 DOI: 10.1111/biom.13683

Supporting Information for "Reader Reaction to " Outcome-adaptive lasso: Variable selection for causal inference " by Shortreed and Ertefaie (2017) " by

Ismaila Baldé [1,*], Yi Archer Yang [2,], Geneviève Lefebvre [1,***]**

[1] Département de mathématiques, Université du Québec à Montréal

[2] Department of Mathematics and Statistics, McGill University

*email: balde.ismaila@courrier.uqam.ca

**email: archer.yang@mcgill.ca

***email: lefebvre.gen@uqam.ca

Web Appendix A. Adaptive elastic net

Elastic net performs simultaneous regularization and variable selection, with a penalty that is a convex combination of the lasso and ridge penalties (Zou and Hastie, 2005). Compared to lasso which selects at most n variables, elastic net is viewed to be better equiped to tackle the $p \gg n$ problem and to deal with grouped selection of correlated variables. Indeed, with correlated features, lasso is known to exhibit variable selection problems as it tends to select only one variable from the group (Zou and Hastie, 2005). The adaptive elastic net method was specifically introduced to analyse high-dimensional data (Zou and Zhang, 2009). It is a mixture of adaptive lasso and elastic net, inheriting good properties from each: the oracle property from the former and the ability to overcome the collinearity problem from the latter (Zou and Zhang, 2009; Ghosh, 2011). Adaptive elastic net has also shown advantages in practice, for example in high-dimensional cancer classification for simultaneous estimation and gene selection (Algamal and Lee, 2015).

Consider a standard linear regression problem along with nonnegative tuning parameters (λ_1, λ_2). Ghosh (2011) defined the adaptive elastic net estimator as

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right), \tag{A1}$$

where $\tilde{w}_j = \left|\tilde{\beta}_j\right|^{-\gamma}$ such that $\gamma > 0$ and $\tilde{\beta}$ is a root n-consistent estimator of β.

Similar to elastic net (Zou and Hastie, 2005), Ghosh (2011) showed the equivalence between adaptive elastic net (that is, Equation (A1)) and an ordinary adaptive lasso problem in some augmented space (see Web Appendix A for the proof):

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y^* - \mathbf{X}^*\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| \right), \tag{A2}$$

where $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $Y^* = \begin{pmatrix} Y \\ 0_p \end{pmatrix}$.

Consequently, $\mathbf{X}^*$ has rank p and adaptive elastic net can potentially select all p predictors when either $p < n$ or $p \geqslant n$.

In what follows, we recall the data augmentation proof of the adaptive elastic net (Ghosh, 2011). Let $\mathbf{I}_p$ be a $p \times p$ identity matrix and $0_p = (0, 0, \ldots, 0)^T \in \mathbb{R}^p$.

For any fixed λ_2 we have:

$$\left\| Y - \mathbf{X}\beta \right\|_2^2 + \lambda_2 \left\| \beta \right\|_2^2 = \left\| Y - \mathbf{X}\beta \right\|_2^2 + \left\| 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \right\|_2^2$$
$$= \left(Y - \mathbf{X}\beta \right)^T \left(Y - \mathbf{X}\beta \right) + \left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \right)^T \left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \right)$$
$$= \begin{pmatrix} Y - \mathbf{X}\beta \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \end{pmatrix}^T \begin{pmatrix} Y - \mathbf{X}\beta \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \end{pmatrix} = \left\| \begin{matrix} Y - \mathbf{X}\beta \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \end{matrix} \right\|_2^2$$
$$= \left\| Y^* - \mathbf{X}^*\beta \right\|_2^2, \quad Y^* = \begin{pmatrix} Y \\ 0_p \end{pmatrix}, \mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}.$$

Thus Equation (A1) becomes Equation (A2)

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y^* - \mathbf{X}^*\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| \right).$$

Web Appendix B. Data augmentation for the proposed penalized iteratively re-weighted least squares GOAL (GOALi)

The GOAL estimator is defined as

$$\hat{\alpha}(GOAL) = \arg\min_{\alpha} \left[\ell_n(\alpha; A, \mathbf{X}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right],$$

where $\hat{w}_j = \left| \hat{\beta}_j^{ols} \right|^{-\gamma}$ such that $\gamma > 1$ and $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A, \beta)} \|Y - \beta_A A - \mathbf{X}\beta\|_2^2$.

The Newton-Raphson update solution of GOAL is obtained as

$$\hat{\alpha}_{PIRLS}(GOAL) = \arg\min_{\alpha} \left[\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right], \qquad \text{(B1)}$$

where $\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) = \frac{1}{2} \sum_{i=1}^{n} t_i \left(z_i - x_i^T \alpha \right)^2, \quad \tilde{p}(x_i) = \frac{1}{1+\exp(-x_i^T \tilde{\alpha})},$

$t_i = \tilde{p}(x_i)[1 - \tilde{p}(x_i)], \quad z_i = x_i^T \tilde{\alpha} + \frac{a_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}, \quad Z = (z_1, \ldots, z_n)^T, \quad \mathbf{T} = \text{diag}(t_1, \ldots, t_n).$

The data augmentation step of the proposed GOAL with PIRLS method (GOALi) is obtained as follows. Let $0_{n \times p}$ be a $n \times p$ null matrix. For any fixed λ_2, we have:

$$\left\| \mathbf{T}^{1/2}(Z-\mathbf{X}\alpha) \right\|_2^2 + \lambda_2 \left\| \alpha \right\|_2^2 = \left\| \mathbf{T}^{1/2}(Z-\mathbf{X}\alpha) \right\|_2^2 + \left\| \mathbf{I}_p^{1/2}\left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha\right) \right\|_2^2$$

$$= \left\| \mathbf{T}^{1/2}(Z-\mathbf{X}\alpha) + 0_{n\times p}\left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha\right) \right\|_2^2 + \left\| 0_{n\times p}^T(Z-\mathbf{X}\alpha) + \mathbf{I}_p^{1/2}\left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha\right) \right\|_2^2$$

$$= \left\| \left(\mathbf{T}^{1/2}, 0_{n\times p}\right) \begin{pmatrix} Z-\mathbf{X}\alpha \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha \end{pmatrix} \right\|_2^2 + \left\| \left(0_{n\times p}^T, \mathbf{I}_p^{1/2}\right) \begin{pmatrix} Z-\mathbf{X}\alpha \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha \end{pmatrix} \right\|_2^2$$

$$= \left\| \begin{pmatrix} \mathbf{T}^{1/2} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I}_p^{1/2} \end{pmatrix} \begin{pmatrix} Z-\mathbf{X}\alpha \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha \end{pmatrix} \right\|_2^2 = \left\| \mathbf{T}^{*1/2}(Z^* - \mathbf{X}^*\alpha) \right\|_2^2,$$

$$Z^* = \begin{pmatrix} Z \\ 0_p \end{pmatrix}, \quad A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}, \quad \mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix} \text{ and } \quad \mathbf{T}^{*1/2} = \begin{pmatrix} \mathbf{T}^{1/2} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I}_p^{1/2} \end{pmatrix}.$$

Thus Equation (B1) becomes

$$\hat{\alpha}_Q(GOAL) = \arg\min_{\alpha} \left[\ell_{Q^*}(\alpha; A^*, \mathbf{X}^*, Z^*, \mathbf{T}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right],$$

where

$$\ell_{Q^*}(\alpha; A^*, \mathbf{X}^*, Z^*, \mathbf{T}^*) = \left\| \mathbf{T}^{*1/2}(Z^* - \mathbf{X}^*\alpha) \right\|_2^2.$$

Web Appendix C. Simulation Results

Web Table C1 presents the bias, standard error (SE) and mean squared error (MSE) for OAL and GOAL estimators under all scenarios in the high-dimensional settings ($p/n = 100/200, 200/500$).

[Table 1 about here.]

Web Table C2 presents the bias, SE and MSE for OAL and GOAL estimators under all scenarios in the low-dimensional settings ($n = 200, 500, 1000$ and $p = 20$).

[Table 2 about here.]

Web Figure C1 displays the bias, SE and MSE for OAL and GOAL estimators under Scenarios 3 and 4 in the high-dimensional settings ($p/n = 100/200, 200/500$).

[Figure 1 about here.]

Web Figures C2, C3, C4 and C5 present the box plots of ATE estimates for OAL and GOAL estimators under all scenarios in the high-dimensional settings ($p/n = 100/200, 200/500$) for $\rho = 0$, 0.2, 0.5 and 0.75, respectively.

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

Web Figures C6, C7, C8 and C9 present the box plots of ATE estimates for OAL and GOAL estimators under all scenarios in the low-dimensional settings ($n = 200, 500, 1000$ with $p = 20$) for $\rho = 0$, 0.2, 0.5 and 0.75, respectively.

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

Web Figures C10-C13 present the wAMD (weighted absolute mean difference) between exposure groups for OAL and GOAL estimators over 1000 simulations. Results for combination ($n = 200$, $p = 100$) are displayed in Web Figures C10 and C11 for $\rho = 0$ and 0.75, respectively, while those for the combination ($n = 200$, $p = 20$) appear in Web Figures C12 and C13 for $\rho = 0$ and 0.75, respectively.

