[Web Appendix]

AIPW: An R Package for Augmented Inverse Probability Weighted Estimation of Average Causal Effects

Yongqi Zhong, Edward H. Kennedy, Lisa M. Bodnar, and Ashley I. Naimi

July 12, 2021

Author affiliations: Department of Epidemiology, Graudate School of Public Health, University of Pittsburgh (Yongqi Zhong, and Lisa M. Bodnar); Department of Data Science and Statistics, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University (Edward H. Kennedy); Department of Epidemiology, Rollins School of Public Health, Emory University (Ashley I. Naimi).

Contents

1 Web Appendix 1: AIPW with missing Outcome

Let R_i be an indicator of whether the outcome for individual i is observed ($R_i = 0$ if missing), and W be all of the covariates from W_Q and W_g. In the presence of missing outcome data, the AIPW estimator in the main text (formula 3) can be written as:

$$\hat{\psi}(a)_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{I(A_i = a, R_i = 1)}{\hat{P}(A = a, R = 1|W_i)}[Y_i - \hat{P}(Y = 1|A_i, W_i, R_i = 1)] + \hat{P}(Y = 1|A \coloneqq a, W_i, R_i = 1)\right\}$$

The propensity scores $\hat{P}(A = a, R = 1|W_i)$ is obtained by estimating the joint probability of treatment and (non)missingness:

$$\hat{P}(A = a, R = 1|W_i) = \hat{P}(R = 1|W_i, A = a)\hat{P}(A = a|W_i),$$

which incorporates missing data mechanism with W. In other words, analyses assume missing at random (MAR) conditional on W, and thus such analyses require W include covariates that render MAR as close to true as possible.

When missing outcomes are detected, the arguments in the AIPW package enabling different covariate sets for the outcome (W_Q) and exposure (W_g) models are disabled. This is because the propensity scores with (non)missing data can be factorized into two ways:

$$\begin{aligned}\hat{P}(A = a, R = 1|W_i) &= \hat{P}(R = 1|W_{Qi}, A = a)\hat{P}(A = a|W_{gi}) \\ &= \hat{P}(R = 1|W_{Qi})\hat{P}(A = a|W_{gi}, R = 1).\end{aligned}$$

In other words, it requires conditioning on both outcome covariates W_Q for missing data mechanism and W_g for exposure mechanism.

2 Web Appendix 2: Derivation of the standard errors of risk ratio and odds ratio for the AIPW estimator

Suppose we have an iid sample $Z_1, ..., Z_n \sim \mathbb{P}$ with $Z = (A, Y, W_Q, W_g)$ where $Y \in \{0, 1\}$. We assume the usual consistency, positivity, and no unmeasured confounding conditions.

Let

$$\widehat{\pi}(a \mid w_g) = \widehat{\mathbb{P}}(A = a \mid W_g = w_g) \text{ and } \widehat{\mu}(w_Q, a) = \widehat{\mathbb{P}}(Y = 1 \mid W_Q = w_Q, A = a)$$

denote estimators of the chance of receiving exposure level $A = a$ given covariate $W_g = w_g$, and the chance of observing outcome $Y = y$ among those with covariates $W_Q = w_Q$ and exposure $A = a$.

Under typical $n^{-1/4}$-type rate conditions, the following estimator is root-n consistent and asymptotically normal

$$\widehat{\mathbb{P}}(Y^a = 1) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbb{1}(A = a)}{\widehat{\pi}(a \mid W_{gi})}\Big\{Y_i - \widehat{\mu}(W_{Qi}, a)\Big\} + \widehat{\mu}(W_{Qi}, a)\right]$$

for the marginal counterfactual probability $\mathbb{P}(Y^a = 1) = \mathbb{E}\{\mathbb{E}(Y \mid X, A = a)\}$. Note we use counterfactual expressions like $\mathbb{P}(Y^a = 1)$ as shorthand, but all the results here follow for the observational expressions $\mathbb{E}\{\mathbb{E}(Y \mid X, A = a)\}$ regardless of whether these equal the corresponding counterfactual expressions.

Therefore the following are estimators for the marginal risk ratio and odds ratio:

$$\widehat{\psi}_{rr} = \frac{\widehat{\mathbb{P}}(Y^1 = 1)}{\widehat{\mathbb{P}}(Y^0 = 1)}$$
$$\widehat{\psi}_{or} = \frac{\widehat{\mathbb{P}}(Y^1 = 1)/\{1 - \widehat{\mathbb{P}}(Y^1 = 1)\}}{\widehat{\mathbb{P}}(Y^0 = 1)/\{1 - \widehat{\mathbb{P}}(Y^0 = 1)\}}$$

Since both the RR and OR are non-negative, normal approximations will work best if we construct confidence intervals on the log scale and then exponentiate.

