Online supplement to "Causal Inference for Nonlinear Outcome Models with Possibly Invalid Instrumental Variables"

Sai Li and Zijian Guo

In Section D, we provide sufficient conditions to verify Condition 4.3. In Section E, we provide further simulation results.

D Verification of Condition 4.3

Let $t_i^* = ((d_i, w_i^\intercal)B^*, v_i)$ and $s_i^* = ((d, w^\intercal)B^*, v_i)$. Let f_{t^*} denote the density of t_i^*. We use $\mathcal{T}^*$ and $\mathcal{T}_v$ to denote the support of the density functions f_{t^*} and f_v, respectively. For a set $\mathcal{T}$, we use $\mathcal{T}^{\text{int}}$ to denote its interior.

We provide some generic examples of f_{t^*} and $q(\cdot)$ such that Condition 4.3 holds. Proposition D.1 provides a sufficient condition for Condition 4.3 (a) and (c). Proposition D.2 provides a sufficient condition for Condition 4.3 (b) when u_i has support $\mathbb{R}$. Proposition D.3 provides a sufficient condition for Condition 4.3 (b) when q is an indicator function.

Proposition D.1 (A sufficient condition for Condition 4.3 (a) and (c))**.** *Suppose that the support of t_i^* is $\mathcal{T}^* = [-a_1, a_1] \times [-a_2, a_2] \times [-a_3, a_3]$ and $\int_{t^* \in (\mathcal{T}^*)^{\text{int}}} f_t(t)dt = 1$, where $a_1, a_2 > 0$ and $|a_3| \leq C < \infty$. Suppose that the density f_{t^*} satisfies*

$$c_1 \leq \inf_{x \in \mathcal{T}_v^{\text{int}}} f_{t^*}((d, w^\intercal)B^*, x) \leq \sup_{x \in \mathcal{T}_v^{\text{int}}} f_{t^*}((d, w^\intercal)B^*, x) \leq c_2$$

for some positive constants c_1 and c_2. Moreover, we assume that $f_{t^*}(t)$ is differentiable and Lipschitz in $\mathcal{T}^*$ and $f_v(v)$ uniformly bounded in $\mathcal{T}_v$.

For any $u \in \{(d, w^\intercal)B^*_{\cdot,1} \pm Ch\} \times \{(d, w^\intercal)B^*_{\cdot,2} \pm Ch\}$ with some sufficiently large constant C, it holds that $|u_1| < a_1$ and $|u_2| < a_2$. Then Condition 4.3 (a) and (c) hold true.

Proof of Proposition D.1. We first verify Condition 4.3 (a). At the end of Section A, we showed that $B = B^*T$ for some invertible $T \in \mathbb{R}^{2\times 2}$. Because T is a constant matrix, we know that $c \leq |T^{-1}| < C < \infty$. Hence, by the linear transformation of density

$$f_t(\tilde{t}) = f_{t^*}(\tilde{t}_{1:2}T^{-1}, \tilde{t}_3)|T^{-1}|. \tag{D.1}$$

As $\mathcal{T}^*$ is convex, above expression implies that $\mathcal{T}$ is also convex, no matter $a_1, a_2 = \infty$ or not. Moreover,

$$\min_i f_t(s_i) \geq \inf_{x \in \mathcal{T}_v^{\text{int}}} f_t((d, w^\intercal)B, x) = \inf_{x \in \mathcal{T}_v^{\text{int}}} f_{t^*}(s_i^*)|T^{-1}| \geq c_0 > 0$$

for some constant $c_0 > 0$. Similarly, one can show that

$$\sup_{x \in \mathcal{T}_v^{\text{int}}} f_t((d, w^\intercal)B, x) \leq C_0 < \infty.$$