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

Web Figures C14-C15 show the proportion of times each covariate was selected over 1000 simulations for inclusion in the PS model for combinations ($n = 200$, $p = 100$) and ($n = 200$, $p = 20$) with $\rho = 0, 0.75$. In the high-dimensional setting ($n = 200$, $p = 100$), all estimators (OAL, GOALn and GOALi) included confounders and predictors of the outcome at similar rates. In this setting and for both correlation values ($\rho = 0, 0.75$), GOALn excluded more pure predictors of the exposure and spurious covariates than OAL. In all scenarios, GOALi included more of these variables than OAL when $\rho = 0$, while the same phenomenon was only observed in Scenario 4 when $\rho = 0.75$. In the low-dimensional setting with $\rho = 0$, OAL and GOAL included all covariates at very similar rates. For the low-dimensional setting with $\rho = 0.75$, GOAL included more or slightly more pure predictors of the exposure and spurious covariates than OAL. Moreover, the confounders and the predictors of the outcome were selected by OAL and GOAL at similar rates, except for Scenario 3 for which GOAL noticeably selected more confounders than OAL. In both high- and low-dimensional settings, the number of covariates selected by the estimators was greater, on average, when $\rho = 0.75$ than when $\rho = 0$. Moreover, for a fixed ρ value, the number of covariates selected by OAL and GOAL was far greater in the high-dimensional setting as compared to the low-dimensional setting.

[Figure 14 about here.]

[Figure 15 about here.]

References

Algamal, Z. Y. and Lee, M. H. (2015). Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in Biology and Medicine* **67**, 136–145.

Ghosh, S. (2011). On the grouped selection and model complexity of the adaptive elastic net. *Statistics and Computing* **21**, 451–462.

Shortreed, S. M. and Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics* **73(4)**, 1111–1122. http://doi.org/10.1111/biom.12679.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B* **67**, 301–320.

Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics* **37**, 1733–1751.

Table C1: Bias (SE; **MSE**) of the IPTW estimator for the average treatment effect (ATE) for OAL, GOALn and GOALi with ratios $p/n = 100/200, 200/500$ under Scenarios 1, 2, 3 and 4 by sections 1, 2, 3 and 4, respectively (results based on 1000 estimates of the ATE).

	$\frac{p}{n}$		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.75$
1	$\frac{100}{200}$	OAL	0.08 (0.21; **0.05**)	0.22 (0.31; **0.14**)	0.43 (0.51; **0.45**)	0.73 (0.69; **1.01**)
		GOALn	0.05 (0.22; **0.05**)	0.00 (0.30; **0.09**)	-0.08 (0.42; **0.18**)	-0.02 (0.49; **0.24**)
		GOALi	0.07 (0.21; **0.05**)	0.10 (0.28; **0.09**)	0.00 (0.40; **0.16**)	0.01 (0.48; **0.23**)
	$\frac{200}{500}$	OAL	0.04 (0.13; **0.02**)	0.12 (0.23; **0.07**)	0.29 (0.39; **0.24**)	0.41 (0.59; **0.51**)
		GOALn	0.01 (0.14; **0.02**)	-0.10 (0.23; **0.06**)	-0.13 (0.31; **0.11**)	-0.13 (0.35; **0.14**)
		GOALi	0.03 (0.13; **0.02**)	0.01 (0.20; **0.04**)	-0.04 (0.28; **0.08**)	-0.10 (0.34; **0.13**)
2	$\frac{100}{200}$	OAL	0.02 (0.18; **0.03**)	0.08 (0.24; **0.06**)	0.18 (0.36; **0.16**)	0.36 (0.52; **0.40**)
		GOALn	-0.01 (0.18; **0.03**)	-0.08 (0.25; **0.07**)	-0.16 (0.32; **0.13**)	-0.09 (0.35; **0.13**)
		GOALi	0.02 (0.18; **0.03**)	0.02 (0.23; **0.05**)	-0.04 (0.31; **0.10**)	-0.04 (0.34; **0.12**)
	$\frac{200}{500}$	OAL	0.00 (0.11; **0.01**)	0.04 (0.16; **0.03**)	0.11 (0.26; **0.08**)	0.20 (0.37; **0.18**)
		GOALn	-0.02 (0.11; **0.01**)	-0.10 (0.17; **0.04**)	-0.18 (0.25; **0.10**)	-0.15 (0.24; **0.08**)
		GOALi	0.00 (0.11; **0.01**)	-0.01 (0.15; **0.02**)	-0.04 (0.20; **0.04**)	-0.09 (0.24; **0.07**)
3	$\frac{100}{200}$	OAL	0.09 (0.20; **0.05**)	0.16 (0.26; **0.09**)	0.30 (0.40; **0.25**)	0.48 (0.52; **0.50**)
		GOALn	0.12 (0.20; **0.06**)	0.02 (0.29; **0.08**)	-0.05 (0.39; **0.15**)	-0.05 (0.46; **0.21**)
		GOALi	0.09 (0.21; **0.05**)	0.10 (0.27; **0.08**)	0.03 (0.38; **0.14**)	-0.01 (0.45; **0.21**)
	$\frac{200}{500}$	OAL	0.04 (0.13; **0.02**)	0.07 (0.19; **0.04**)	0.20 (0.31; **0.14**)	0.29 (0.46; **0.29**)
		GOALn	0.06 (0.15; **0.03**)	-0.07 (0.21; **0.05**)	-0.13 (0.33; **0.13**)	-0.15 (0.34; **0.14**)
		GOALi	0.04 (0.13; **0.02**)	0.01 (0.19; **0.03**)	-0.01 (0.27; **0.07**)	-0.09 (0.35; **0.13**)
4	$\frac{100}{200}$	OAL	0.05 (0.19; **0.04**)	0.19 (0.30; **0.12**)	0.52 (0.52; **0.55**)	1.03 (0.70; **1.54**)
		GOALn	0.01 (0.20; **0.04**)	-0.06 (0.32; **0.11**)	-0.12 (0.49; **0.26**)	-0.03 (0.57; **0.33**)
		GOALi	0.05 (0.19; **0.04**)	0.06 (0.29; **0.09**)	-0.01 (0.46; **0.21**)	-0.04 (0.57; **0.33**)
	$\frac{200}{500}$	OAL	0.02 (0.11; **0.01**)	0.14 (0.22; **0.07**)	0.42 (0.41; **0.34**)	0.68 (0.69; **0.93**)
		GOALn	-0.02 (0.12; **0.01**)	-0.11 (0.23; **0.07**)	-0.17 (0.36; **0.16**)	-0.20 (0.46; **0.25**)
		GOALi	0.01 (0.11; **0.01**)	0.00 (0.20; **0.04**)	-0.04 (0.34; **0.11**)	-0.17 (0.44; **0.23**)

Table C2: Bias (SE; **MSE**) of the IPTW estimator for the average treatment effect (ATE) for OAL, GOALn and GOALi with fixed $p = 20$ and increasing $n = 200, 500, 1000$ under Scenarios 1, 2, 3 and 4 by sections 1, 2, 3 and 4, respectively (results based on 1000 estimates of the ATE).

	n		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.75$
1		OAL	0.04 (0.19; **0.04**)	0.15 (0.28; **0.10**)	0.41 (0.44; **0.36**)	0.65 (0.59; **0.78**)
	200	GOALn	0.01 (0.18; **0.03**)	0.03 (0.25; **0.06**)	0.03 (0.33; **0.11**)	0.02 (0.44; **0.20**)
		GOALi	0.01 (0.18; **0.03**)	0.02 (0.24; **0.06**)	0.03 (0.32; **0.10**)	0.02 (0.42; **0.18**)
		OAL	0.03 (0.12; **0.02**)	0.12 (0.17; **0.04**)	0.31 (0.32; **0.20**)	0.54 (0.47; **0.51**)
	500	GOALn	0.01 (0.12; **0.01**)	0.02 (0.15; **0.02**)	0.01 (0.22; **0.05**)	0.01 (0.29; **0.08**)
		GOALi	0.01 (0.12; **0.01**)	0.02 (0.15; **0.02**)	0.00 (0.22; **0.05**)	0.01 (0.29; **0.08**)
		OAL	0.01 (0.08; **0.01**)	0.09 (0.13; **0.02**)	0.26 (0.26; **0.14**)	0.46 (0.40; **0.37**)
	1000	GOALn	0.00 (0.08; **0.01**)	0.01 (0.10; **0.01**)	-0.01 (0.17; **0.03**)	-0.01 (0.23; **0.05**)
		GOALi	0.00 (0.08; **0.01**)	0.01 (0.10; **0.01**)	-0.01 (0.17; **0.03**)	-0.01 (0.23; **0.05**)
2		OAL	0.00 (0.15; **0.02**)	0.04 (0.20; **0.04**)	0.18 (0.30; **0.12**)	0.35 (0.44; **0.32**)
	200	GOALn	-0.01 (0.15; **0.02**)	-0.01 (0.19; **0.04**)	-0.01 (0.24; **0.06**)	0.00 (0.31; **0.10**)
		GOALi	-0.01 (0.15; **0.02**)	0.00 (0.19; **0.04**)	-0.01 (0.24; **0.06**)	0.01 (0.30; **0.09**)
		OAL	0.01 (0.11; **0.01**)	0.03 (0.12; **0.01**)	0.11 (0.21; **0.06**)	0.25 (0.31; **0.16**)
	500	GOALn	0.01 (0.11; **0.01**)	0.00 (0.11; **0.01**)	-0.02 (0.16; **0.02**)	-0.01 (0.20; **0.04**)
		GOALi	0.01 (0.11; **0.01**)	0.00 (0.11; **0.01**)	-0.02 (0.16; **0.02**)	-0.01 (0.20; **0.04**)
		OAL	0.00 (0.07; **0.00**)	0.02 (0.08; **0.01**)	0.10 (0.15; **0.03**)	0.19 (0.25; **0.10**)
	1000	GOALn	0.00 (0.07; **0.00**)	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.14; **0.02**)
		GOALi	0.00 (0.07; **0.00**)	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.14; **0.02**)
3		OAL	0.04 (0.19; **0.04**)	0.10 (0.25; **0.07**)	0.25 (0.32; **0.17**)	0.40 (0.44; **0.35**)
	200	GOALn	0.03 (0.19; **0.04**)	0.03 (0.25; **0.06**)	0.04 (0.31; **0.10**)	0.01 (0.41; **0.17**)
		GOALi	0.03 (0.19; **0.04**)	0.03 (0.24; **0.06**)	0.04 (0.30; **0.09**)	0.03 (0.40; **0.16**)
		OAL	0.01 (0.12; **0.01**)	0.07 (0.16; **0.03**)	0.19 (0.27; **0.11**)	0.33 (0.34; **0.22**)
	500	GOALn	0.01 (0.12; **0.01**)	0.01 (0.15; **0.02**)	-0.01 (0.23; **0.05**)	0.01 (0.28; **0.08**)
		GOALi	0.01 (0.12; **0.01**)	0.01 (0.15; **0.02**)	-0.01 (0.23; **0.05**)	0.01 (0.28; **0.08**)
		OAL	0.00 (0.08; **0.01**)	0.04 (0.12; **0.02**)	0.16 (0.21; **0.07**)	0.29 (0.28; **0.16**)
	1000	GOALn	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.18; **0.03**)	0.00 (0.22; **0.05**)
		GOALi	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.17; **0.03**)	0.00 (0.22; **0.05**)
4		OAL	0.01 (0.17; **0.03**)	0.14 (0.26; **0.09**)	0.46 (0.44; **0.41**)	0.92 (0.62; **1.23**)
	200	GOALn	0.00 (0.17; **0.03**)	0.03 (0.24; **0.06**)	0.03 (0.36; **0.13**)	0.03 (0.52; **0.28**)
		GOALi	0.00 (0.17; **0.03**)	0.03 (0.24; **0.06**)	0.03 (0.36; **0.13**)	0.02 (0.51; **0.26**)
		OAL	0.02 (0.11; **0.01**)	0.09 (0.15; **0.03**)	0.38 (0.31; **0.24**)	0.77 (0.52; **0.86**)
	500	GOALn	0.01 (0.11; **0.01**)	0.01 (0.14; **0.02**)	0.01 (0.23; **0.06**)	0.00 (0.35; **0.12**)
		GOALi	0.01 (0.11; **0.01**)	0.01 (0.14; **0.02**)	0.01 (0.23; **0.05**)	0.00 (0.35; **0.12**)
		OAL	0.01 (0.07; **0.01**)	0.08 (0.12; **0.02**)	0.34 (0.25; **0.18**)	0.72 (0.42; **0.70**)
	1000	GOALn	0.00 (0.07; **0.01**)	0.01 (0.10; **0.01**)	0.00 (0.17; **0.03**)	-0.01 (0.27; **0.08**)
		GOALi	0.00 (0.07; **0.01**)	0.01 (0.10; **0.01**)	0.00 (0.17; **0.03**)	-0.01 (0.27; **0.08**)

