Simulation results in the main text of the Double calibration

September 24, 2024

1 Case I: Sparse PS & Dense OR

First, we consider the case where the PS is a sparse logistic-linear model and the OR is a dense nonlinear model:

$$T|X = x \sim \text{Bernoulli}\left(\pi(x) = \psi(x^\top \gamma)\right),$$
$$Y|X = x, T = t \sim (r(x,t), 1),$$

where $\psi(z) \equiv 1/(1+\exp(-z))$ in this section. In particular, we choose γ to satisfy the following: $\|\gamma\|_2 = 1$, $\gamma_j \sim \text{Uniform}([1,2])$ for $j = 1, \cdots, s$ and $\gamma_{s+1} = \cdots = \gamma_p = 0$ so only the first s coordinates of γ are active. For the outcome model $r(x,t)$, we first transform the covariates X nonlinearly as follows:

$$\tilde{X}_1 \equiv \texttt{bs}(X_1, 100)^\top \cdot (1, 1/2, 1/3, \cdots, 1/100), \tilde{X}_2 \equiv 2/(1+\exp(-X_2)),$$
$$\tilde{X}_3 \equiv \exp(X_3/2), \tilde{X}_4 \equiv X_4/(1+\exp(X_3)), \tilde{X}_5 = X_4 X_5/10,$$

and then scale $\tilde{X}_2, \cdots, \tilde{X}_5$ to have zero mean and unit variance. Here $\texttt{bs}(x, 100)$ denotes the B-spline transformation of x with 100 degrees-of-freedom. We then define the outcome model $r(x,t)$ as

$$r(x,t) \equiv (|\tilde{x}_1 + \tilde{x}_2 - \tilde{x}_3/2 + \tilde{x}_4/3 - \tilde{x}_5/4| + 0.05)^{-1} + x^\top \beta - t,$$

where $\|\beta\|_2 \equiv 1$ and $\beta_j \propto j^{-1}$ for $j = 1, \cdots, p$.

The simulation results are displayed in Figure 1 and Table 1.

Tables 1 below summarize the coverage probabilities and the lengths of 95% confidence intervals, Mean/Median Absolute Biases, standard errors, and Root-Mean-Squared-Errors (RMSE) of different estimators for the Sparse PS & Dense OR simulation setting from the main text.

Table 1: Mean Absolute Bias (MAB†), Median Absolute Bias (MAB‡), standard error (std), root-mean-squared error (RMSE), Coverage probability of 95% confidence intervals (Coverage), length of 95% confidence intervals (CI length) and Computation time (in minutes) for estimating τ under the *sparse PS & dense OR* setting from the main text.

Setting	measure	g-formula	IPW	AIPW	TMLE	ARB	hdCBPS	RCAL	RCAL*	DCal
	MAB†	1.34	1.06	0.98	0.70	0.53	0.78	0.78	0.80	0.47
	MAB‡	1.39	1.08	1.00	0.71	0.50	0.80	0.80	0.82	0.41
$s = 10$,	std	0.02	0.58	0.31	0.67	0.49	0.50	0.32	0.33	0.53
$p = 400$,										
$n = 200$										