For the derivative of f_t, by (D.1),

$$\begin{aligned}\max_{1\leq i\leq n} \sup_{t_0 \in \mathcal{N}_h(s_i)\cap\mathcal{T}} \|\nabla f_t(t_0)\|_\infty &= \max_{1\leq i\leq n} \sup_{t_0 \in \mathcal{N}_h(s_i)\cap\mathcal{T}} \left\| \nabla f_{t^*}((t_0)_{1:2}T^{-1}, (t_0)_3) \begin{pmatrix} T^{-1} & 0 \\ 0 & 1 \end{pmatrix} \right\|_\infty |T^{-1}| \\ &\leq \left\| \sup_{\tilde{t}^*_{1:2} \in (d,w^\intercal)B^* \pm hT^{-1}, \tilde{t}^*_3 \in \mathcal{T}_v} \frac{\partial f_{t^*}(\tilde{t}^*)}{\partial \tilde{t}^*} \right\|_\infty (1 + 2\|T^{-1}\|_{\max})|T^{-1}|.\end{aligned}$$

As T^{-1} has bounded norms, the interval $(d, w^\intercal)B^* \pm hT^{-1}$ is inside $\{(d, w^\intercal)B^*_{\cdot,1} \pm Ch\} \times \{(d, w^\intercal)B^*_{\cdot,2} \pm Ch\}$. As $\{(d, w^\intercal)B^*_{\cdot,1} \pm Ch\} \times \{(d, w^\intercal)B^*_{\cdot,2} \pm Ch\}$ is a subset of $[-a_1, a_1] \times [-a_2, a_2]$ and f_{t^*} is differentiable and Lipschitz in $\mathcal{T}^*$, we have

$$\max_{1\leq i\leq n} \sup_{t_0 \in \mathcal{N}_h(s_i)\cap\mathcal{T}} \|\nabla f_t(t_0)\|_\infty \leq C_3 < \infty.$$

The convexity of $\mathcal{T}_v = [-a_3, a_3]$ is obvious.

For Condition 4.3 (c), since the evaluation point $|(d, w^\intercal)B^*_{\cdot,1} \pm Ch| \leq a_1$ and $|(d, w^\intercal)B^*_{\cdot,2} \pm Ch| \leq a_2$, we know that $((d, w^\intercal)B + \Delta^\intercal, v)^\intercal \in \mathcal{T}$ for any $\Delta \in \mathbb{R}^2$ satisfying $\|\Delta\|_\infty \leq h$ and for any $v \in \mathcal{T}_v$. $\square$

Proposition D.2 (A sufficient condition for Condition 4.3 (b))**.** *Assume that v_i has a compact support $\mathcal{T}_v$. The function $q(\cdot,\cdot) : \mathbb{R}^2 \to [0, 1]$ is twice differentiable and its first two derivatives are uniformly bounded. The random variable $q(d\beta + w^\intercal\kappa, u_i)$ is away from zero and one at some point u_0 such that $f_u(u_0|w_i^\intercal\eta, v_i) > 0$ for any $v_i \in \mathcal{T}_v$. Moreover, assume that the conditional density $f_u(u|w^\intercal\eta, v)$ comes from a location-scale family such that*

$$f_u(u|w^\intercal\eta, v) = \frac{1}{\sigma(w^\intercal\eta, v)} f_0\left(\frac{u - \mu(w^\intercal\eta, v)}{\sigma(w^\intercal\eta, v)}\right),$$

where $f_0(\cdot)$, $\mu(w^\intercal\eta, v) = \mathbb{E}[u|w^\intercal\eta, v]$, and $\sigma^2(w^\intercal\eta, v) = Var(u|w^\intercal\eta, v)$ are all twice-differentiable and their first two derivatives are uniformly bounded. Then Condition 4.3 (b) holds true.