Let $\varphi_a(Z; \pi, \mu) = \frac{\mathbb{1}(A=a)}{\pi(a|W_{gi})}\Big\{Y_i - \mu(W_{Qi}, a)\Big\} + \mu(W_{Qi}, a)$ denote the uncentered influence function for $\mathbb{P}(Y^a = 1) = \mathbb{E}\{\mathbb{E}(Y \mid X, A = a)\}$ so that

$$\widehat{\mathbb{P}}(Y^a = 1) = \frac{1}{n}\sum_{i=1}^{n}\varphi_a(Z_i; \widehat{\pi}, \widehat{\mu})$$

Also let

$$\Sigma = \operatorname{cov}\begin{pmatrix}\varphi_0(Z; \pi, \mu) \\ \varphi_1(Z; \pi, \mu)\end{pmatrix} = \begin{pmatrix}\operatorname{var}\{\varphi_0(Z; \pi, \mu)\} & \operatorname{cov}\{\varphi_0(Z; \pi, \mu), \varphi_1(Z; \pi, \mu)\} \\ \operatorname{cov}\{\varphi_0(Z; \pi, \mu), \varphi_1(Z; \pi, \mu)\} & \operatorname{var}\{\varphi_1(Z; \pi, \mu)\}\end{pmatrix}$$

denote the covariance matrix of the influence functions, with elements $\Sigma = \begin{pmatrix}\Sigma_{00} & \Sigma_{01} \\ \Sigma_{01} & \Sigma_{11}\end{pmatrix}$.

An estimate of the covariance matrix is simply given by

$$\widehat{\Sigma} = \begin{pmatrix}\widehat{\operatorname{var}}\{\varphi_0(Z; \widehat{\pi}, \widehat{\mu})\} & \widehat{\operatorname{cov}}\{\varphi_0(Z; \widehat{\pi}, \widehat{\mu}), \varphi_1(Z; \widehat{\pi}, \widehat{\mu})\} \\ \widehat{\operatorname{cov}}\{\varphi_0(Z; \widehat{\pi}, \widehat{\mu}), \varphi_1(Z; \widehat{\pi}, \widehat{\mu})\} & \widehat{\operatorname{var}}\{\varphi_1(Z; \widehat{\pi}, \widehat{\mu})\}\end{pmatrix}$$

where $\widehat{\operatorname{cov}}$ and $\widehat{\operatorname{var}}$ are just empirical covariances/variances.

Then under usual $n^{-1/4}$-type rate conditions on $(\widehat{\pi}, \widehat{\mu})$ we have

$$\sqrt{n}\left\{\begin{pmatrix}\widehat{\mathbb{P}}(Y^0=1)\\ \widehat{\mathbb{P}}(Y^1=1)\end{pmatrix} - \begin{pmatrix}\mathbb{P}(Y^0=1)\\ \mathbb{P}(Y^1=1)\end{pmatrix}\right\} \rightsquigarrow N(0,\Sigma)$$

Therefore by the delta method, we have

$$\sqrt{n}\Big(\log\widehat{\psi}_{rr} - \log\psi_{rr}\Big) \rightsquigarrow N\left(0, \begin{pmatrix}\frac{-1}{\mathbb{P}(Y^0=1)}\\ \frac{1}{\mathbb{P}(Y^1=1)}\end{pmatrix}^{\mathrm{T}} \Sigma \begin{pmatrix}\frac{-1}{\mathbb{P}(Y^0=1)}\\ \frac{1}{\mathbb{P}(Y^1=1)}\end{pmatrix}\right)$$

so that a 95% CI for ψ_{rr} is given by

$$\exp\left\{\log\widehat{\psi}_{rr} \pm \frac{1.96}{\sqrt{n}}\sqrt{\sum_{a=0}^{1}\frac{\widehat{\Sigma}_{aa}}{\widehat{\mathbb{P}}(Y^a=1)^2} - \frac{2\widehat{\Sigma}_{01}}{\widehat{\mathbb{P}}(Y^0=1)\widehat{\mathbb{P}}(Y^1=1)}}\right\}$$

