

Web-based Supplementary Materials for “Bayesian Effect Estimation

Accounting for Adjustment Uncertainty” by Chi Wang, Giovanni

Parmigiani, and Francesca Dominici

In Web Appendix A, we provide detailed MCMC algorithms for BAC and TBAC. In Web Appendices B-E, we provide additional simulation results to further evaluate the performances of TBAC and BAC under various situations. A summary of these simulation results is provided in Section 4.1 of the paper.

Web Appendix A: Details of the MCMC Algorithms

BAC

The posterior samples of $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y})$ are obtained by iteratively sampling from $P(\boldsymbol{\alpha}^X | \beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^Y, D)$, $P(\boldsymbol{\alpha}^Y | \beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^X, D)$ and $P(\beta^{\boldsymbol{\alpha}^Y} | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, D)$. The three full conditionals are

$$\begin{aligned} 1) P(\boldsymbol{\alpha}^X | \beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^Y, D) &\stackrel{\text{using A3}}{=} P(\boldsymbol{\alpha}^X | \boldsymbol{\alpha}^Y, \mathbf{X}) = \frac{P(\mathbf{X} | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y) P(\boldsymbol{\alpha}^X | \boldsymbol{\alpha}^Y)}{P(\mathbf{X} | \boldsymbol{\alpha}^Y)} \stackrel{\text{using A2}}{=} \frac{P(\mathbf{X} | \boldsymbol{\alpha}^X) P(\boldsymbol{\alpha}^X | \boldsymbol{\alpha}^Y)}{P(\mathbf{X} | \boldsymbol{\alpha}^Y)} \\ &\propto P(\mathbf{X} | \boldsymbol{\alpha}^X) P(\boldsymbol{\alpha}^X | \boldsymbol{\alpha}^Y), \end{aligned}$$

where, based on Raftery et al. (1997),

$$\begin{aligned} P(\mathbf{X} | \boldsymbol{\alpha}^X) &= \frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2} \Gamma(\frac{\nu}{2}) |\mathbf{I}_n + \phi^2 \mathbf{W}_{\boldsymbol{\alpha}^X} \boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X} \mathbf{W}_{\boldsymbol{\alpha}^X}'|^{1/2}} \\ &\quad \times [\lambda\nu + (\mathbf{X} - \mathbf{W}_{\boldsymbol{\alpha}^X} \boldsymbol{\mu}_{0\boldsymbol{\alpha}^X})' (\mathbf{I}_n + \phi^2 \mathbf{W}_{\boldsymbol{\alpha}^X} \boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X} \mathbf{W}_{\boldsymbol{\alpha}^X}')^{-1} (\mathbf{X} - \mathbf{W}_{\boldsymbol{\alpha}^X} \boldsymbol{\mu}_{0\boldsymbol{\alpha}^X})]^{-\frac{\nu+n}{2}}, \end{aligned}$$

and $\mathbf{W}_{\boldsymbol{\alpha}^X}$ is the design matrix of the exposure regression, $\mathbf{I}_n$ is the $n \times n$ identity matrix, n is the sample size.

$$\begin{aligned} 2) P(\boldsymbol{\alpha}^Y | \beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^X, D) &\stackrel{\text{using A1}}{=} P(\boldsymbol{\alpha}^Y | \boldsymbol{\alpha}^X, \tilde{\mathbf{Y}}) = \frac{P(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y) P(\boldsymbol{\alpha}^Y | \boldsymbol{\alpha}^X)}{P(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}^X)} \stackrel{\text{using A4}}{=} \frac{P(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}^Y) P(\boldsymbol{\alpha}^Y | \boldsymbol{\alpha}^X)}{P(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}^X)} \\ &\propto P(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}^Y) P(\boldsymbol{\alpha}^Y | \boldsymbol{\alpha}^X). \end{aligned}$$