Proof of Proposition D.2. We first show that $g(s_i)$ is uniformly bounded away from zero and one. By (3) and (9),

$$g(s_i) = \mathbb{E}[y_i|d_i = d, w_i = w, v_i = v_i] = \int q(d\beta + w^\intercal\kappa, u_i) f_u(u_i|w^\intercal\eta, v_i) du_i.$$

Since $q(d\beta + w^\intercal\eta, u_i)$ is Lipschitz in u_i,

$$|q(d\beta + w^\intercal\eta, u_i) - q(d\beta + w^\intercal\eta, u_0)| \leq C|u_i - u_0|$$

for some constant C. Hence, for any

$$|u_i - u_0| \leq \frac{1 - q(d\beta + w^\intercal\eta, u_0)}{2C},$$

$$q(d\beta + w^\intercal\eta, u_i) \leq q(d\beta + w^\intercal\eta, u_0) + \frac{1 - q(d\beta + w^\intercal\eta, u_0)}{2} \leq c_1 < 1. \tag{D.2}$$

Therefore,

$$\int q(d\beta + w^\intercal\kappa, u_i) f_u(u_i|w^\intercal\eta, v_i) du_i \leq \int_{|u_i - u_0| > \frac{1 - q(d\beta + w^\intercal\eta, u_0)}{2C}} f_u(u_i|w^\intercal\eta, v_i) du_i$$
$$+ c_1 \int_{|u_i - u_0| \leq \frac{1 - q(d\beta + w^\intercal\eta, u_0)}{2C}} f_u(u_i|w^\intercal\eta, v_i) du_i,$$

where the last step is due to $q(\cdot) \leq 1$ and (D.2).

Because $f_u(u_0|w^\intercal\eta, v_i) > 0$ $\forall v_i \in \mathcal{T}_v$ and $\mathcal{T}_v$ is compact, there exists a constant c_0 such that $f_u(u_0|w^\intercal\eta, v_i) \geq c_0 > 0$ $\forall v \in \mathcal{T}_v$. Using the Lipschitz property of $f_u(u_i|w^\intercal\eta, v_i)$ in u_i, it is easy to show that

$$\int_{|u_i - u_0| \leq \frac{1 - q(d\beta + w^\intercal\eta, u_0)}{2C}} f_u(u_i|w^\intercal\eta, v_i) du_i \geq c_2 > 0$$

and hence

$$
\begin{aligned}
g(s_i) &= \int q(d\beta + w^\intercal \kappa, u_i) f_u(u_i|w^\intercal \eta, v_i) du_i \\
&\leq 1 - \int_{|u_i - u_0| \leq \frac{1-q(d\beta + w^\intercal \eta, u_0)}{2C}} f_u(u_i|w^\intercal \eta, v_i) du_i + c_1 \int_{|u_i - u_0| \leq \frac{1-q(d\beta + w^\intercal \eta, u_0)}{2C}} f_u(u_i|w^\intercal \eta, v_i) du_i \\
&\leq 1 - (1-c_1)c_2 < 1
\end{aligned}
$$

uniformly in v_i. Similarly one can show that $g(s_i)$ is bounded away from zero uniformly in s_i.

Next, we show the Lipschitz property of g. Let $s_i^* = ((d, w^\intercal)B^*, v_i)^\intercal$. We first show the Lipschitz property of g at s_i is implied by the Lipschitz property of g^* and s_i^*. For $M = 2$, T is invertible and

$$
\frac{\partial g(s_i)}{\partial s_i} = \frac{\partial g^*(s_i^*)}{\partial s_i} = \frac{\partial g^*(s_i^*)}{\partial s_i^*} \frac{\partial s_i^*}{\partial s_i} = \frac{\partial g^*(s_i^*)}{\partial s_i^*} T^{-1}.
$$

As the columns of B and B^* are normalized,

$$
\| \frac{\partial s_i^*}{\partial s_i} \|_2 \leq C < \infty.
$$

Same arguments hold for $\partial g(s_i)/\partial (s_i)_2$. Using the above arguments, we arrive at

$$
\| \frac{\partial g(s_i)}{\partial s_i} \|_2 \leq \| \frac{\partial g^*(s_i^*)}{\partial s_i} \|_2 C
$$

for some constant $C > 0$.