Table 1 continued from previous page

	RMSE	1.38	1.09	1.03	0.79	0.64	0.88	0.87	0.88	0.60
	CP	0.00%	57.00%	15.00%	90.5%	92.00%	53.50%	32.00%	32.00%	96.00%
	Length	0.06	2.28	1.21	2.64	1.92	1.98	1.27	1.29	2.09
	Time						0.08	0.62	0.16	0.02
	MAB†	1.35	1.02	0.92	0.56	0.48	0.60	0.77	0.79	0.40
	MAB‡	1.41	1.02	0.93	0.54	0.44	0.56	0.80	0.80	0.33
$s = 10$,	std	0.01	0.45	0.25	0.53	0.37	0.46	0.24	0.25	0.43
$p = 800$,	RMSE	1.39	1.05	0.97	0.63	0.57	0.74	0.84	0.85	0.50
$n = 400$	CP	0.50%	28.50%	8.50%	93.5%	82.00%	60.50%	15.50%	16.00%	91.50%
	Length	0.05	1.77	0.97	2.06	1.45	1.81	0.94	0.97	1.69
	Time						0.10	4.33	0.97	0.06
	MAB†	1.11	0.91	0.70	0.32	0.40	0.32	0.61	0.64	0.33
	MAB‡	1.11	0.93	0.71	0.28	0.39	0.23	0.59	0.58	0.30
$s = 10$,	std	0.02	0.35	0.19	0.41	0.27	0.44	0.20	0.19	0.33
$p = 1000$,	RMSE	1.15	0.93	0.75	0.39	0.47	0.44	0.67	0.71	0.40
$n = 800$	CP	0.00%	18.00%	11.50%	97.5%	72.00%	88.00%	22.00%	15.00%	89.00%
	Length	0.09	1.38	0.75	1.62	1.05	1.71	0.77	0.73	1.28
	Time						0.64	60.77	6.84	0.16
	MAB†	0.87	0.81	0.51	0.20	0.37	0.20	0.45	0.45	0.34
	MAB‡	0.86	0.82	0.51	0.18	0.37	0.14	0.43	0.45	0.34
$s = 10$,	std	0.03	0.29	0.15	0.30	0.19	0.34	0.15	0.15	0.24
$p = 1000$,	RMSE	0.89	0.83	0.55	0.25	0.41	0.26	0.50	0.51	0.38
$n = 1600$	CP	0.00%	15.50%	18.00%	99.5%	49.00%	98.50%	24.12%	25.00%	75.50%
	Length	0.10	1.12	0.59	1.19	0.73	1.33	0.58	0.59	0.93
	Time						3.37	437.60	22.48	0.36
	MAB†	1.44	1.15	1.08	0.66	0.51	0.84	0.84	0.88	0.47
	MAB‡	1.48	1.15	1.11	0.62	0.48	0.87	0.82	0.87	0.36
$s = 20$,	std	0.01	0.56	0.29	0.96	0.52	0.49	0.32	0.32	0.56
$p = 400$,	RMSE	1.49	1.20	1.15	0.77	0.62	0.96	0.95	1.00	0.59
$n = 200$	CP	0.00%	38.00%	11.50%	98.5%	95.00%	47.00%	32.50%	30.50%	96.00%
	Length	0.05	2.21	1.15	3.75	2.04	1.92	1.27	1.25	2.18
	Time						0.10	0.59	0.15	0.02
	MAB†	1.36	1.06	0.94	0.40	0.40	0.59	0.70	0.72	0.34
	MAB‡	1.38	1.05	0.94	0.35	0.36	0.53	0.69	0.74	0.29
$s = 20$,	std	0.02	0.41	0.22	0.71	0.38	0.43	0.23	0.24	0.42
$p = 800$,	RMSE	1.39	1.09	0.99	0.50	0.49	0.71	0.79	0.80	0.44
$n = 400$	CP	0.00%	15.00%	6.50%	97.5%	92.00%	64.00%	22.00%	23.00%	93.00%
	Length	0.06	1.61	0.87	2.79	1.49	1.68	0.92	0.93	1.65
	Time						0.20	5.15	1.01	0.06
	MAB†	1.26	0.98	0.80	0.29	0.33	0.36	0.62	0.64	0.28
	MAB‡	1.27	0.98	0.81	0.25	0.29	0.27	0.59	0.60	0.23
$s = 20$,	std	0.02	0.33	0.18	0.58	0.28	0.41	0.17	0.18	0.34
$p = 1000$,	RMSE	1.28	1.00	0.84	0.38	0.40	0.47	0.68	0.70	0.35
$n = 800$	CP	0.00%	10.00%	7.50%	100.00%	85.50%	82.50%	12.00%	13.50%	95.50%