Similarly the delta method also gives

$$\sqrt{n}\Big(\log\widehat{\psi}_{or} - \log\psi_{or}\Big) \rightsquigarrow N\left(0, \begin{pmatrix}\frac{-1}{\mathbb{P}(Y^0=1)\mathbb{P}(Y^0=0)}\\ \frac{1}{\mathbb{P}(Y^1=1)\mathbb{P}(Y^1=0)}\end{pmatrix}^{\mathrm{T}} \Sigma \begin{pmatrix}\frac{-1}{\mathbb{P}(Y^0=1)\mathbb{P}(Y^0=0)}\\ \frac{1}{\mathbb{P}(Y^1=1)\mathbb{P}(Y^1=0)}\end{pmatrix}\right)$$

so that a 95% CI for ψ_{or} is given by

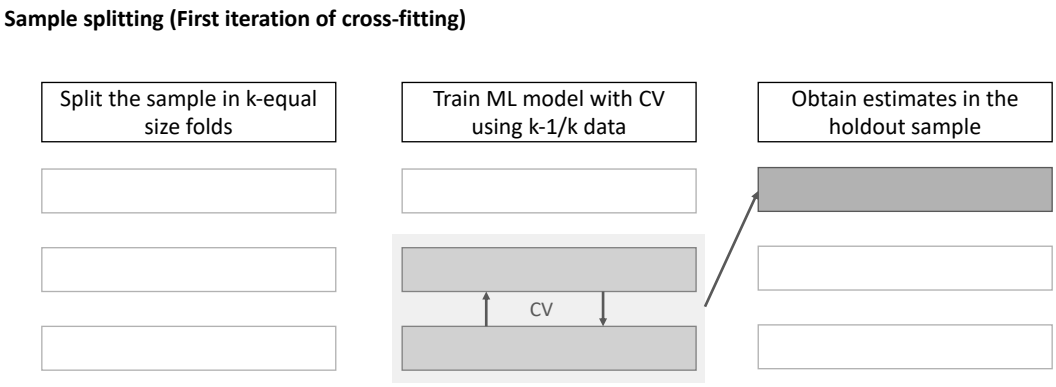
$$\exp\left\{\log\widehat{\psi}_{or} \pm \frac{1.96}{\sqrt{n}}\sqrt{\sum_{a=0}^{1}\frac{\widehat{\Sigma}_{aa}}{\widehat{\mathbb{P}}(Y^a=1)^2\widehat{\mathbb{P}}(Y^a=0)^2} - \frac{2\widehat{\Sigma}_{01}}{\widehat{\mathbb{P}}(Y^0=1)\widehat{\mathbb{P}}(Y^0=0)\widehat{\mathbb{P}}(Y^1=1)\widehat{\mathbb{P}}(Y^1=0)}}\right\}$$

3 Web Appendix 3: Example code that can be used to implement an augmented inverse probability weighted estimator via the AIPW package using the simulated RCT data available in the package.

```
library(AIPW)
library(SuperLearner)
set.seed(1234)
#load simulated dataset (RCT)
data(eager_sim_rct)
#Specify SuperLearner libraries
sl.lib = c("SL.gam","SL.earth","SL.ranger","SL.xgboost")
#Create a vector of covariates
Cov = c("loss_num","age", "time_try_pregnant","BMI","meanAP")
#create a new AIPW object called AIPW_SL
AIPW_SL <- AIPW$new(Y = eager_sim_rct$sim_Y,
                    A = eager_sim_rct$sim_Tx,
                    W.g = eager_sim_rct$eligibility,
                    W.Q = subset(eager_sim_rct,select=Cov), #covariates
                    Q.SL.library = sl.lib, #outcome model
                    g.SL.library = sl.lib, #exposure model
                    k_split = 10, #num of folds for cross-fitting
                    verbose=TRUE)
#fit the data stored in the AIPW_SL object
AIPW_SL$fit()
#summarise the results using truncated propensity scores
AIPW_SL$summary(g.bound = 0.025)
```