We are left to establish the Lipschitz property of g^* at s_i^*. Notice that $q((s_i^*)_1, u_i) f_u(u_i|(s_i^*)_2, (s_i^*)_3)$ is Lebesgue-integrable because $q(\cdot, \cdot) \in [0, 1]$ and $f_u(\cdot)$ is a density function. In addition, $\sup_{x \in \mathbb{R}^2} |q'(x)| \leq C < \infty$ and $C f_u(u_i|(s_i^*)_2, (s_i^*)_3)$ is Lebesgue-integrable with respect to u_i. Hence, we change the order of differentiation and integration to get that

$$
\frac{\partial g^*(s_i^*)}{\partial (s_i^*)_1} = \int \nabla q((s_i^*)_1, u_i) f_u(u_i|(s_i^*)_2, (s_i^*)_3) du_i
$$

and hence

$$
\sup_{s_i^*} | \frac{\partial g^*(s_i^*)}{\partial (s_i^*)_1} | \leq C < \infty.
$$

Similarly, we can show that

$$\sup_{s_i^*} \left| \frac{\partial^2 g^*(s_i^*)}{\{\partial (s_i^*)_1\}^2} \right| \leq C < \infty.$$

For the partial derivatives with respect to $((s_i^*)_2, (s_i^*)_3)$, by our assumption on $f_u(u|w^\intercal \eta, v)$, we can use change of variable to arrive at

$$g(s_i^*) = \int q((s_i^*)_1, \sigma_i x + \mu_i) f_0(x) dx,$$

where $\mu(w_i^\intercal \eta, v_i)$ is abbreviated as μ_i and $\sigma(w_i^\intercal \eta, v_i)$ is abbreviated as σ_i, and

$$\int f_0(x) dx = \int f_u(u|w_i^\intercal \eta, v_i) du = 1.$$

Using similar arguments as above, the conditions of Proposition D.1 imply that

$$|\frac{\partial g^*(s_i^*)}{\partial (w_i^\intercal \eta, v_i)}| \leq C \int (|x| + 1) f_0(x) dx \leq C' < \infty.$$

As a result, we can change the order of differentiation and integration to get

$$\sup_{s_i^*} \| \frac{\partial g^*(s_i^*)}{\partial (w_i^\intercal \eta, v_i)} \|_2 \leq C' < \infty.$$

Similarly, we can show that

$$\sup_{s_i^*} \left\| \frac{\partial^2 g^*(s_i^*)}{\{\partial (w_i^\intercal \eta, v_i)\}^{\otimes 2}} \right\|_2 \leq C'' < \infty \text{ and } \sup_{s_i^*} \left\| \frac{\partial^2 g^*(s_i^*)}{\partial (s_i^*)_1 \partial (w_i^\intercal \eta, v_i)} \right\|_2 \leq C'' < \infty.$$

□

Proposition D.3 (Second sufficient condition for Condition 4.3 (b)). *Assume that v_i has a compact support $\mathcal{T}_v$ and*

$$q(d\beta + w^\intercal \kappa, u_i) = \mathbb{1}(d\beta + w^\intercal \kappa + u_i \geq c)$$

for fixed some constant c. Then

$$g^*(s_i^*) = \mathbb{P}(u_i \geq c - d\beta - w^\intercal \kappa | w_i^\intercal \eta = w^\intercal \eta, v_i = v).$$

If g^* satisfies Condition 4.3(b), then g satisfies Condition 4.3(b).

Proof of Proposition D.3. The proof is obvious and is omitted here. □

E Further simulation results

We consider two more simulation scenarios.