Figure C1: Absolute bias (circle), standard error (square) and mean squared error (triangle) of IPTW estimator for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 3 and 4 (based on 1000 IPTW estimates). The ratios $p/n = 100/200$, $200/500$ are presented in rows 1 and 2, respectively, for Scenario 3 and in rows 3 and 4, respectively, for Scenario 4.

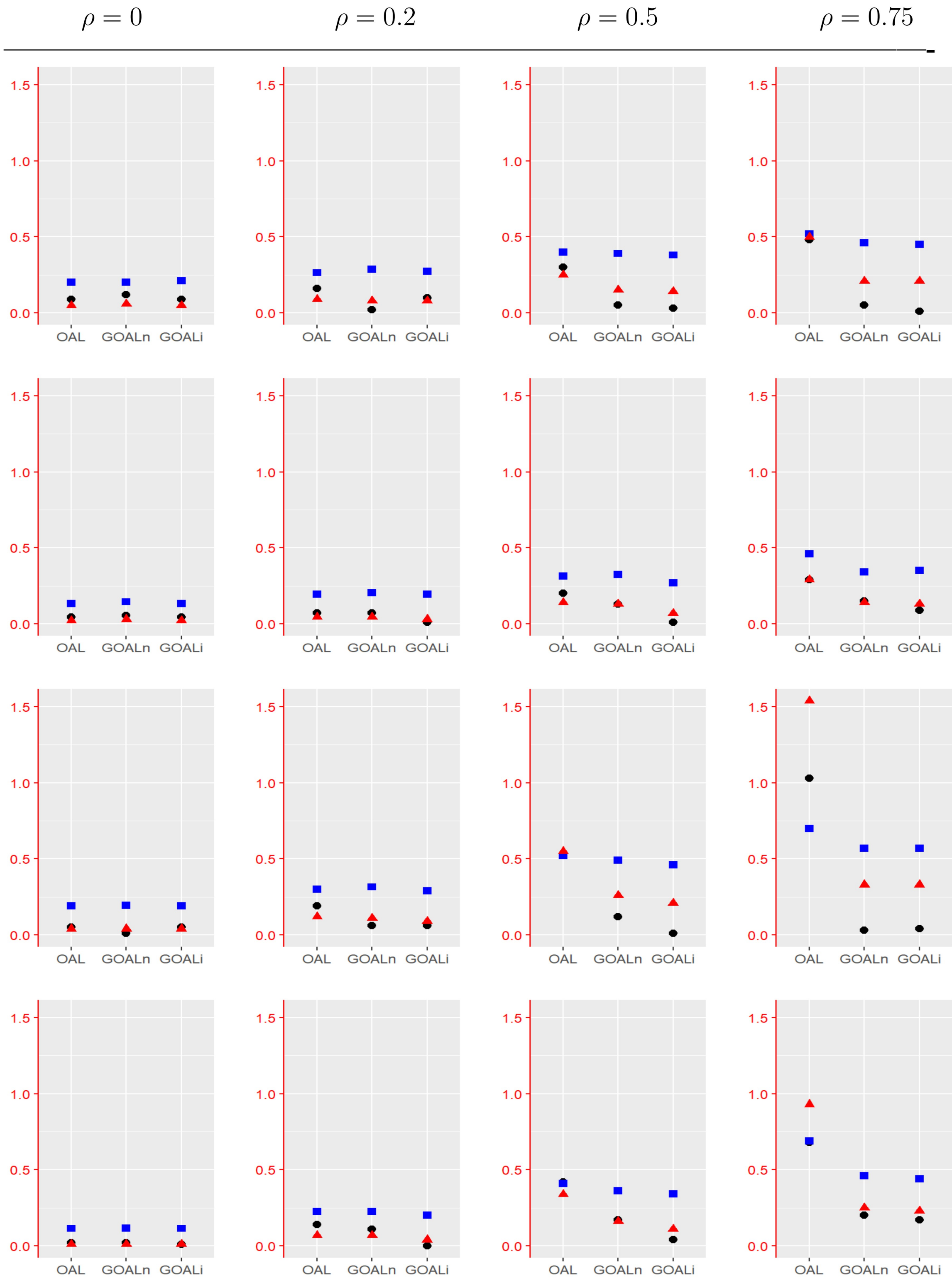

Figure C2: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0$. The true value of ATE is indicated with dotted line (ATE=0).

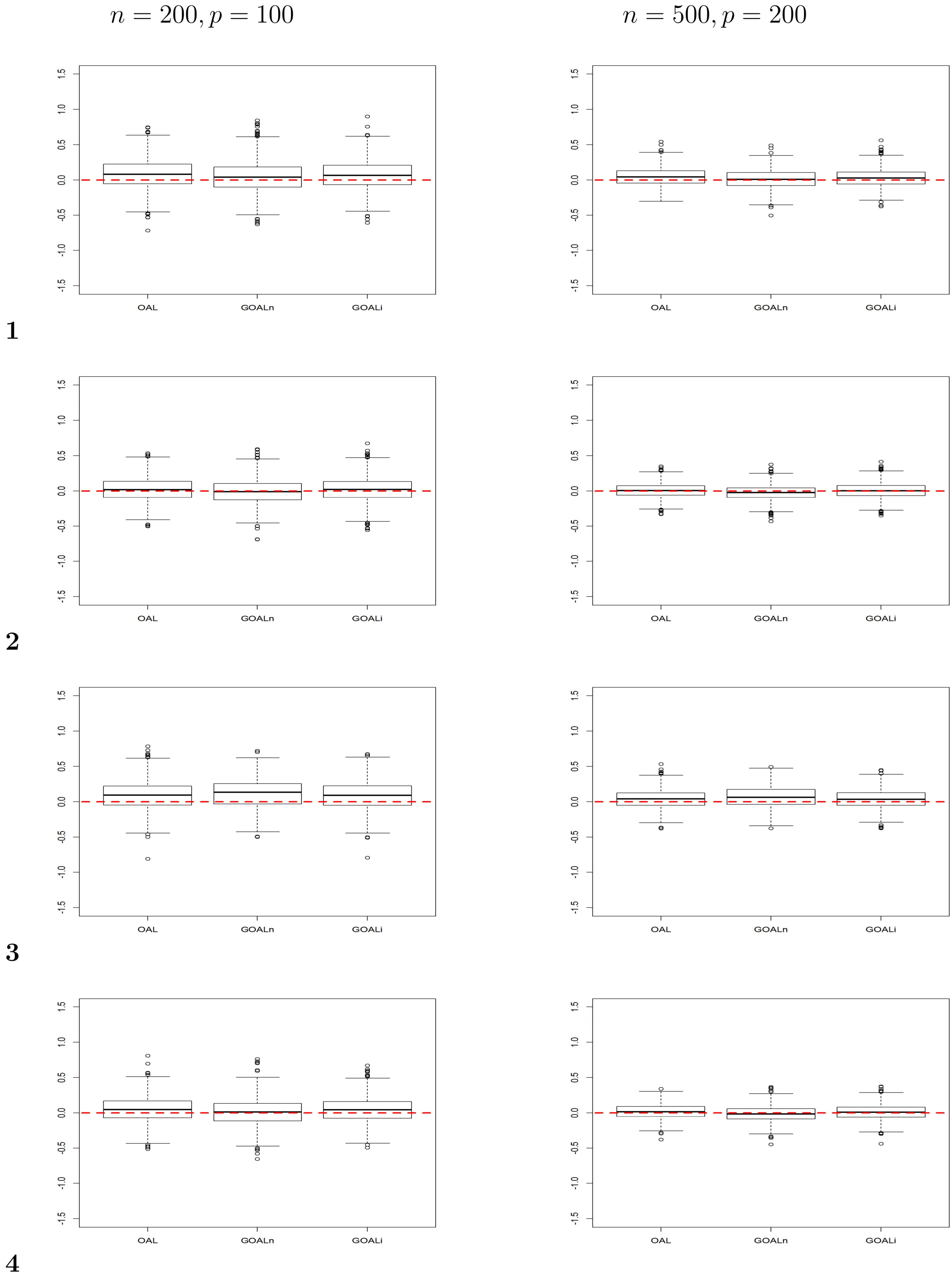

Figure C3: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0.2$. The true value of ATE is indicated with dotted line (ATE=0).