Table 1 continued from previous page

	Length	0.08	1.29	0.72	2.28	1.11	1.61	0.68	0.69	1.34
	Time						0.30	43.52	7.18	0.16
	MAB†	1.03	0.89	0.61	0.32	0.26	0.25	0.46	0.48	0.22
	MAB‡	1.00	0.90	0.60	0.29	0.24	0.21	0.46	0.46	0.19
$s = 20$,	std	0.02	0.26	0.14	0.46	0.20	0.34	0.14	0.14	0.25
$p = 1000$,	RMSE	1.05	0.91	0.65	0.38	0.30	0.31	0.52	0.53	0.27
$n = 1600$	CP	0.00%	7.50%	6.00%	99.5%	82.50%	92.50%	20.92%	19.50%	93.50%
	Length	0.09	1.03	0.56	1.79	0.77	1.32	0.57	0.56	0.97
	Time						1.87	513.17	24.26	0.34
	MAB†	1.24	1.09	1.00	0.60	0.51	0.71	0.77	0.79	0.47
	MAB‡	1.25	1.09	1.04	0.56	0.45	0.67	0.78	0.81	0.40
$s = 50$,	std	0.02	0.55	0.27	1.26	0.52	0.45	0.32	0.32	0.54
$p = 400$,	RMSE	1.29	1.13	1.07	0.74	0.62	0.84	0.87	0.89	0.59
$n = 200$	CP	0.50%	43.00%	10.00%	99%	94.00%	62.00%	31.00%	29.50%	96.50%
	Length	0.09	2.15	1.08	4.96	2.04	1.77	1.26	1.26	2.11
	Time						0.06	0.45	0.11	0.02
	MAB†	1.14	0.98	0.85	0.42	0.37	0.44	0.58	0.61	0.34
	MAB‡	1.17	0.98	0.87	0.36	0.31	0.36	0.58	0.63	0.27
$s = 50$,	std	0.02	0.41	0.21	0.97	0.38	0.38	0.24	0.24	0.41
$p = 800$,	RMSE	1.17	1.01	0.89	0.54	0.45	0.55	0.66	0.68	0.43
$n = 400$	CP	0.00%	17.00%	7.00%	99.5%	94.00%	79.50%	37.00%	31.50%	95.50%
	Length	0.10	1.60	0.82	3.82	1.50	1.48	0.93	0.93	1.61
	Time						0.05	4.81	0.88	0.06
	MAB†	1.06	0.92	0.73	0.40	0.25	0.32	0.49	0.51	0.23
	MAB‡	1.07	0.93	0.73	0.32	0.20	0.30	0.49	0.50	0.18
$s = 50$,	std	0.03	0.31	0.16	0.78	0.28	0.30	0.18	0.18	0.32
$p = 1000$,	RMSE	1.08	0.94	0.76	0.51	0.31	0.38	0.54	0.56	0.29
$n = 800$	CP	0.00%	8.50%	4.00%	100%	93.50%	89.50%	23.00%	20.00%	97.50%
	Length	0.11	1.22	0.63	3.05	1.11	1.16	0.69	0.70	1.25
	Time						0.19	48.06	6.01	0.16
	MAB†	0.94	0.84	0.60	0.48	0.19	0.26	0.45	0.46	0.18
	MAB‡	0.94	0.88	0.61	0.45	0.16	0.25	0.47	0.48	0.16
$s = 50$,	std	0.03	0.25	0.13	0.62	0.20	0.23	0.13	0.13	0.24
$p = 1000$,	RMSE	0.95	0.86	0.62	0.57	0.23	0.30	0.48	0.49	0.22
$n = 1600$	CP	0.00%	7.00%	5.50%	98%	94.50%	85.50%	13.50%	12.00%	98.50%
	Length	0.11	0.99	0.51	2.44	0.80	0.91	0.51	0.51	0.94
	Time						0.62	328.59	26.79	0.34