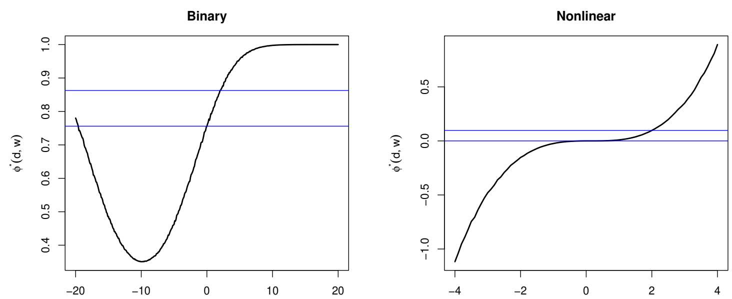

Figure E.1: The curves correspond to the functions $\phi^*(d, w)$ in scenarios (iii) and (iv) for $w = (0, \ldots, 0, 0.1)^\intercal \in \mathbb{R}^7$. The blue lines correspond to the true values for $d = -1$ and $d = 2$ in each scenario and the difference between two blue lines is CATE$(-1, 2|w)$.

E.1 Binary outcome model (iii)

(iii)

$$\mathbb{P}(y_i = 1|d_i, w_i, u_i) = \text{logit}\left(d_i\beta + w_i^\intercal\kappa + u_i + (d_i\beta + w_i^\intercal\kappa + u_i)^2/3\right),$$

with $\beta = 0.25$ and $\kappa = \eta = (0, 0, 0, 0, 0, 0.4, 0.2)^\intercal$. The unmeasured confounder u_i is generated as $u_i = \exp(0.25v_i + w_i^\intercal\eta) + \xi_i,\ \xi_i \sim U[-1, 1]$.

In Table E.1, we report the inference results CATE$(-1, 2|w)$ in binary outcome model (iii). The pattern is similar to that in Table 1 for binary outcome model (i).

E.2 The orthogonality assumption between strengths and direct effects of IVs

In this subsection, we report results in binary outcome models (i) and (iii) with $\kappa = \eta = (0, 0, 0, 0, 0, 0.4, -0.4)^\intercal$. All other parameters remain the same. Hence,

	n	c_γ	SpotIV MAE	COV	SE	MT	Valid-CF MAE	COV	SE	Oracle MAE	COV	SE
Normal	500	0.4	0.084	0.932	0.12	1	0.173	0.566	0.10	0.058	0.960	0.09
	500	0.6	0.070	0.916	0.07	0.98	0.151	0.442	0.07	0.045	0.978	0.07
	500	0.8	0.055	0.928	0.08	1	0.116	0.542	0.06	0.046	0.972	0.07
	1000	0.4	0.059	0.910	0.09	1	0.191	0.324	0.08	0.045	0.920	0.06
	1000	0.6	0.046	0.938	0.07	1	0.162	0.190	0.05	0.036	0.952	0.05
	1000	0.8	0.037	0.950	0.06	1	0.123	0.240	0.05	0.030	0.966	0.05
Unif	500	0.4	0.089	0.948	0.12	1	0.182	0.542	0.09	0.058	0.928	0.12
	500	0.6	0.064	0.928	0.09	1	0.154	0.406	0.07	0.041	0.960	0.07
	500	0.8	0.051	0.954	0.08	1	0.130	0.444	0.06	0.041	0.982	0.07
	1000	0.4	0.062	0.922	0.09	1	0.194	0.280	0.07	0.042	0.946	0.07
	1000	0.6	0.043	0.964	0.07	1	0.162	0.148	0.05	0.033	0.972	0.05
	1000	0.8	0.038	0.968	0.06	1	0.132	0.212	0.05	0.027	0.968	0.05

Table E.1: Inference for $\mathrm{CATE}(-1,2|w)$ in the binary outcome model (iii). The "MAE", "COV" and "SE" columns report the median absolute errors of $\widehat{\mathrm{CATE}}(-1,2|w)$, the empirical coverages of the confidence intervals and the average of estimated standard errors of the point estimators, respectively. The "MT" column reports the proportion of passing the majority rule testing in 500 replications. The columns indexed with "SpotIV" and "Valid-CF" correspond to the proposed method and the method assuming valid IVs, respectively. The columns indexed with "Oracle" correspond to the method which knows $\mathcal{V}$ as a priori.