4 Web Figure 1: Implementation of sample splitting and cross-fitting

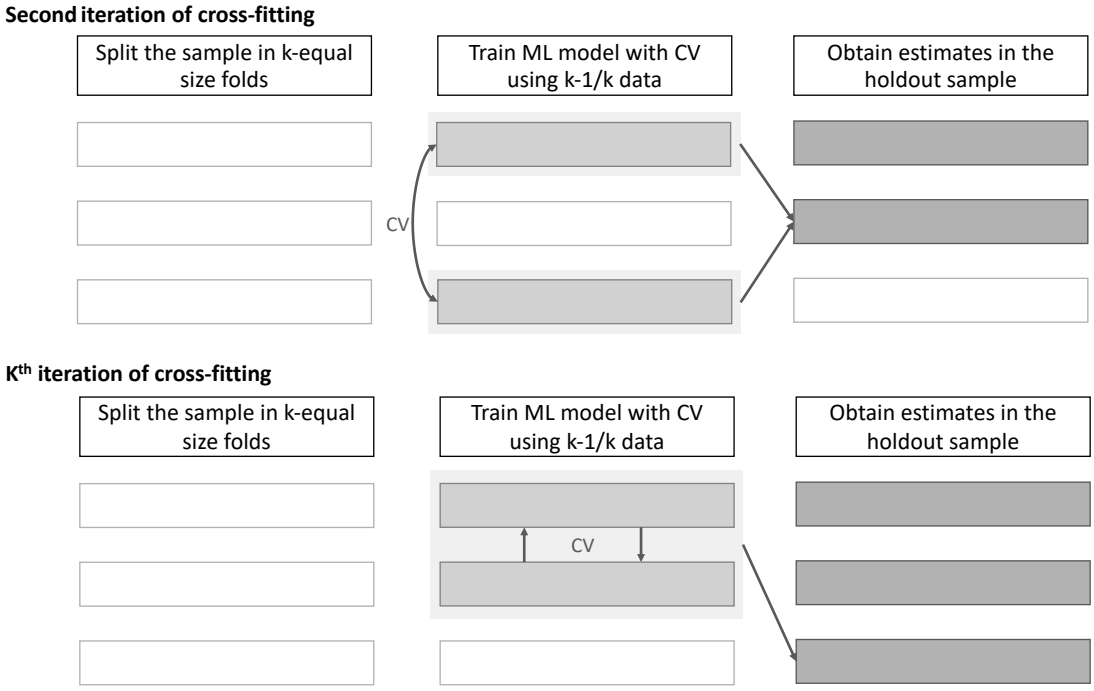
To implement sample splitting, one needs to subset the input data into k equal-size folds randomly, then fit the exposure and the outcome models with $(k-1)/k$ data, and finally use the fitted models to estimate propensity scores and outcome model predictions with the $1/k$ held-out sample.[1?] (Figure S1a)



ML, Stacking machine learning; CV, Cross-validation; k=3 in this example.

Web Figure 1a: Illustration of sample splitting

Cross-fitting is a more efficient version of sample splitting.[1] While sample splitting only uses $1/k$ of the sample for estimating propensity score and outcome model, cross-fitting iterates the process sample-splitting k times until estimates of the exposure and outcome for all observations are obtained.(Figure S1b).



ML, Stacking machine learning; CV, Cross-validation; k=3 in this example.

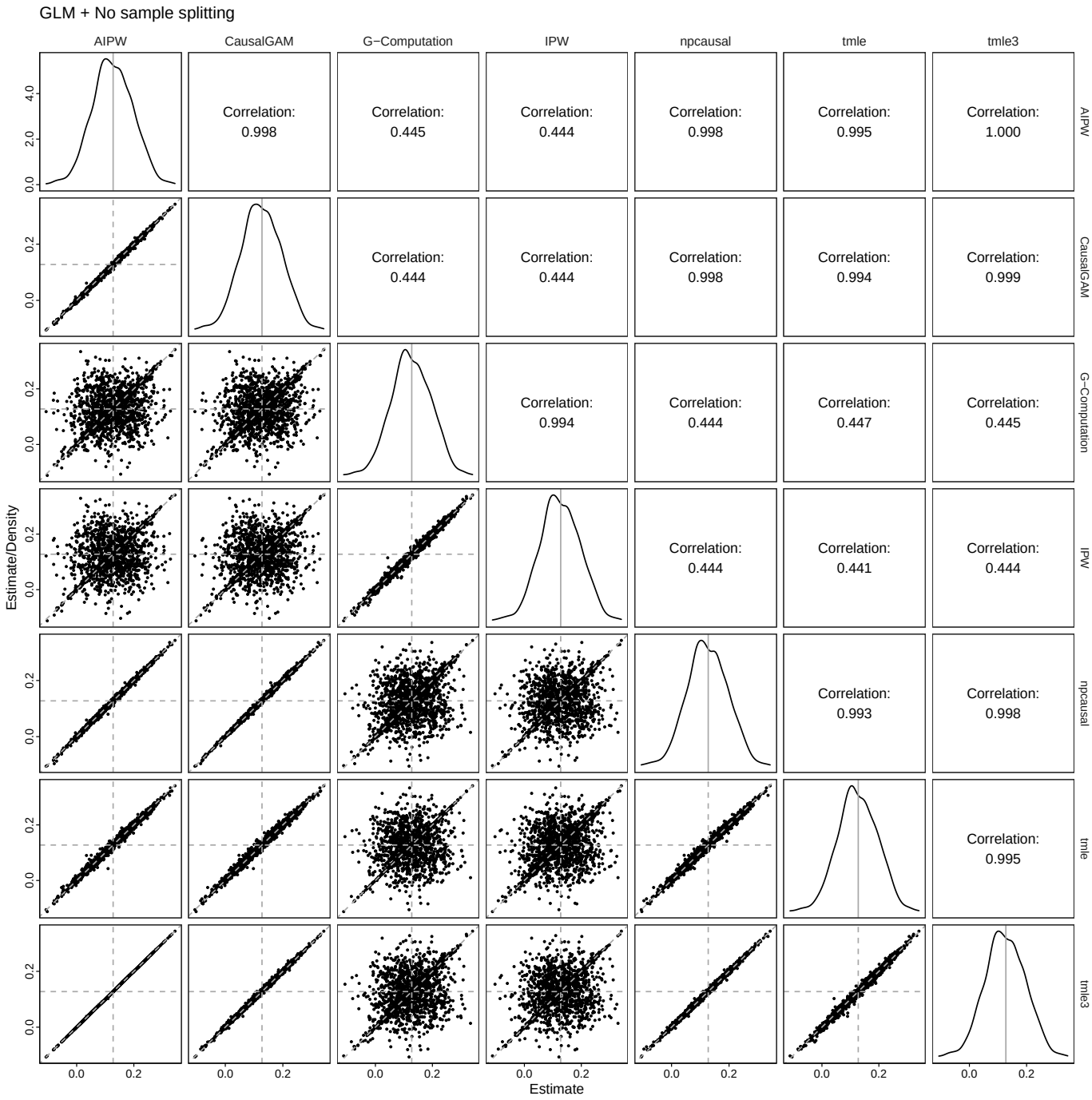
Web Figure 1b: Illustration of cross-fitting