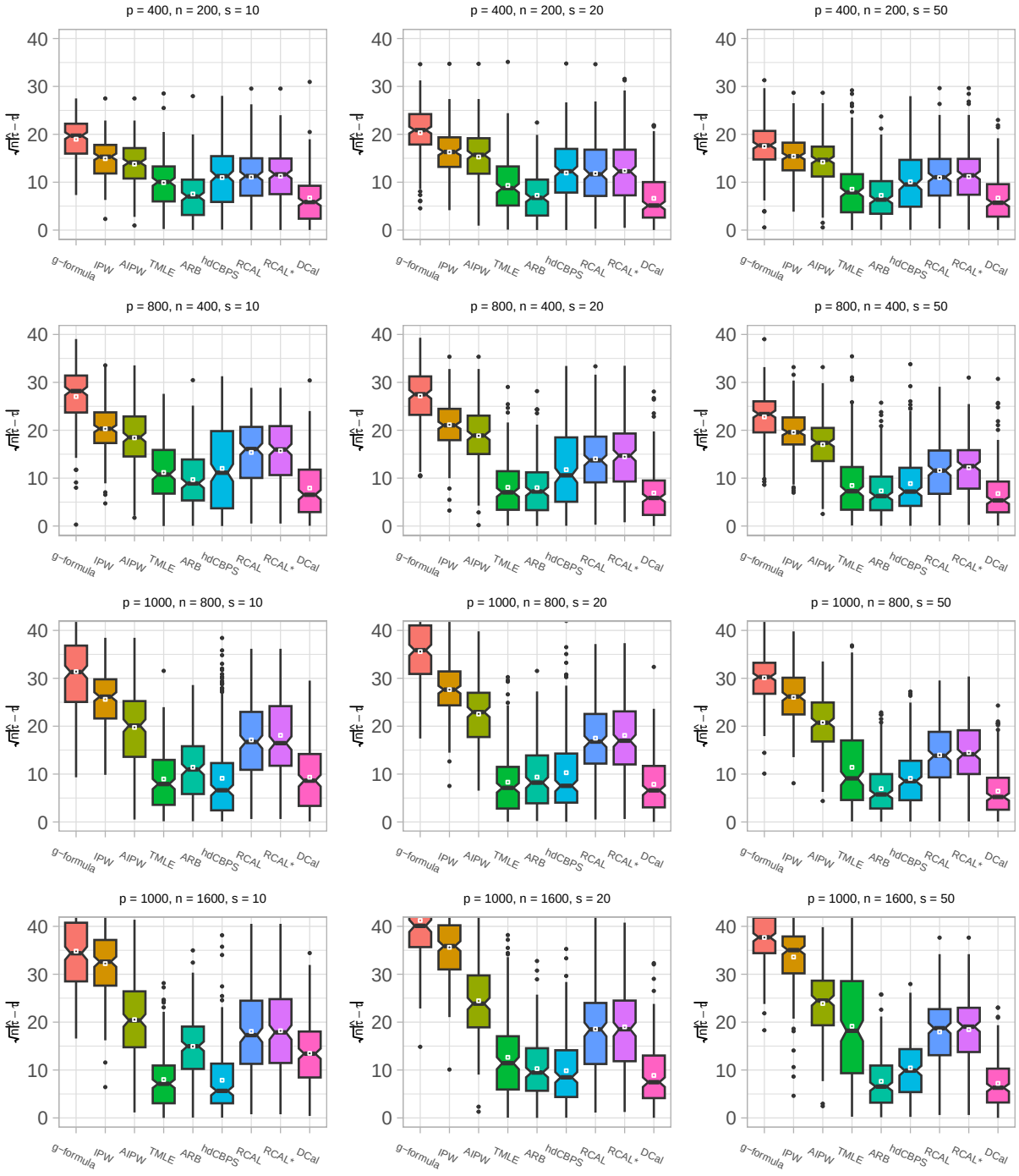

Figure 1: Boxplots of the $\sqrt{n}$-scaled estimation error $\sqrt{n}|\hat{\tau} - \tau|$ under different sample size n, dimension p and sparsity level s under the *Sparse PS & Dense OR* setting. The white dots correspond to mean.