Let $\mathbf{W}_{\boldsymbol{\alpha}^Y}$ be the design matrix of the outcome regression and suppose the observations of X are placed in the first column of $\mathbf{W}_{\boldsymbol{\alpha}^Y}$. Let $\tilde{\mathbf{W}}_{\boldsymbol{\alpha}^Y}$ be the second to the $(M+1)$ th columns

of $\mathbf{W}_{\alpha^Y}$, based on Raftery et al. (1997),

$$\begin{aligned} P(\tilde{\mathbf{Y}}|\alpha^Y) &= \frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma(\frac{\nu}{2})|\mathbf{I}_n + \phi^2\tilde{\mathbf{W}}_{\alpha^Y}\tilde{\Sigma}_{0\alpha^Y}\tilde{\mathbf{W}}_{\alpha^Y}'|^{1/2}} \\ &\quad \times [\lambda\nu + (\tilde{\mathbf{Y}} - \tilde{\mathbf{W}}_{\alpha^Y}\tilde{\boldsymbol{\mu}}_{0\alpha^Y})'(\mathbf{I}_n + \phi^2\tilde{\mathbf{W}}_{\alpha^Y}\tilde{\Sigma}_{0\alpha^Y}\tilde{\mathbf{W}}_{\alpha^Y}')^{-1}(\tilde{\mathbf{Y}} - \tilde{\mathbf{W}}_{\alpha^Y}\tilde{\boldsymbol{\mu}}_{0\alpha^Y})]^{-\frac{\nu+n}{2}}, \end{aligned}$$

where $\tilde{\boldsymbol{\mu}}_{0\alpha^Y}$ is the second to the $(M+1)$ th elements of $\boldsymbol{\mu}_{0\alpha^Y}$ and $\tilde{\Sigma}_{0\alpha^Y}$ is the second to the $(M+1)$ th rows and the second to the $(M+1)$ th columns of $\Sigma_{0\alpha^Y}$.

$$3) P(\beta^{\alpha^Y}|\alpha^X, \alpha^Y, D) \stackrel{\text{using A3}}{=} P(\beta^{\alpha^Y}|\alpha^Y, D).$$

Based on Bernardo and Smith (2000), we obtain

$$\beta^{\alpha^Y}|\alpha^Y, D \sim t_{n+\nu}(\beta_{n\alpha^Y}, \sigma_{n\alpha^Y}^2),$$

where $\beta_{n\alpha^Y}$ is the first element of $\boldsymbol{\theta}_{n\alpha^Y}$, $\sigma_{n\alpha^Y}^2$ is the (1,1) element of $S_{n\alpha^Y}(\mathbf{W}_{\alpha^Y}'\mathbf{W}_{\alpha^Y} + \Sigma_{0\alpha^Y}^{-1})^{-1}$, and

$$\boldsymbol{\theta}_{n\alpha^Y} = (\mathbf{W}_{\alpha^Y}'\mathbf{W}_{\alpha^Y} + \Sigma_{0\alpha^Y}^{-1})^{-1}(\Sigma_{0\alpha^Y}^{-1}\boldsymbol{\mu}_{0\alpha^Y} + \mathbf{W}_{\alpha^Y}'\mathbf{Y})$$

$$S_{n\alpha^Y} = (n + \nu)^{-1}[\nu\lambda + (\mathbf{Y} - \mathbf{W}_{\alpha^Y}\boldsymbol{\theta}_{n\alpha^Y})'\mathbf{Y} + (\boldsymbol{\mu}_{0\alpha^Y} - \boldsymbol{\theta}_{n\alpha^Y})'\Sigma_{0\alpha^Y}^{-1}\boldsymbol{\mu}_{0\alpha^Y}].$$

TBAC

We implement two separate MCMC algorithms to draw from $P(\alpha^X|\mathbf{X})$ and from $P(\beta^{\alpha^Y}, \alpha^Y|D)$.

First, we use the *MC³* method (Madigan and York, 1995) to sample from $P(\alpha^X|\mathbf{X})$ and count the appearance frequency of each α^X . Second, we draw a posterior sample of $(\alpha^Y, \beta^{\alpha^Y})$ to approximate $P(\alpha^Y, \beta_*|D)$. Notice that