In the **AIPW** package with the **SuperLearner**, when $k\_split = 2$, 10-fold cross-validation (CV) will be used for training stacking machine learning algorithms; when $k\_split \geq 3$, $k\_split - 1$ fold CV will be used (e.g., 2-fold CV is used for $k\_split = 3$), with the CV-fold assignment remains the same throughout cross-fitting. With the **sl3** package, 10-fold CV will be used regardless of k_split.

References

1. Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 2018;21(1):C1–C68. doi:10.1111/ectj.12097. Publisher: Oxford Academic.

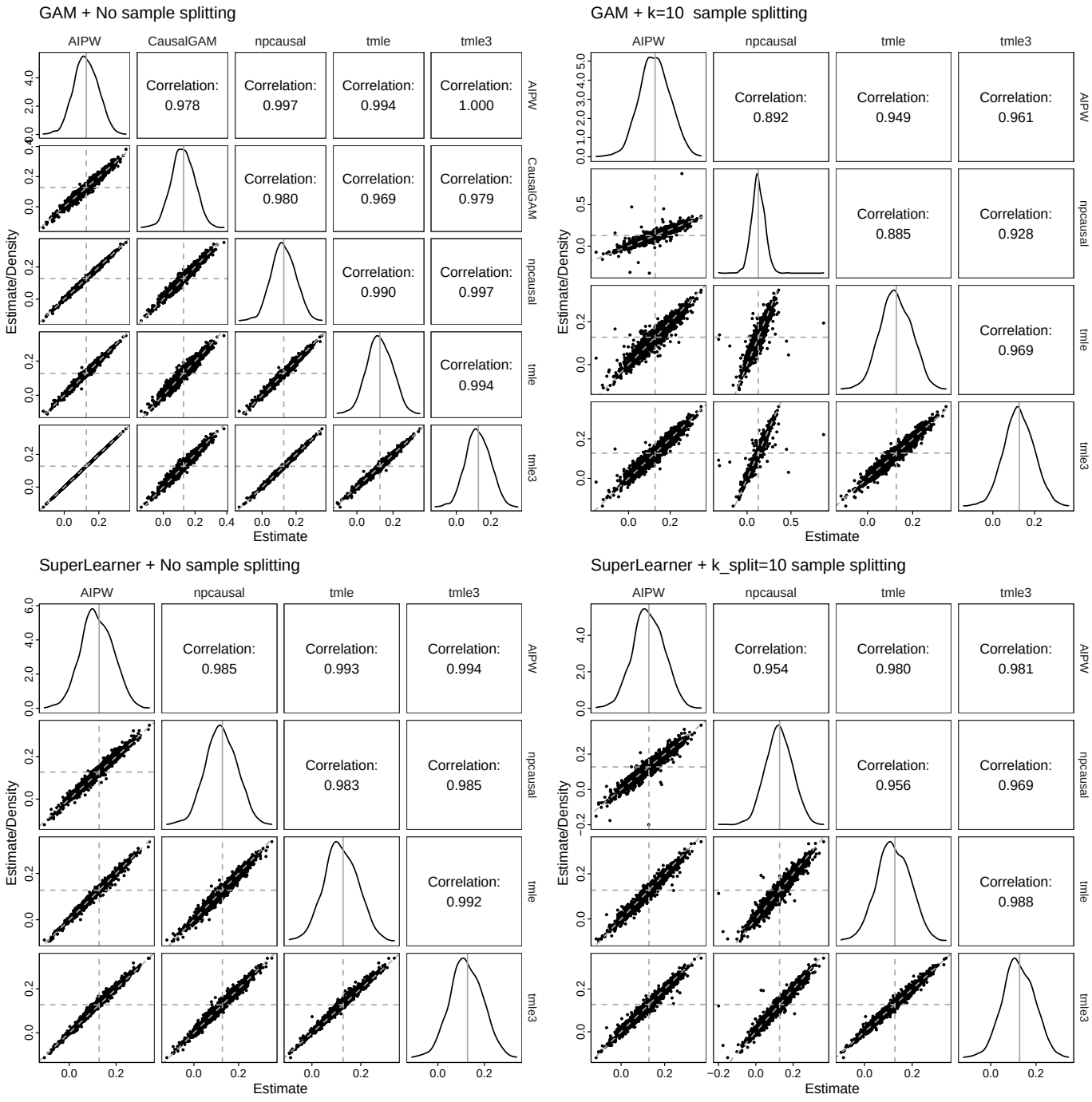
5 Web Figure 2: Pairwise comparison of risk difference estimates with the true data generating functions using different methods