2 Case II: Dense PS & Sparse OR

Next, we consider the case where the PS is a dense nonlinear model bounded between 0 and 1 and the OR is a sparse linear model:

$$T|X = x \sim \text{Bernoulli}\left(\pi(x)\right),$$
$$Y|X = x, T = t \sim \left(r(x,t), 1\right).$$

Here we take the PS model to be

$$\pi(X) \equiv \min\{\max\{\psi(\check{X}_1 - \frac{1}{2}\check{X}_2 + \frac{1}{4}\check{X}_3 - \frac{1}{8}\check{X}_4 + \gamma^\top X), 0.05\}, 0.95\},$$

where $\check{X}_1 := e^{0.5X_1}$, $\check{X}_2 := 10 + X_2/(1+e^{X_1})$, $\check{X}_3 := (0.05X_1X_3 + 0.6)^2$, $\check{X}_4 := (X_2 + X_4 + 10)^2$, and $\gamma_j \propto j^{-1}$ with $\|\gamma\|_2 \equiv 2$. We take the OR model to be

$$r(X,t) \equiv \beta^\top X - 1 + 2t$$

where we first draw $\beta_j \sim \text{Uniform}([1,2])$ for $j = 1, \cdots, s$ and set $\beta_j \equiv 0$ for $j = s+1, \cdots, p$. We eventually normalize β such that $\|\beta\|_2 = 1$.

The simulation results are displayed in Figure 2, and Table 2.

Table 2: Mean Absolute Bias (MAB†), Median Absolute Bias (MAB‡), standard error (std), root-mean-squared error (RMSE), Coverage probability of 95% confidence intervals (Coverage), length of 95% confidence intervals (CI length) and Computation time (in minutes) for estimating τ under the *dense PS & sparse OR* setting from the main text.

Setting	**measure**	**g-formula**	**IPW**	**AIPW**	**TMLE**	**ARB**	**hdCBPS**	**RCAL**	**RCAL***	**DCal**
	MAB†	0.47	1.03	0.30	0.20	0.18	0.19	0.29	0.29	0.17
	MAB‡	0.48	1.05	0.27	0.17	0.15	0.15	0.27	0.29	0.14
$s = 10$,	std	0.14	0.56	0.21	0.25	0.18	0.32	0.20	0.20	0.26
$p = 400$,	RMSE	0.51	1.08	0.35	0.26	0.22	0.25	0.34	0.34	0.21
$n = 200$	CP	17.50%	37.50%	66.00%	87.00%	87.50%	94.00%	68.50%	71.50%	99.00%
	Length	0.56	2.21	0.81	0.99	0.71	1.25	0.77	0.77	1.03
	Time						1.40	0.45	0.11	0.02
	MAB†	0.35	0.73	0.18	0.13	0.11	0.13	0.18	0.18	0.13
	MAB‡	0.35	0.75	0.17	0.11	0.08	0.10	0.15	0.16	0.09
$s = 10$,	std	0.11	0.69	0.19	0.17	0.13	0.24	0.14	0.14	0.23
$p = 800$,	RMSE	0.37	0.82	0.22	0.16	0.14	0.17	0.21	0.21	0.18
$n = 400$	CP	13.50%	56.00%	83.00%	91.00%	92.00%	98.50%	80.50%	80.00%	98.00%
	Length	0.42	2.69	0.73	0.66	0.50	0.94	0.56	0.56	0.90
	Time						3.10	2.71	0.55	0.07
	MAB†	0.28	0.52	0.12	0.12	0.07	0.09	0.11	0.11	0.10
	MAB‡	0.28	0.49	0.11	0.11	0.06	0.07	0.10	0.10	0.09
$s = 10$,	std	0.08	0.84	0.16	0.13	0.09	0.16	0.11	0.11	0.21
$p = 1000$,	RMSE	0.29	0.62	0.15	0.14	0.08	0.11	0.13	0.14	0.13
$n = 800$	CP	7.00%	79.00%	92.50%	83.00%	97.50%	99.50%	84.00%	85.00%	100.00%