$\gamma^\intercal\kappa = \gamma^\intercal\eta = 0$ in the current setting and the orthogonality assumption holds. The results are provided in Table E.2 and Table E.3.

E.3 Nonlinear outcome model (iv)

In this subsection, we examine the performance of our proposal when $p_\eta = 2$. Specifically, we set

$$y_i = (d_i\beta + z_i^T\kappa)u_i,$$

where $u_i = \exp(z_i^T\eta_1 + 0.25v_i) + z_i^T\eta_2 + v_i + \xi_i$ with $\xi_i \sim U[-1,1]$. We set $\beta = 0.25$, $\kappa = (0,0,0,0,0,0.4,0.2)^\intercal$, $\eta_1 = (0,0,0,0,0,0.4,0)^\intercal$, and $\eta_2 = (0,0,0,0,0,0,0.2)^\intercal$. The corresponding B^* and Θ^* are

$$B^* = \begin{pmatrix} 0.25 & 0 & 0 \\ \kappa & \eta_1 & \eta_2 \end{pmatrix} \quad \text{and} \quad \Theta^* = \begin{pmatrix} 0.25\gamma + \kappa & \eta_1 & \eta_2 \end{pmatrix}.$$

The results are reported in Table E.4.

			SpotIV				Valid-CF			Oracle		
	n	c_γ	MAE	COV	SE	MT	MAE	COV	SE	MAE	COV	SE
Normal	500	0.4	0.084	0.978	0.13	1	0.344	0.192	0.11	0.068	0.948	0.10
	500	0.6	0.063	0.982	0.11	1	0.269	0.138	0.08	0.059	0.948	0.09
	500	0.8	0.053	0.984	0.10	1	0.229	0.154	0.07	0.046	0.952	0.08
	1000	0.4	0.056	0.970	0.10	1	0.359	0.032	0.08	0.049	0.946	0.08
	1000	0.6	0.043	0.976	0.08	1	0.239	0.022	0.06	0.036	0.958	0.06
	1000	0.8	0.039	0.984	0.07	1	0.220	0.018	0.05	0.035	0.974	0.06
Unif	500	0.4	0.080	0.978	0.14	1	0.350	0.152	0.10	0.069	0.952	0.11
	500	0.6	0.068	0.988	0.11	1	0.270	0.128	0.08	0.060	0.954	0.09
	500	0.8	0.053	0.984	0.10	1	0.222	0.128	0.07	0.051	0.956	0.08
	1000	0.4	0.061	0.970	0.10	1	0.351	0.028	0.08	0.048	0.950	0.08
	1000	0.6	0.044	0.986	0.08	1	0.273	0.012	0.06	0.041	0.970	0.07
	1000	0.8	0.080	0.986	0.07	1	0.224	0.028	0.05	0.037	0.966	0.06

Table E.2: Inference for $\mathrm{CATE}(-1, 2|w)$ in the binary outcome model (i) when the orthogonality assumption holds. The "MAE", "COV" and "SE" columns report the median absolute errors of $\widehat{\mathrm{CATE}}(-1, 2|w)$, the empirical coverages of the confidence intervals and the average of estimated standard errors of the point estimators, respectively. The "MT" column reports the proportion of passing the majority rule testing in 500 replications. The columns indexed with "SpotIV" and "Valid-CF" correspond to the proposed method and the method assuming valid IVs, respectively. The columns indexed with "Oracle" correspond to the method which knows $\mathcal{V}$ as a priori.