$$\begin{aligned} P(\alpha^Y, \beta^{\alpha^Y}|D) &\propto P(\mathbf{Y}|\alpha^Y, \beta^{\alpha^Y})P(\beta^{\alpha^Y}|\alpha^Y)P(\alpha^Y) \\ &= P(\mathbf{Y}|\alpha^Y, \beta^{\alpha^Y})P(\beta^{\alpha^Y}|\alpha^Y)\sum_{\alpha^X} P(\alpha^Y|\alpha^X)P(\alpha^X|\mathbf{X}) \\ &= \sum_{\alpha^X} P_{\alpha^X}(\alpha^Y, \beta^{\alpha^Y}|\mathbf{Y})P(\alpha^X|\mathbf{X}), \end{aligned}$$

where $P_{\alpha^X}(\alpha^Y, \beta^{\alpha^Y}|\mathbf{Y}) = P(\mathbf{Y}|\alpha^Y, \beta^{\alpha^Y})P(\beta^{\alpha^Y}|\alpha^Y)P(\alpha^Y|\alpha^X)$ is the joint posterior of $(\alpha^Y, \beta^{\alpha^Y})$ with prior on α^Y specified as $P(\alpha^Y|\alpha^X)$. For each given α^X , we draw a sample

of $(\boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y})$ from $P_{\boldsymbol{\alpha}^X}(\boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y} | \mathbf{Y})$ by iteratively sampling from $P_{\boldsymbol{\alpha}^X}(\boldsymbol{\alpha}^Y | \beta^{\boldsymbol{\alpha}^Y}, \mathbf{Y})$ and $P_{\boldsymbol{\alpha}^X}(\beta^{\boldsymbol{\alpha}^Y} | \boldsymbol{\alpha}^Y, \mathbf{Y})$. These two full conditionals can be derived as follows:

- 1) $P_{\boldsymbol{\alpha}^X}(\boldsymbol{\alpha}^Y | \beta^{\boldsymbol{\alpha}^Y}, \mathbf{Y}) \stackrel{\text{using A1'}}{=} P_{\boldsymbol{\alpha}^X}(\boldsymbol{\alpha}^Y | \tilde{\mathbf{Y}}) \propto P(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}^Y) P(\boldsymbol{\alpha}^Y | \boldsymbol{\alpha}^X)$
- 2) $P_{\boldsymbol{\alpha}^X}(\beta^{\boldsymbol{\alpha}^Y} | \boldsymbol{\alpha}^Y, \mathbf{Y})$ follows the t -distribution $t_{n+\nu}(\beta_{n\boldsymbol{\alpha}^Y}, \sigma_{n\boldsymbol{\alpha}^Y}^2)$

We take the sample size equal to the frequency of $\boldsymbol{\alpha}^X$ in the Markov chain from $P(\boldsymbol{\alpha}^X | \mathbf{X})$ and combine samples from different $\boldsymbol{\alpha}^X$ s together to obtain the sample of $(\boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y})$ from their joint posterior $P(\boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y} | D)$.

Web Appendix B: Simulation Results in the Presence of Predictors Only

Correlated with X

Consider the true model: $Y_i = \beta X_i + \delta_1^Y U_{1i} + \delta_2^Y U_{2i} + \epsilon_i^Y$, where $i = 1, \dots, 1000$, and ϵ_i^Y are independent $N(0, 1)$. $(X_i, U_{1i}, U_{2i}, U_{3i}, U_{4i})$ are independent normal vectors with mean zero and a covariance matrix, $\Sigma = (\sigma_{kl})_{5 \times 5}$, where $\sigma_{kk} = 1, k = 1, \dots, 5$, $\sigma_{12} = \sigma_{14} = \sigma_{21} = \sigma_{24} = \rho$, $\sigma_{15} = \sigma_{51} = \rho/2$, and all other σ_{kl} 's equal to zero. Under this scenario, U_3 and U_4 are two predictors that are only correlated with X but not with Y (given X). The set of potential confounders $\mathbf{U}$ includes $U_1, \dots, U_4$ as well as 49 additional independent $N(0, 1)$ random variables. In our simulation, ρ is set to 0.6 and $\beta = \delta_1^Y = \delta_2^Y = 0.1$. Five hundred independent data sets were generated. We applied odds priors in both BAC and TBAC, and chose the value of dependence parameter ω to be 2, 4, 10 or ∞ . The results are summarized in Table S1.

For both BAC and TBAC, the standard errors of estimates increase as ω increases. