply the dimension of datasets and columns type of estimators.

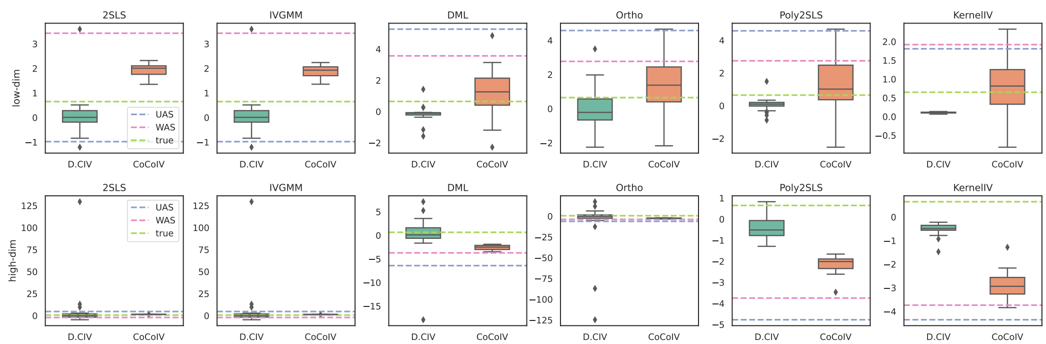

Figure 6: Experiment results of ATE on non-linear response function. Green line refers to true ATE 0.65. Rows imply the dimension of datasets and columns type of estimators.

Ortho, the performance of our model varies with the dataset dimension, which may be due to the higher sensitivity of these estimators, as noted in Sec. 5.1.

Fig. 6 presents the estimation of ATEs with a non-linear response function. For DML, Ortho, Poly2SLS, and KernelIV, while the estimates are more accurate than those of D.CIV, our model exhibits higher variance in low-dimensional datasets. This can be partly attributed to D.CIV's better compatibility with Poly2SLS and KernelIV, as observed in the linear case. However, in high-dimensional datasets, our model demonstrates better performance, particularly in the cases of 2SLS, IVGMM, DML, and Ortho, while producing fewer extreme values. This suggests the potential of our model to effectively handle more challenging high-dimensional scenarios.