Table 2 continued from previous page

	Length	0.31	3.31	0.64	0.50	0.36	0.64	0.41	0.41	0.82
	Time						7.47	24.41	2.72	0.19
	MAB†	0.24	0.42	0.09	0.10	0.05	0.06	0.07	0.07	0.07
	MAB‡	0.24	0.33	0.08	0.09	0.04	0.05	0.06	0.06	0.06
$s = 10$,	std	0.06	0.89	0.14	0.07	0.06	0.12	0.08	0.08	0.18
$p = 1000$,	RMSE	0.25	0.53	0.12	0.12	0.06	0.08	0.09	0.09	0.09
$n = 1600$	CP	2.00%	98.50%	98.00%	66.50%	96.50%	99.50%	90.67%	89.50%	100.00%
	Length	0.23	3.48	0.56	0.28	0.25	0.46	0.30	0.30	0.70
	Time						22.58	254.20	5.99	0.44
	MAB†	0.47	1.24	0.31	0.22	0.16	0.21	0.29	0.30	0.16
	MAB‡	0.46	1.27	0.28	0.20	0.14	0.17	0.27	0.27	0.13
$s = 20$,	std	0.19	0.72	0.25	0.25	0.18	0.34	0.23	0.23	0.30
$p = 400$,	RMSE	0.51	1.30	0.36	0.28	0.21	0.27	0.34	0.35	0.21
$n = 200$	CP	31.50%	41.50%	76.00%	83.50%	90.00%	94.00%	78.50%	77.50%	99.00%
	Length	0.73	2.82	0.98	0.98	0.71	1.34	0.92	0.92	1.17
	Time						2.06	0.46	0.12	0.02
	MAB†	0.40	0.90	0.20	0.16	0.12	0.13	0.21	0.21	0.13
	MAB‡	0.39	0.97	0.19	0.14	0.10	0.10	0.20	0.20	0.11
$s = 20$,	std	0.14	0.83	0.21	0.18	0.13	0.25	0.17	0.17	0.25
$p = 800$,	RMSE	0.42	1.00	0.24	0.20	0.15	0.17	0.25	0.25	0.16
$n = 400$	CP	13.50%	56.00%	84.50%	84.00%	92.00%	98.00%	78.00%	81.00%	98.50%
	Length	0.54	3.24	0.81	0.70	0.51	0.98	0.66	0.66	0.99
	Time						4.52	2.67	0.51	0.07
	MAB†	0.30	0.65	0.13	0.12	0.07	0.09	0.14	0.14	0.10
	MAB‡	0.29	0.63	0.12	0.11	0.06	0.06	0.13	0.13	0.08
$s = 20$,	std	0.10	0.98	0.18	0.12	0.09	0.17	0.12	0.12	0.23
$p = 1000$,	RMSE	0.31	0.77	0.16	0.15	0.08	0.11	0.16	0.17	0.14
$n = 800$	CP	11.50%	74.50%	95.50%	79.50%	98.50%	98.00%	85.43%	85.00%	100.00%
	Length	0.40	3.86	0.71	0.49	0.37	0.68	0.48	0.49	0.90
	Time						10.19	24.30	2.36	0.21
	MAB†	0.24	0.46	0.09	0.11	0.05	0.06	0.08	0.08	0.08
	MAB‡	0.24	0.35	0.07	0.10	0.04	0.05	0.07	0.08	0.07
$s = 20$,	std	0.07	1.05	0.15	0.07	0.06	0.13	0.09	0.09	0.19
$p = 1000$,	RMSE	0.24	0.58	0.12	0.12	0.06	0.08	0.10	0.10	0.10
$n = 1600$	CP	6.50%	96.00%	98.50%	59.50%	95.50%	100.00%	92.46%	92.50%	100.00%
	Length	0.29	4.11	0.59	0.28	0.25	0.49	0.35	0.35	0.75
	Time						27.20	248.58	5.03	0.41
	MAB†	0.50	1.39	0.38	0.28	0.22	0.28	0.34	0.34	0.20
	MAB‡	0.48	1.43	0.36	0.28	0.20	0.25	0.34	0.33	0.17
$s = 50$,	std	0.23	0.79	0.28	0.24	0.18	0.31	0.27	0.27	0.32
$p = 400$,	RMSE	0.54	1.45	0.42	0.33	0.27	0.34	0.39	0.39	0.25
$n = 200$	CP	44.50%	43.50%	77.50%	70.00%	80.50%	92.00%	83.50%	79.50%	97.00%
	Length	0.89	3.11	1.09	0.92	0.71	1.22	1.05	1.05	1.24