	n	c_γ	SpotIV MAE	SpotIV COV	SpotIV SE	SpotIV Votes	Valid-CF MAE	Valid-CF COV	Valid-CF SE	Oracle MAE	Oracle COV	Oracle SE
Normal	500	0.4	0.069	0.984	0.12	1	0.298	0.208	0.10	0.057	0.954	0.09
	500	0.6	0.053	0.972	0.09	0.98	0.241	0.124	0.08	0.047	0.954	0.07
	500	0.8	0.053	0.984	0.10	1	0.298	0.208	0.10	0.046	0.952	0.08
	1000	0.4	0.053	0.980	0.09	1	0.304	0.048	0.08	0.045	0.940	0.07
	1000	0.6	0.043	0.976	0.08	1	0.251	0.028	0.06	0.036	0.958	0.06
	1000	0.8	0.037	0.978	0.06	1	0.202	0.024	0.05	0.033	0.956	0.05
Unif	500	0.4	0.069	0.982	0.12	1	0.301	0.164	0.10	0.066	0.942	0.09
	500	0.6	0.050	0.984	0.09	1	0.233	0.140	0.07	0.049	0.944	0.08
	500	0.8	0.044	0.994	0.08	1	0.301	0.164	0.10	0.041	0.962	0.07
	1000	0.4	0.054	0.982	0.09	1	0.201	0.164	0.10	0.047	0.934	0.07
	1000	0.6	0.041	0.986	0.07	1	0.233	0.140	0.05	0.038	0.948	0.06
	1000	0.8	0.031	0.984	0.06	1	0.198	0.134	0.07	0.031	0.960	0.05

Table E.3: Inference for CATE$(-1, 2|w)$ in the binary outcome model (iii) when the orthogonality assumption holds. The "MAE", "COV" and "SE" columns report the median absolute errors of $\widehat{\mathrm{CATE}}(-1, 2|w)$, the empirical coverages of the confidence intervals and the average of estimated standard errors of the point estimators, respectively. The "MT" column reports the proportion of passing the majority rule testing in 500 replications. The columns indexed with "SpotIV" and "Valid-CF" correspond to the proposed method and the method assuming valid IVs, respectively. The columns indexed with "Oracle" correspond to the method which knows $\mathcal{V}$ as a priori.

	n	c_γ	SpotIV MAE	SpotIV COV	SpotIV SE	SpotIV MT	TSHT MAE	TSHT COV	TSHT SE	Oracle MAE	Oracle COV	Oracle SE
Normal	500	0.4	0.079	0.984	0.17	1	0.268	0.167	0.08	0.093	0.960	0.16
	500	0.6	0.092	0.992	0.16	0.99	0.079	0.633	0.05	0.084	0.986	0.13
	500	0.8	0.087	0.968	0.15	1	0.065	0.587	0.04	0.071	0.968	0.12
	1000	0.4	0.084	1	0.15	1	0.049	0.843	0.06	0.081	0.928	0.12
	1000	0.6	0.074	0.986	0.12	1	0.046	0.720	0.04	0.056	0.966	0.09
	1000	0.8	0.064	0.950	0.10	1	0.049	0.550	0.03	0.052	0.958	0.08
Unif	500	0.4	0.093	0.997	0.18	1	0.305	0.190	0.07	0.104	0.963	0.19
	500	0.6	0.091	1	0.18	1	0.062	0.703	0.05	0.091	0.980	0.15
	500	0.8	0.091	0.980	0.16	1	0.067	0.543	0.04	0.071	0.968	0.13
	1000	0.4	0.054	0.982	0.09	1	0.054	0.797	0.05	0.047	0.934	0.07
	1000	0.6	0.041	0.986	0.07	1	0.052	0.637	0.03	0.038	0.948	0.06
	1000	0.8	0.031	0.984	0.06	1	0.050	0.527	0.03	0.031	0.960	0.05

Table E.4: Inference for CATE$(-1, 2|w)$ in the nonlinear outcome (iv). The "MAE", "COV" and "SE" columns report the median absolute errors of $\widehat{\mathrm{CATE}}(-1, 2|w)$, the empirical coverages of the confidence intervals and the average of estimated standard errors of the point estimators, respectively. The "MT" column reports the proportion of passing the majority rule testing in 500 replications. The columns indexed with "SpotIV" and "TSHT" correspond to the proposed method and the method assuming valid IVs, respectively. The columns indexed with "Oracle" correspond to the method which knows $\mathcal{V}$ as a priori.