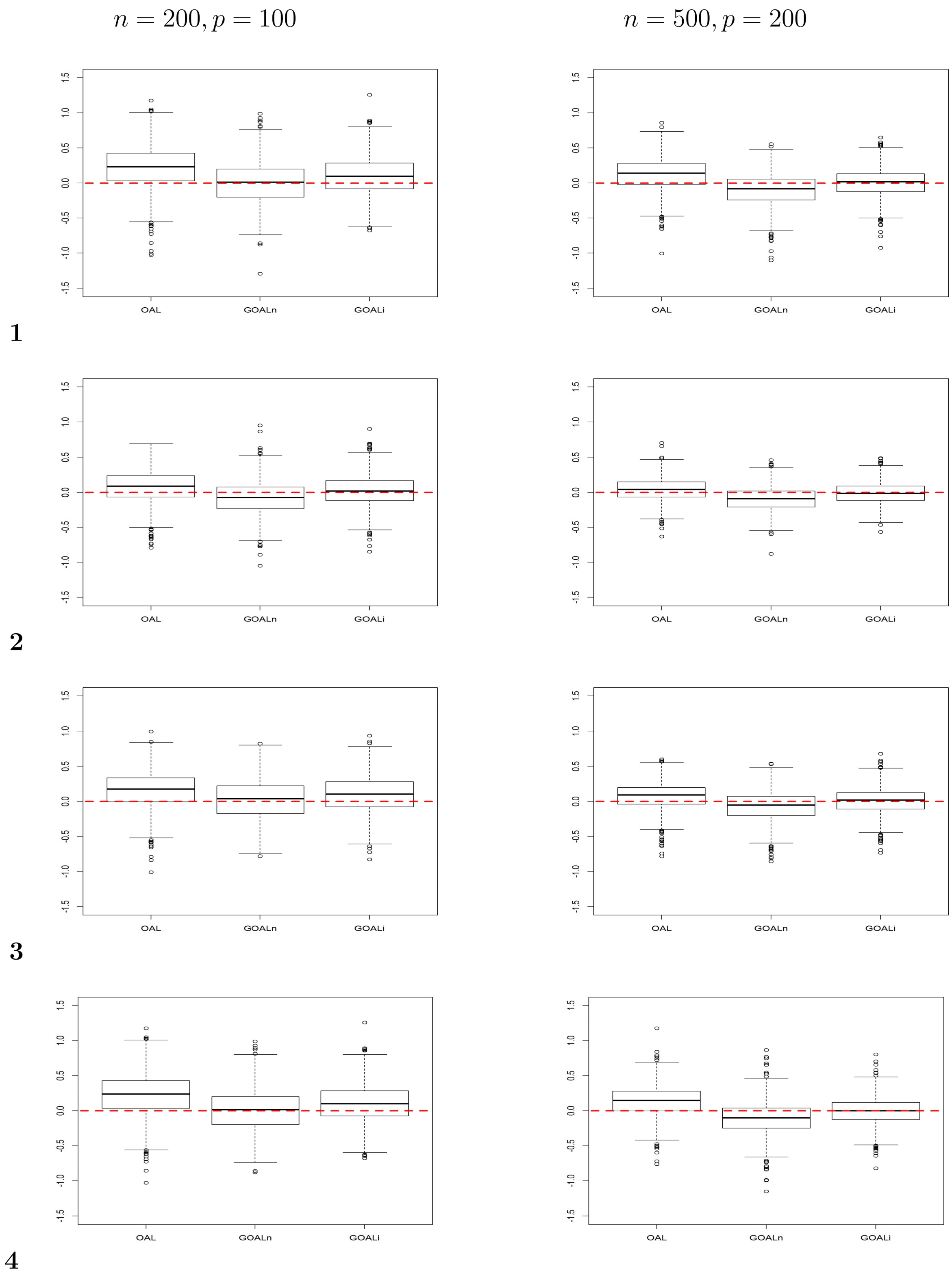

Figure C4: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0.5$. The true value of ATE is indicated with dotted line (ATE=0).

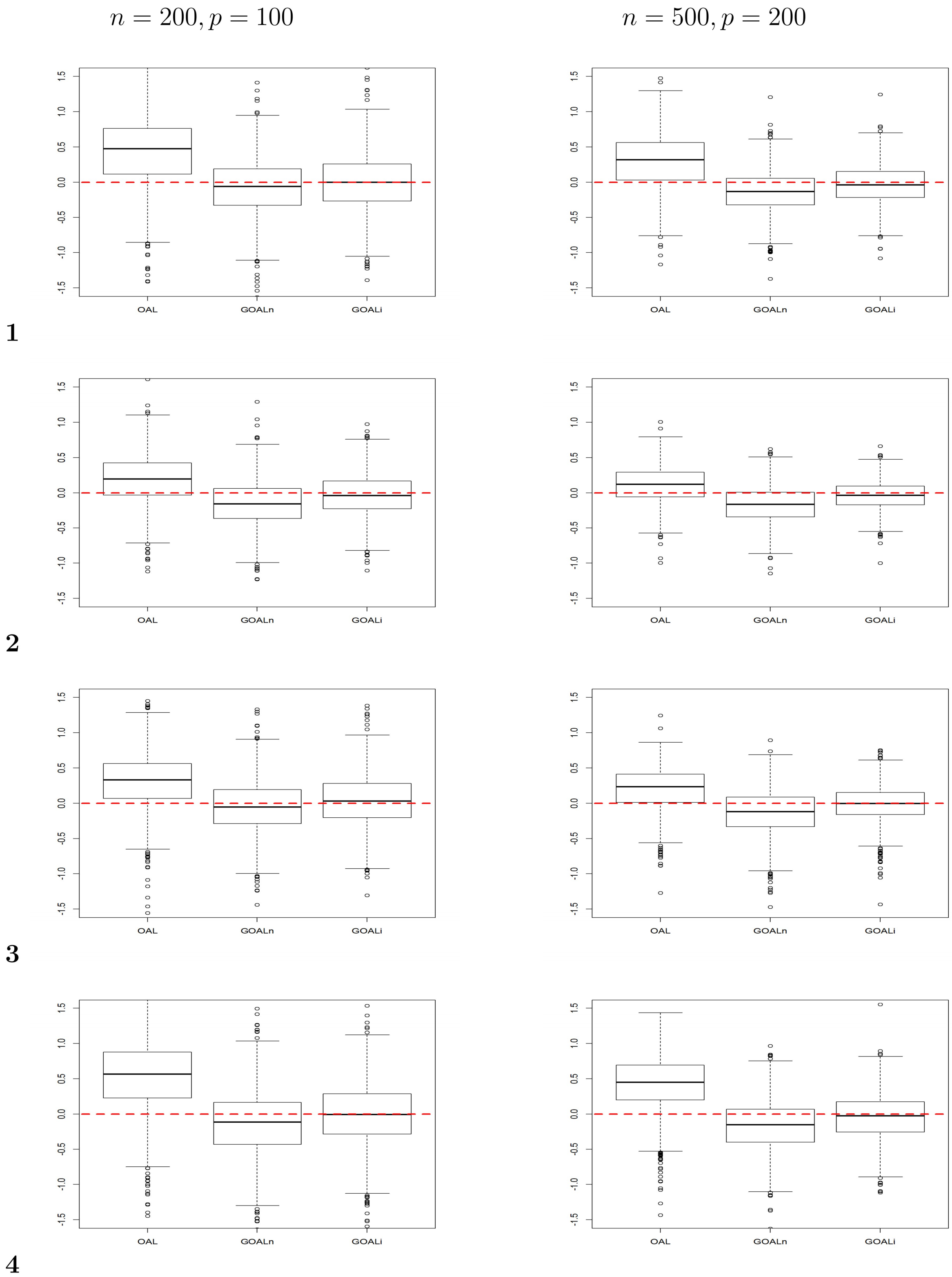

Figure C5: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0.75$. The true value of ATE is indicated with dotted line (ATE=0).