Table 2 continued from previous page

	Time						2.08	0.47	0.12	0.02
	MAB†	0.39	1.04	0.22	0.18	0.13	0.14	0.23	0.24	0.12
	MAB‡	0.39	1.07	0.21	0.15	0.12	0.12	0.22	0.22	0.11
$s = 50$,	std	0.17	0.86	0.23	0.16	0.13	0.25	0.19	0.19	0.27
$p = 800$,	RMSE	0.41	1.13	0.26	0.22	0.15	0.18	0.27	0.27	0.16
$n = 400$	CP	32.50%	54.00%	87.00%	76.50%	91.00%	98.50%	87.00%	85.00%	99.50%
	Length	0.65	3.37	0.89	0.64	0.51	0.97	0.76	0.76	1.05
	Time						4.55	2.67	0.51	0.08
	MAB†	0.32	0.75	0.14	0.14	0.08	0.09	0.17	0.17	0.11
	MAB‡	0.31	0.72	0.13	0.13	0.06	0.08	0.15	0.16	0.09
$s = 50$,	std	0.12	0.99	0.19	0.12	0.09	0.18	0.14	0.14	0.24
$p = 1000$,	RMSE	0.33	0.88	0.17	0.17	0.10	0.12	0.19	0.19	0.14
$n = 800$	CP	20.00%	71.50%	93.00%	72.00%	91.00%	99.00%	86.00%	85.50%	100.00%
	Length	0.48	3.87	0.74	0.46	0.37	0.70	0.55	0.55	0.93
	Time						10.09	24.24	2.26	0.18
	MAB†	0.24	0.44	0.09	0.11	0.05	0.08	0.09	0.09	0.08
	MAB‡	0.24	0.38	0.08	0.10	0.04	0.07	0.09	0.09	0.06
$s = 50$,	std	0.09	1.06	0.15	0.07	0.06	0.13	0.10	0.10	0.20
$p = 1000$,	RMSE	0.24	0.55	0.12	0.13	0.06	0.10	0.11	0.11	0.09
$n = 1600$	CP	14.50%	96.50%	99.00%	59.00%	92.50%	99.00%	92.00%	92.50%	100.00%
	Length	0.35	4.17	0.61	0.27	0.25	0.51	0.40	0.40	0.77
	Time						27.19	244.39	4.58	0.43