Abbreviations: IPW, Inverse Probability Weighting; AIPW, Augmented Inverse Probability Weighting; GLM, Generalized Linear Model; GAM, Generalized Additive Model; TMLE, Targeted Maximum Likelihood Estimation.

Diagonal panels are the density plots of estimates from each package, lower diagonal panels are scatter plots of estimates between two packages, and upper diagonal panels are Pearson correlations of estimates between two packages. In the scatter plots, horizontal and vertical lines refer to $RD_{true} = 0.128$, and diagonal lines are references with a slope = 1 and an intercept of 0. Please refer to the main text for the details.

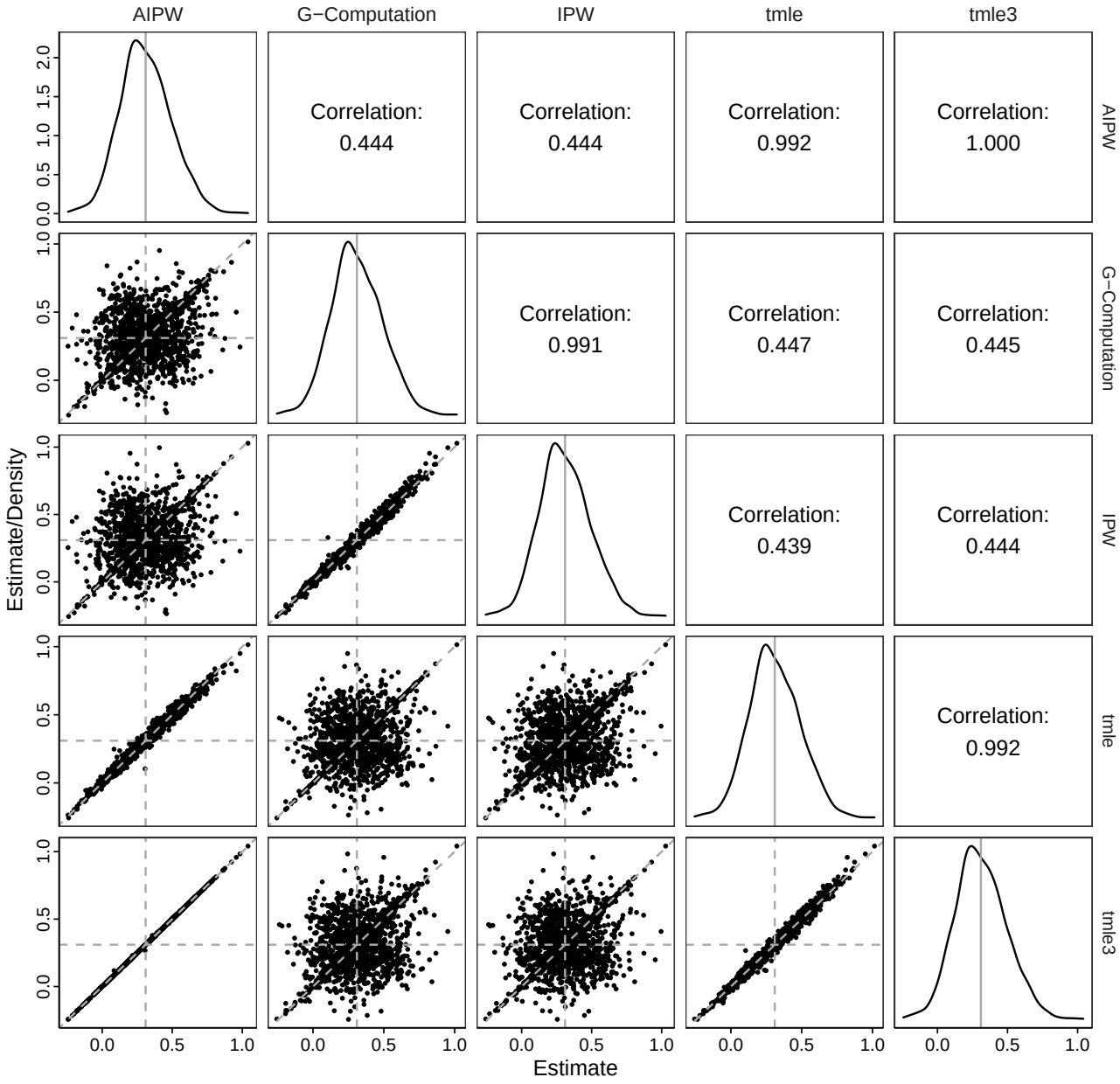
6 Web Figure 3: Pairwise comparison of risk difference estimates using doubly robust packages with different estimation methods



Abbreviations: IPW, Inverse Probability Weighting; AIPW, Augmented Inverse Probability Weighting; GLM, Generalized Linear Model; GAM, Generalized Additive Model; TMLE, Targeted Maximum Likelihood Estimation.

Diagonal panels are the density plots of estimates from each package, lower diagonal panels are scatter plots of estimates between two packages, and upper diagonal panels are Pearson correlations of estimates between two packages. In the scatter plots, horizontal and vertical lines refer to $RD_{true} = 0.128$, and diagonal lines are references with a slope = 1 and an intercept of 0. Please refer to the main text for the details.

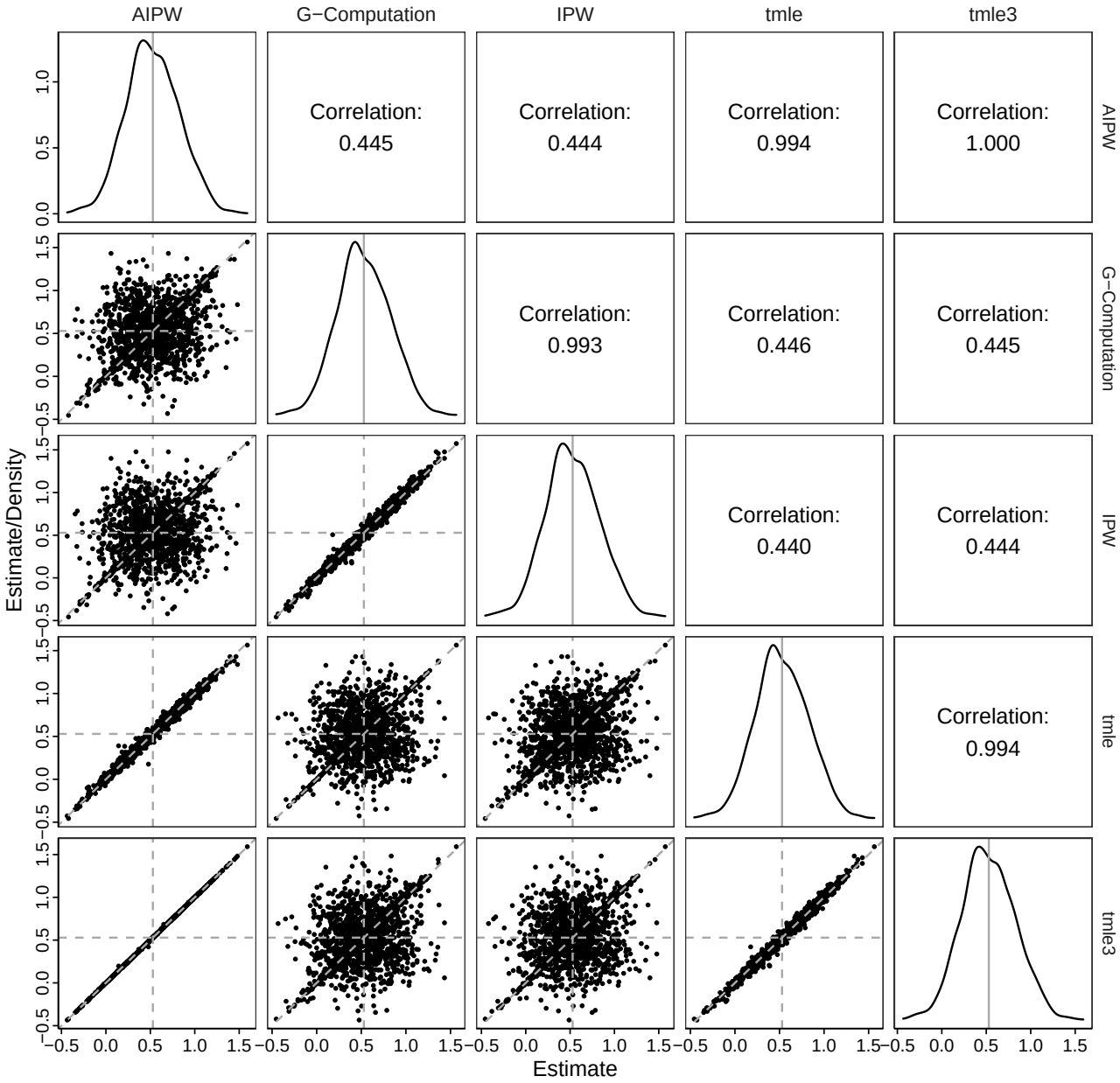
7 Web Figure 4: Pairwise comparison of log(risk ratio) estimates with the true data generating functions using different methods