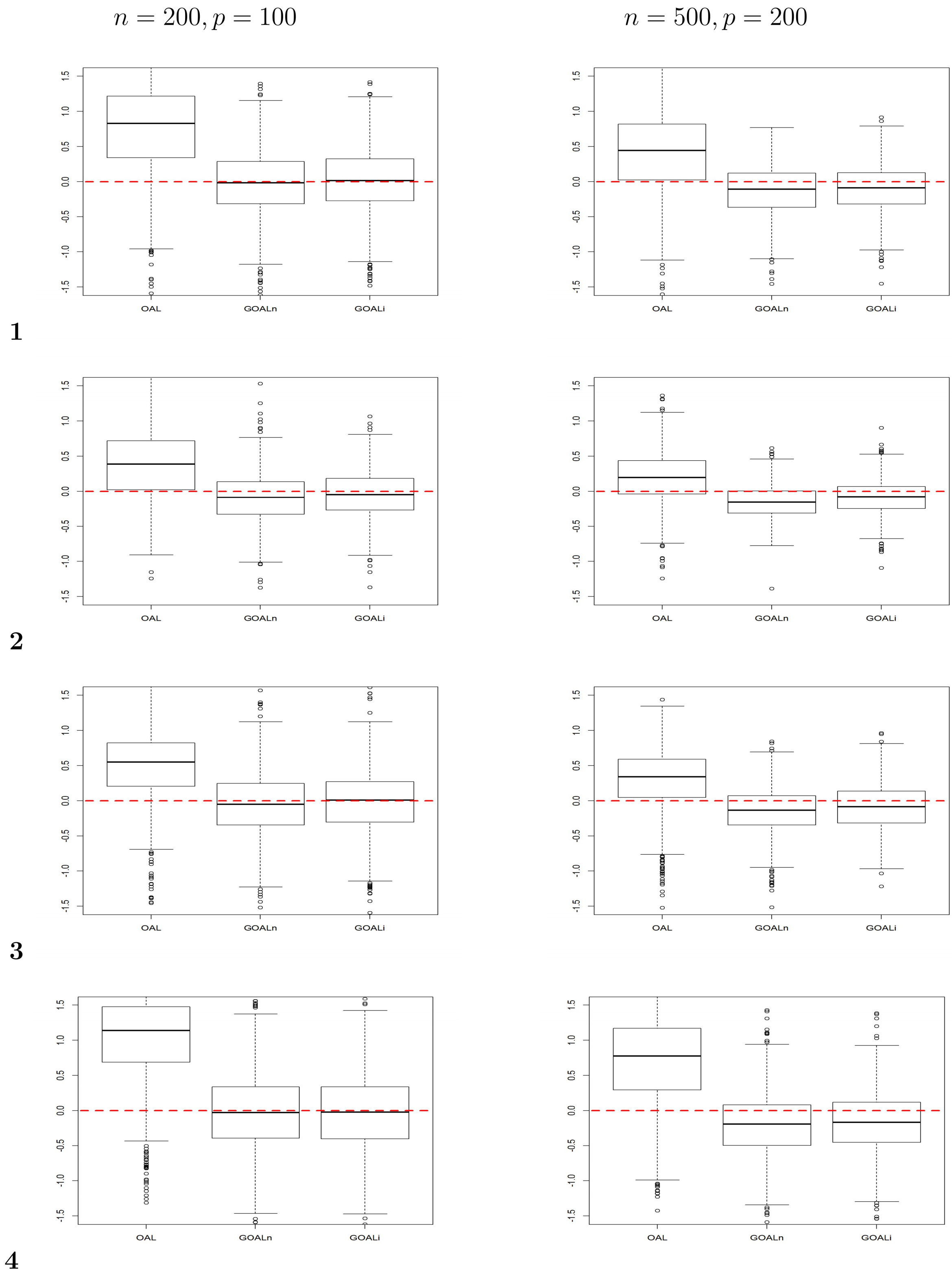

Figure C6: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0$. The true value of ATE is indicated with dotted line (ATE=0).

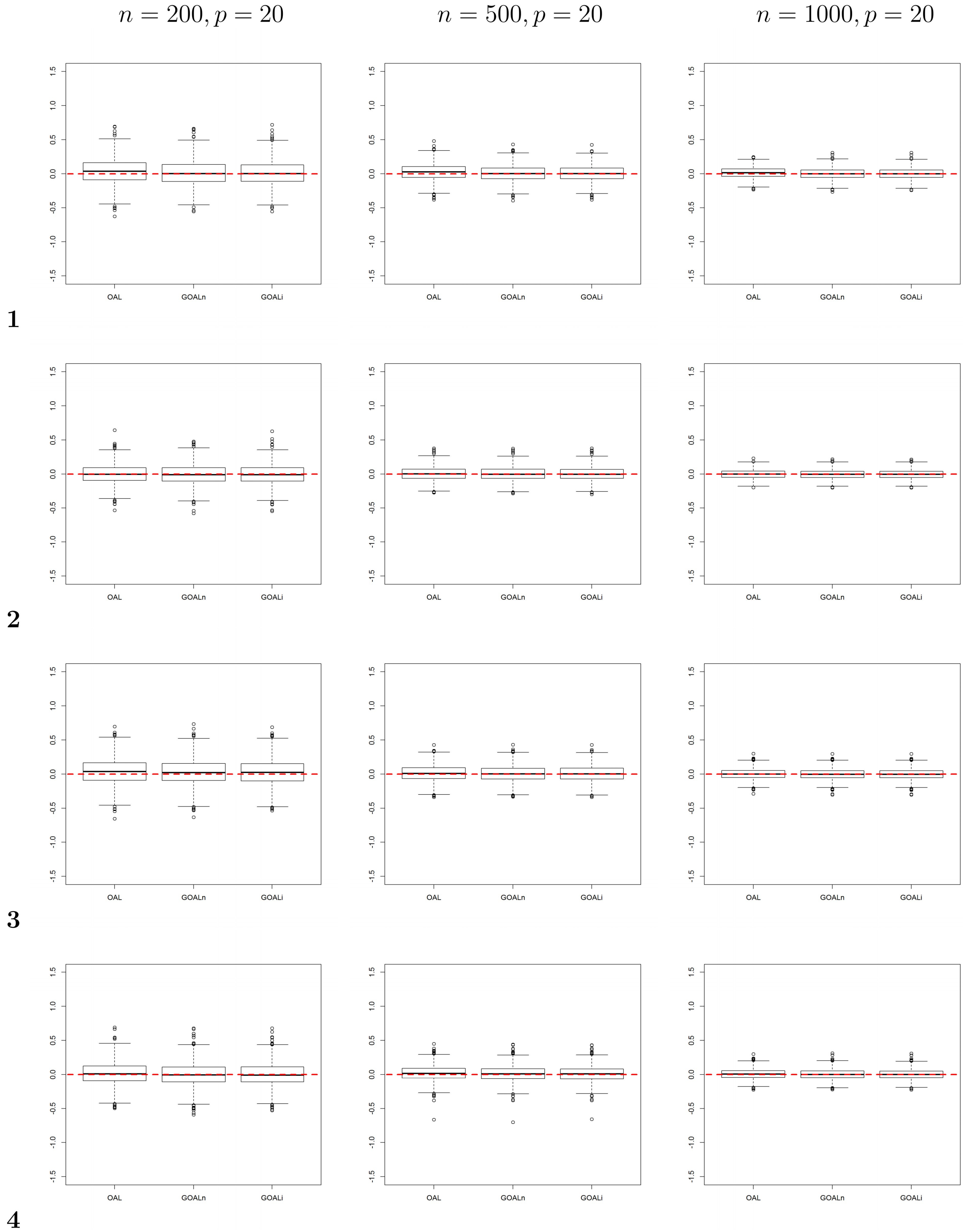

Figure C7: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0.2$. The true value of ATE is indicated with dotted line (ATE=0).