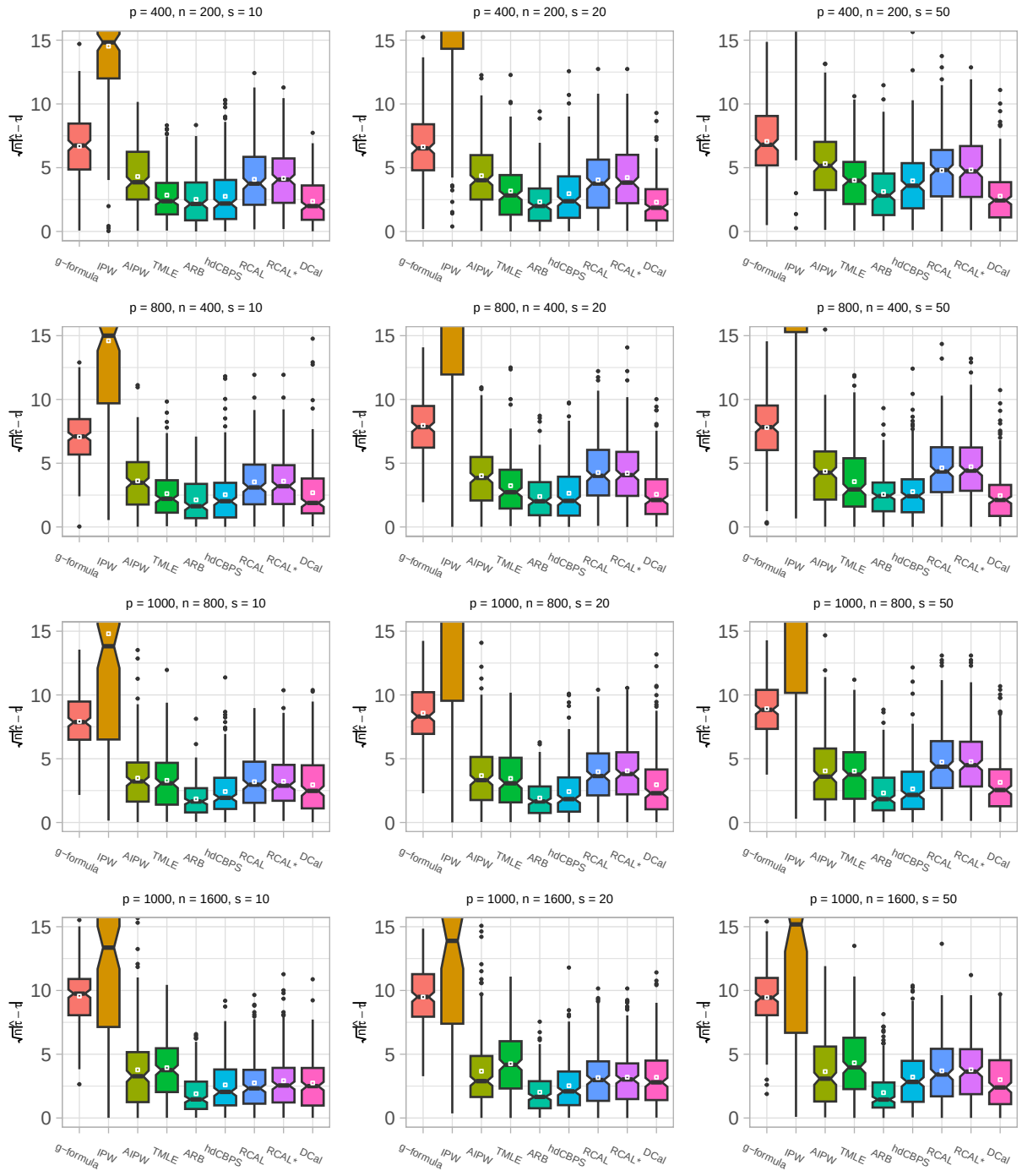

Figure 2: Boxplots of the $\sqrt{n}$-scaled estimation error $\sqrt{n}|\hat{\tau} - \tau|$ under different sample size n, dimension p and sparsity level s under the *Dense PS & Sparse OR* setting. The white dots correspond to mean.