Abbreviations: IPW, Inverse Probability Weighting; AIPW, Augmented Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; RR, Risk Ratio.

Diagonal panels are the density plots of estimates from each package, lower diagonal panels are scatter plots of estimates between two packages, and upper diagonal panels are Pearson correlations of estimates between two packages. In the scatter plots, horizontal and vertical lines refer to $log(RR_{true}) = 0.31$, and diagonal lines are references with a slope = 1 and an intercept of 0

8 Web Figure 5: Pairwise comparison of log(odds ratio) estimates with the true data generating functions using different methods



Abbreviations: IPW, Inverse Probability Weighting; AIPW, Augmented Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; OR, Odds Ratio.

Diagonal panels are the density plots of estimates from each package, lower diagonal panels are scatter plots of estimates between two packages, and upper diagonal panels are Pearson correlations of estimates between two packages. In the scatter plots, horizontal and vertical lines refer to $log(OR_{true}) = 0.53$, and diagonal lines are references with a slope = 1 and an intercept of 0

9 Web Table 1: Performance of the AIPW package in estimating the average treatment effect [log(RR) and log(OR)] using true generalized linear model without cross-fitting in a simulated observational study based on EAGeR[a]

Package/Method	Bias (SE)	MSE	MeanCIwidth	Coverage (SE)[b,c]
log(RR)				
G-Computation	0.001 (0.004)	0.032	0.707	96.2 (0.4)
IPW	0.001 (0.004)	0.033	0.719	96.4 (0.4)
AIPW	0.001 (0.004)	0.033	0.675	94.8 (0.5)
tmle	0.001 (0.004)	0.032	0.671	94.8 (0.5)
tmle3	0.001 (0.004)	0.033	0.687	95.0 (0.5)
log(RR)				
G-Computation	-0.002 (0.007)	0.087	1.171	95.7 (0.5)
IPW	-0.002 (0.007)	0.090	1.195	96.1 (0.4)
AIPW	-0.002 (0.007)	0.090	1.119	94.8 (0.5)
tmle	-0.002 (0.007)	0.087	1.114	94.8 (0.5)
tmle3	-0.002 (0.007)	0.090	1.141	95.1 (0.5)

Abbreviations: EAGeR, the Effects of Aspirin in Gestation and Reproduction trial; SE, Standard Error; MSE, Mean Squared Error; CI, Confidence Intervals; MeanCIwidth, Mean width of 95% Confidence Interval; MeanRuntimeSec, Mean Runtime in Seconds; IPW, Inverse Probability Weighting; AIPW, Augmented Inverse Probability Weighting; RR, Risk Ratio; OR, Odds Ratio.

[a] Sample size (n) = 200; Number of simulation (nSim) = 2000; $log(RR_{true}) = 0.31$; $log(OR_{true}) = 0.53$; Numbers within parentheses are Monte Carlo SEs of the performance indicator estimates.

[b] Values of confidence interval coverage and its standard errors are expressed as percentages.

[c] Asymptotic SEs were used for CI calculation in AIPW, tmle and tmle3. CIs for G-Computation and IPW were obtained by 200 bootstraps and sandwich estimators, respectively.