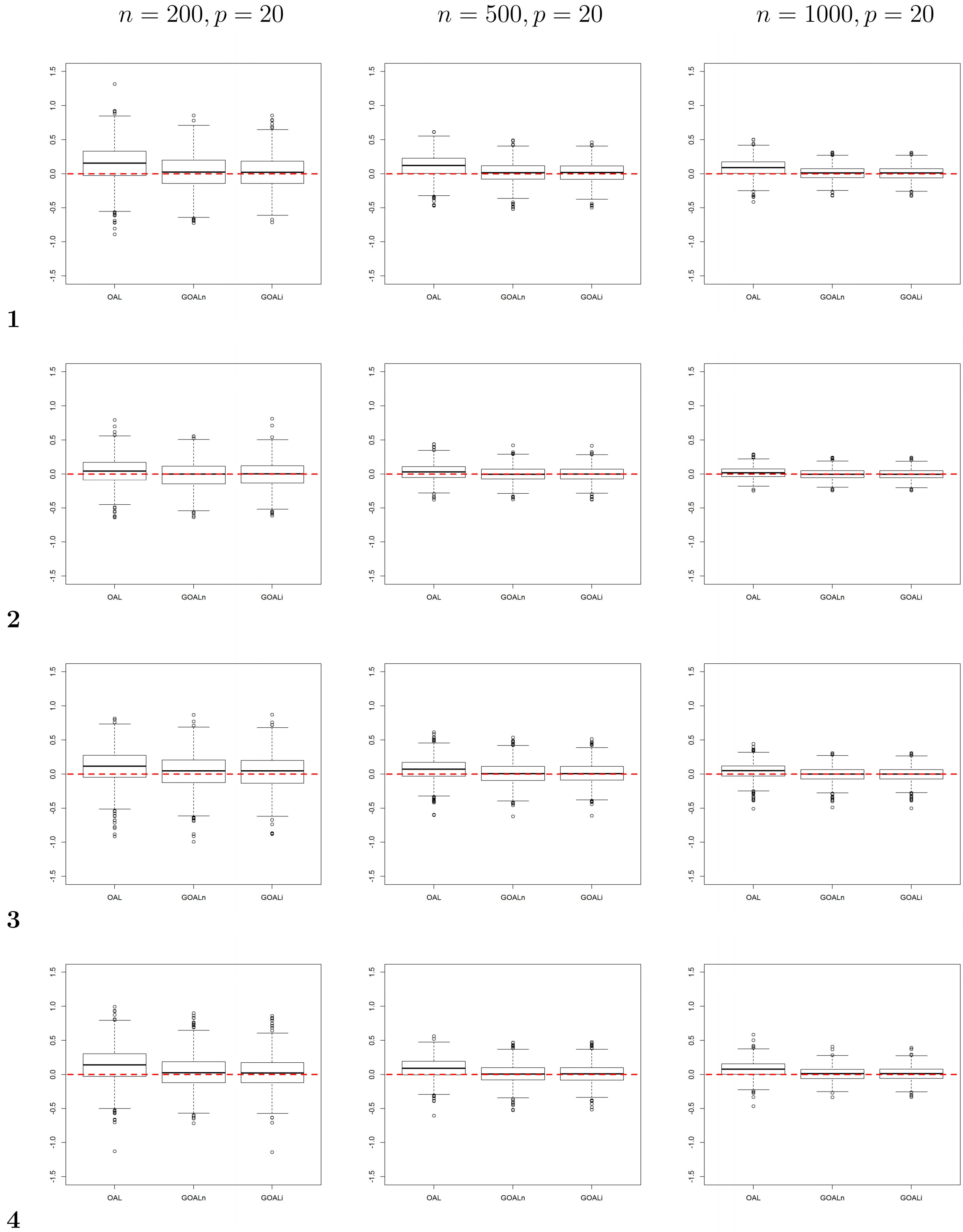

Figure C8: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0.5$. The true value of ATE is indicated with dotted line (ATE=0).

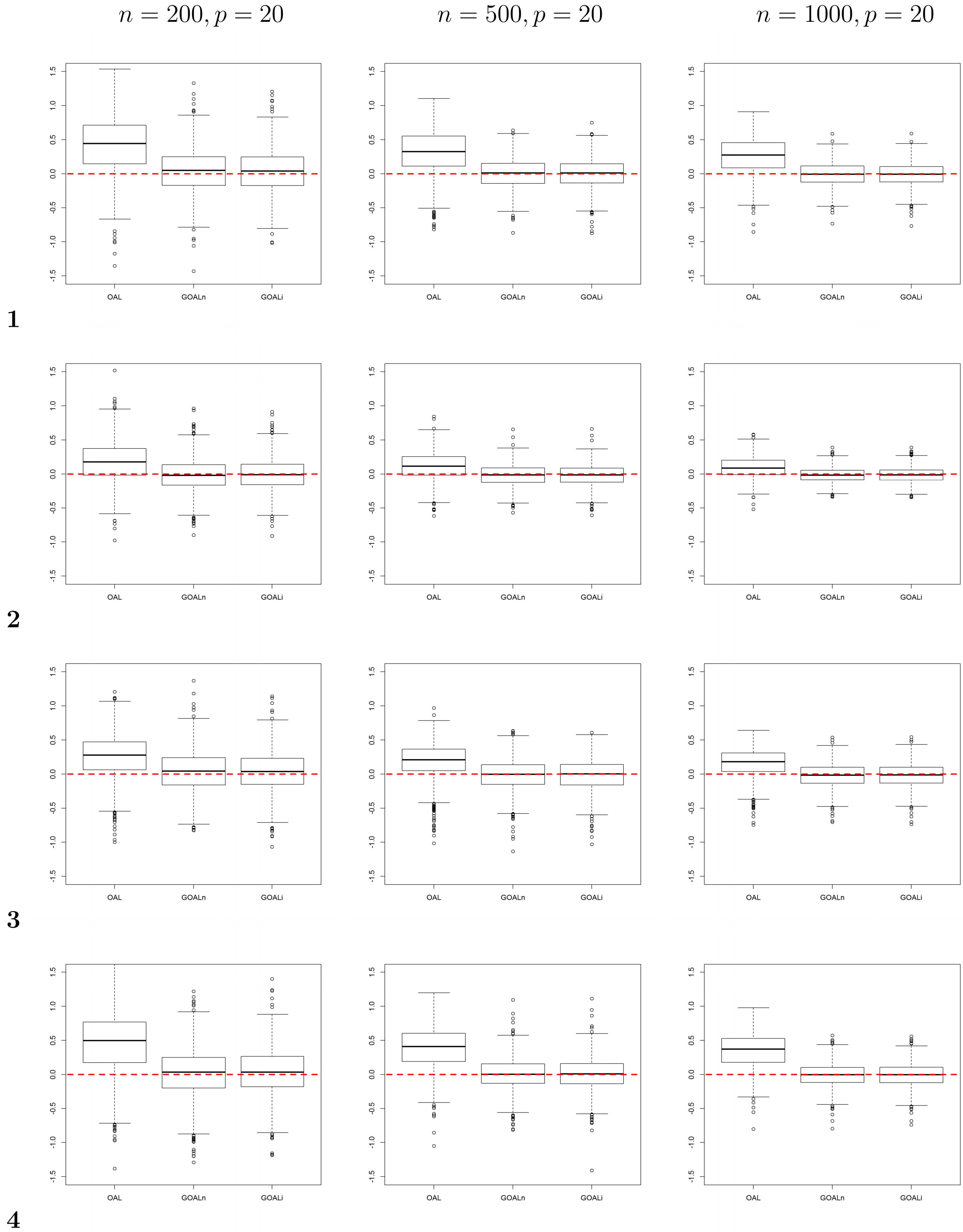

Figure C9: Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0.75$. The true value of ATE is indicated with dotted line (ATE=0).

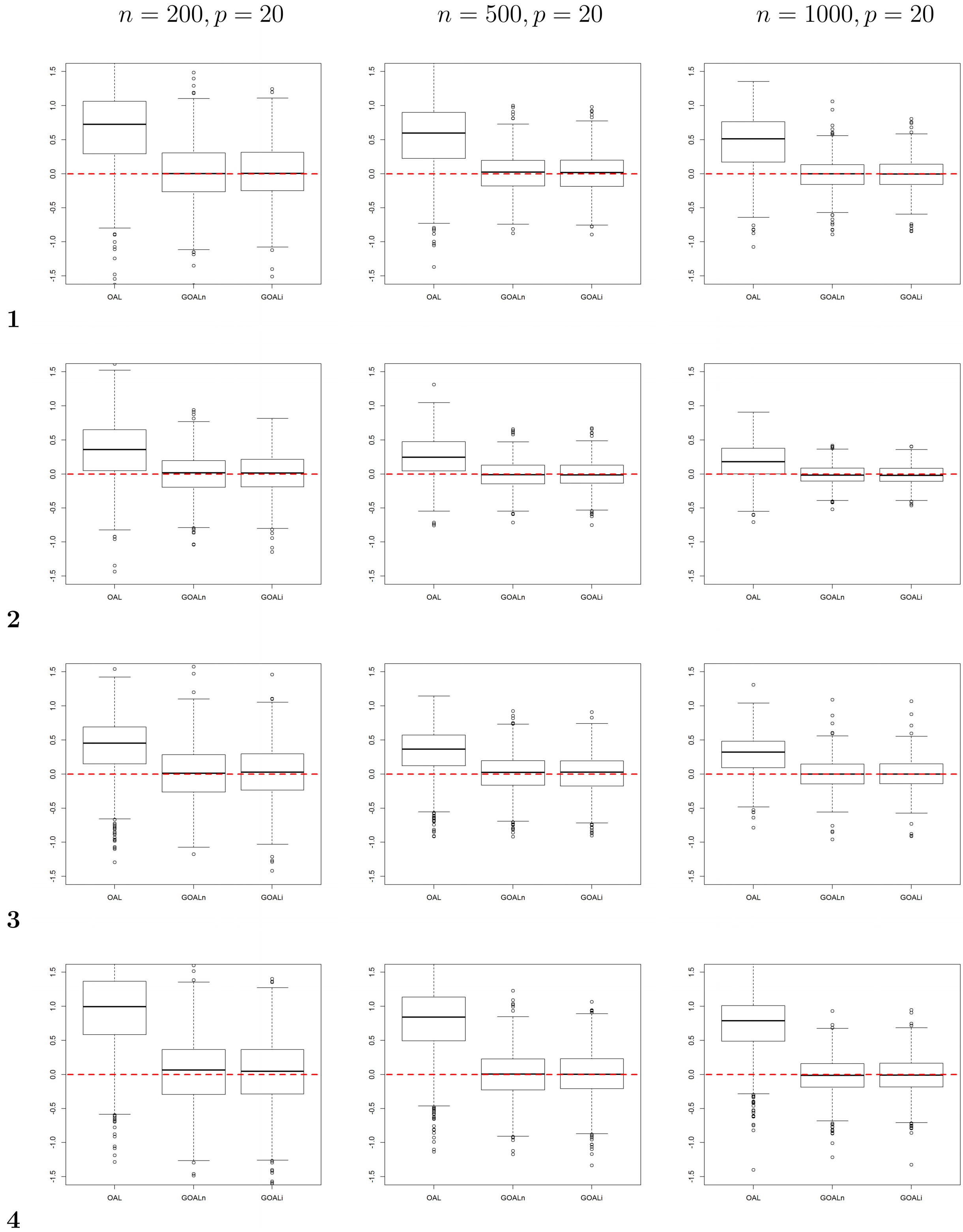

Figure C10: Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 100$ and $\rho = 0$ under Scenarios 1, 2, 3 and 4 (by row).

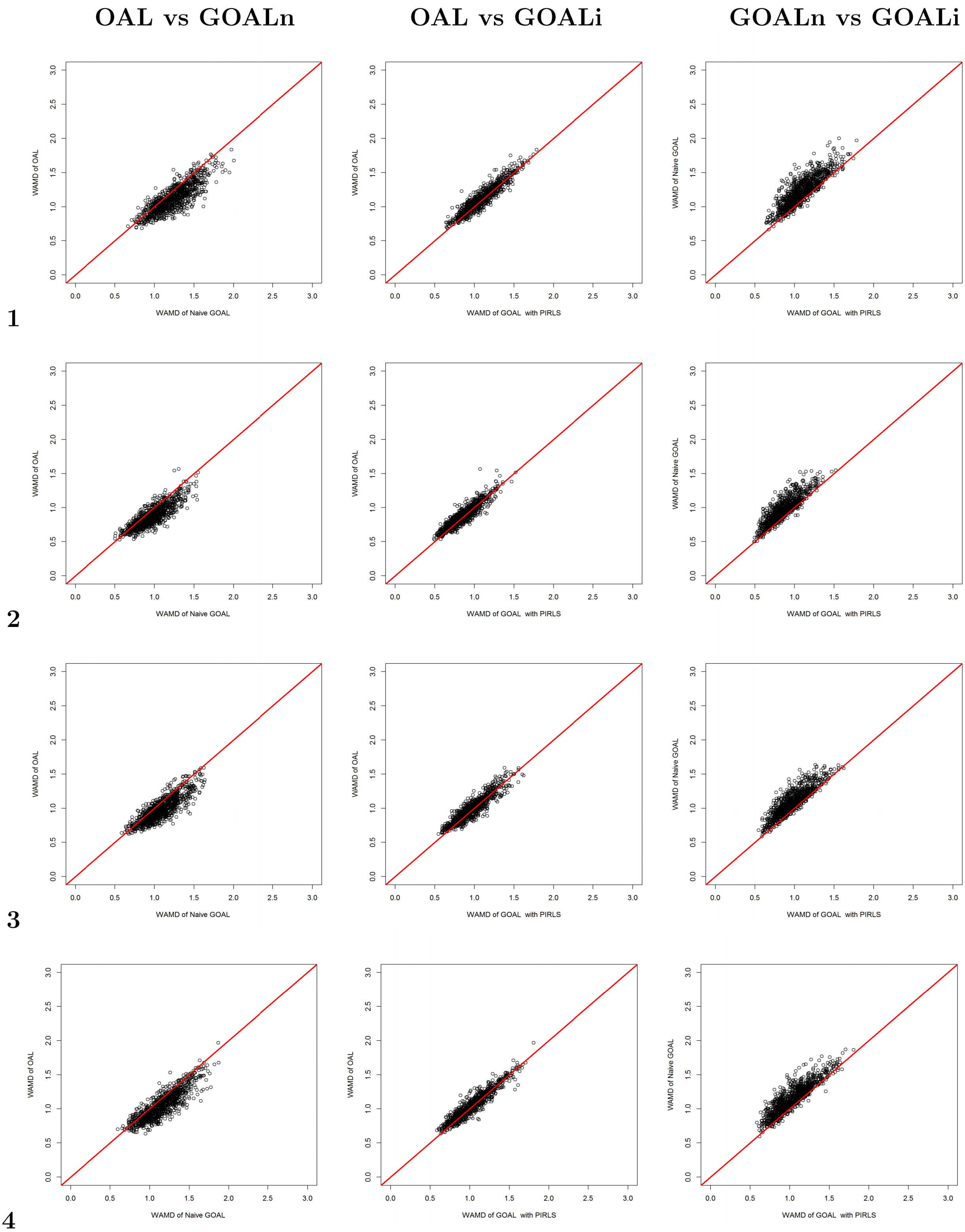

Figure C11: Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 100$ and $\rho = 0.75$ under Scenarios 1, 2, 3 and 4 (by row).

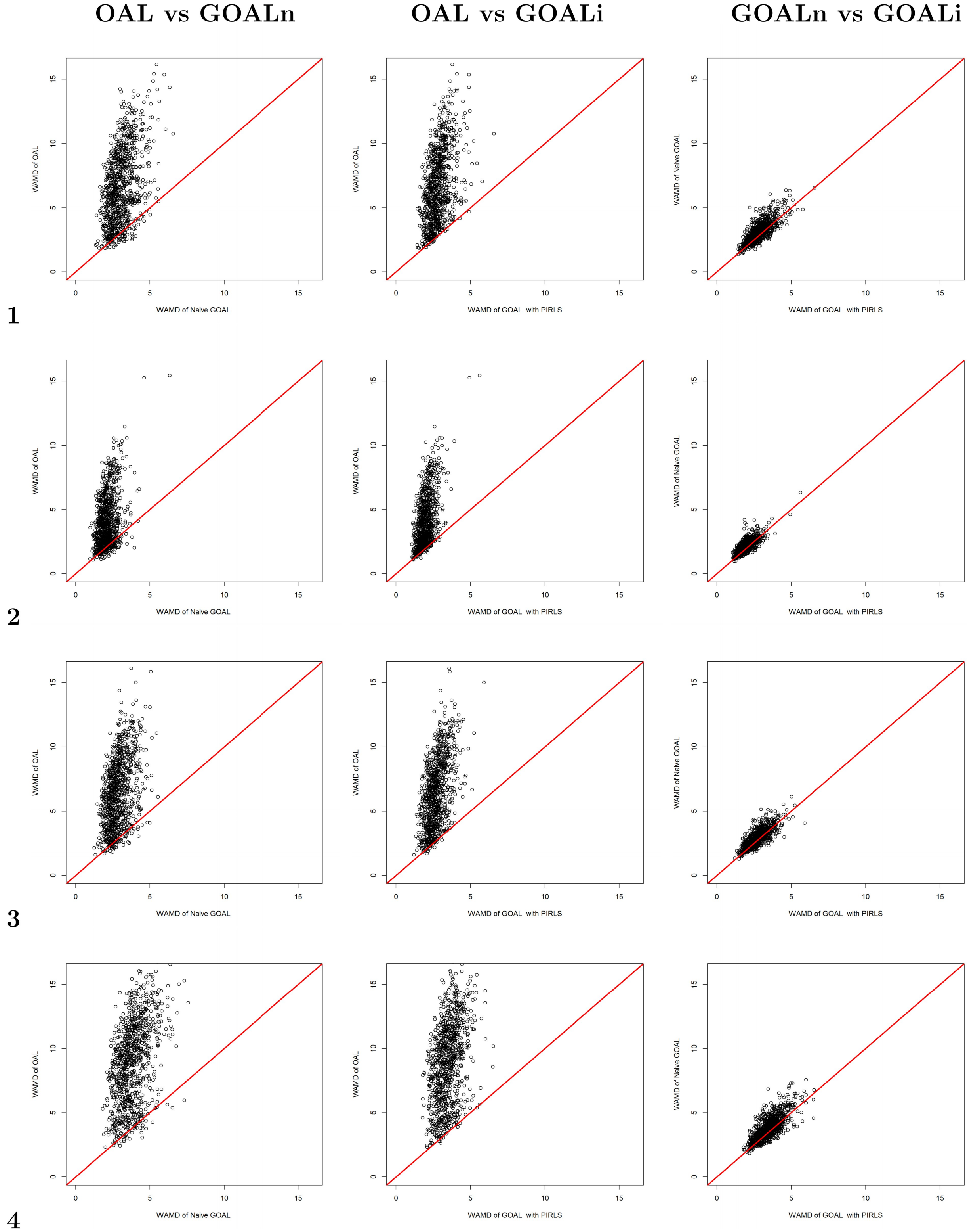

Figure C12: Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 20$ and $\rho = 0$ under Scenarios 1, 2, 3 and 4 (by row).

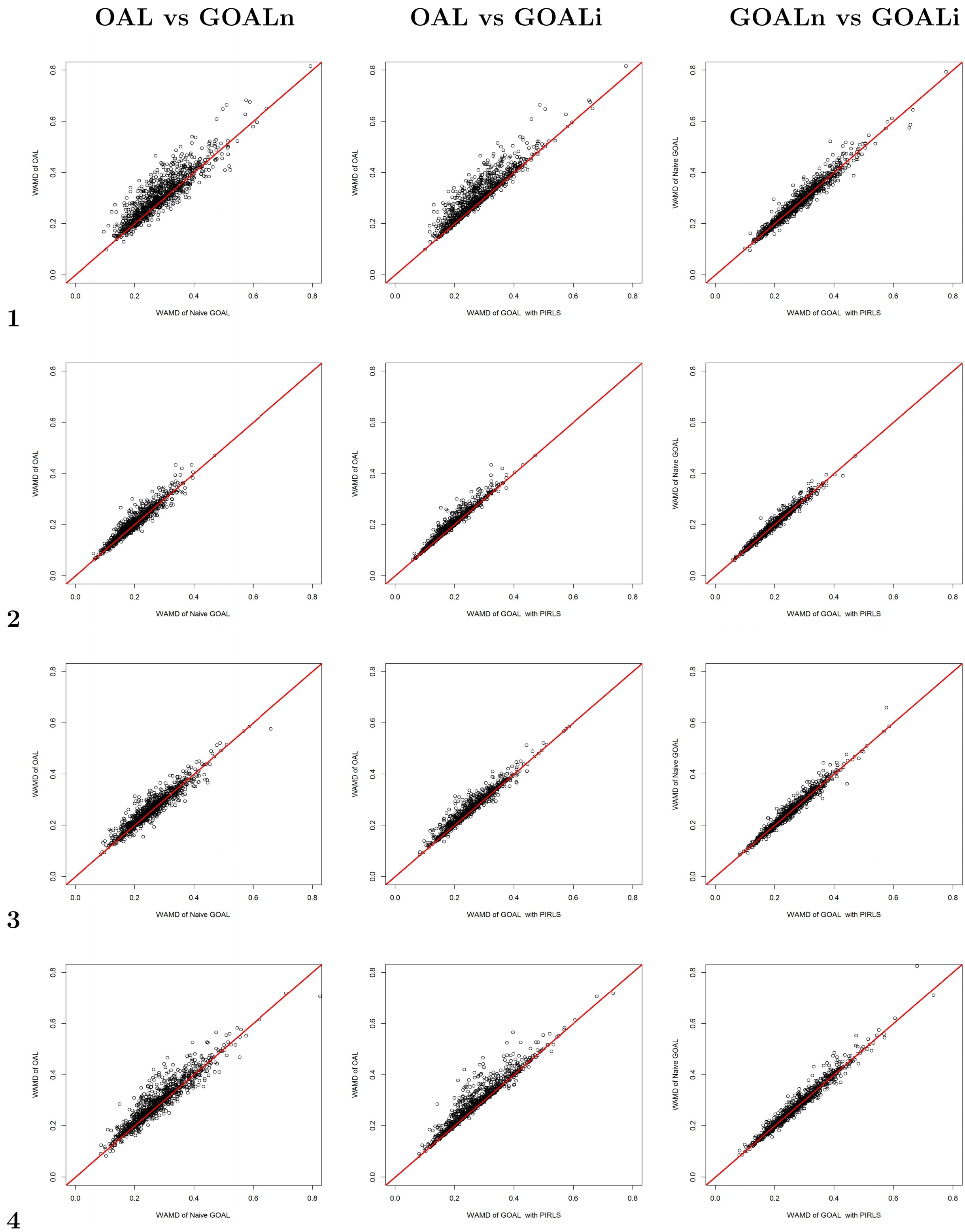

Figure C13: Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 20$ and $\rho = 0.75$ under Scenarios 1, 2, 3 and 4 (by row).

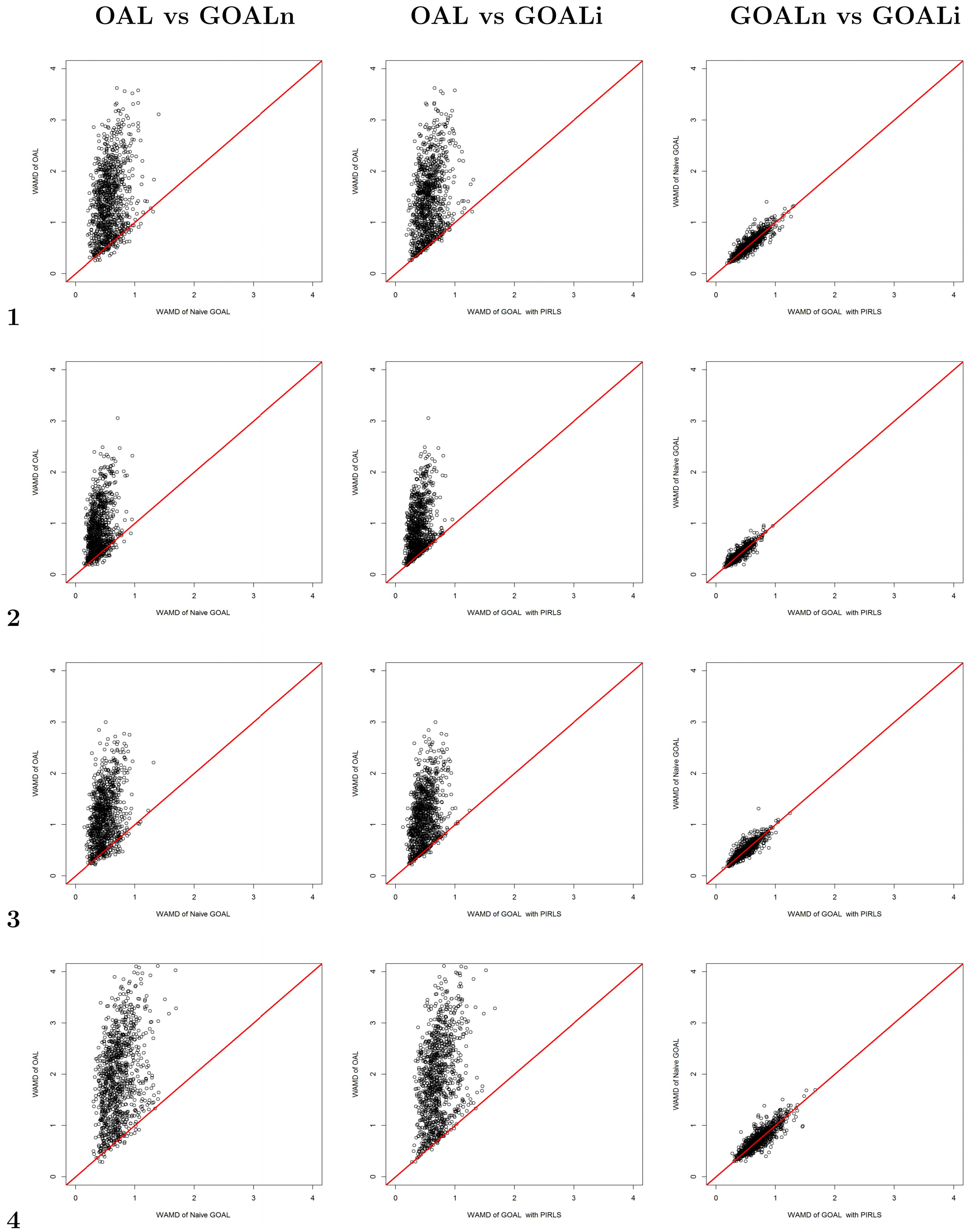

Figure C14: Probability of covariate being included in the propensity score (PS) model for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) with $n = 200$ and $p = 100$.

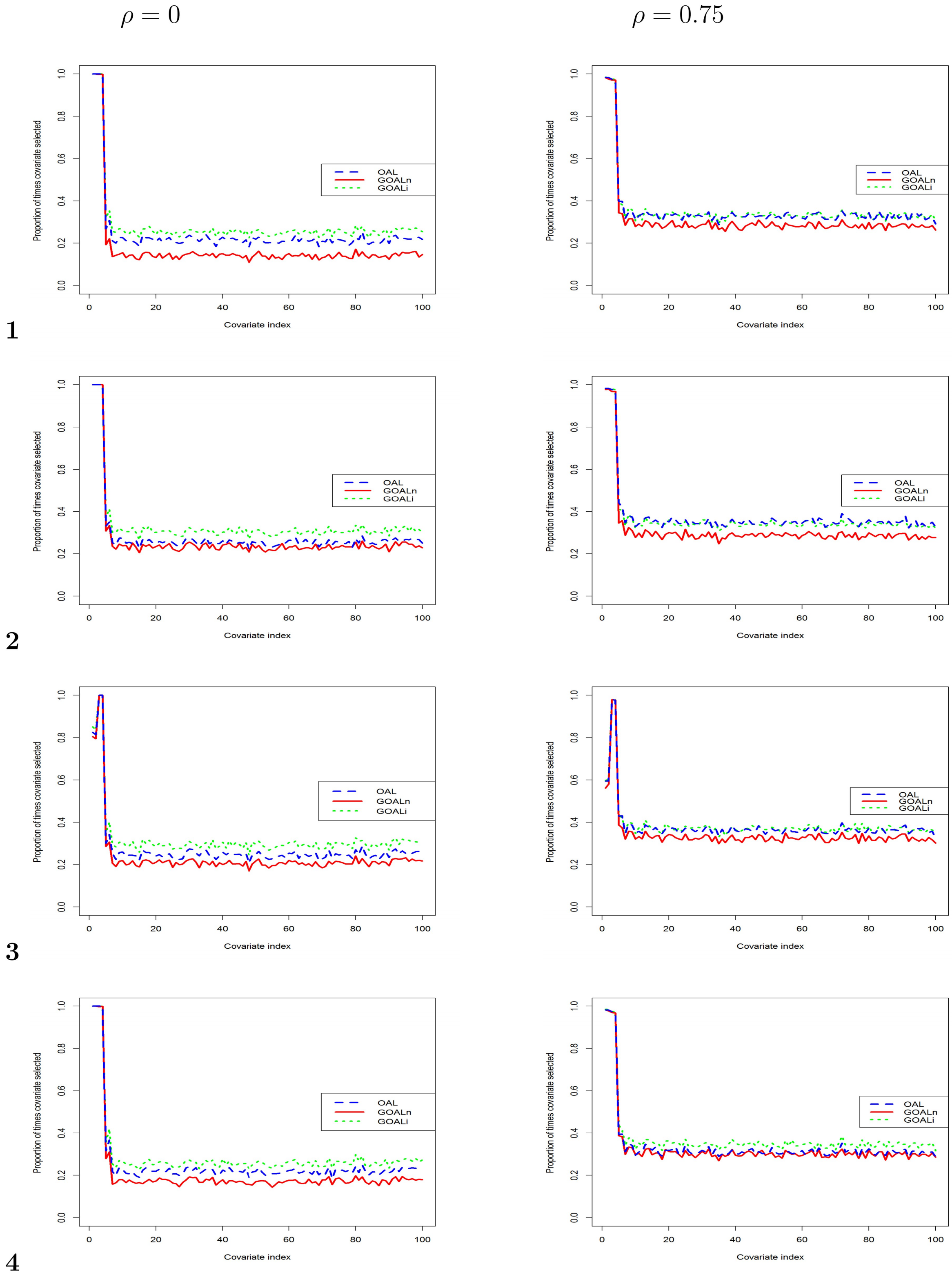

Figure C15: Probability of covariate being included in the propensity score (PS) model for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) with $n = 200$ and $p = 20$.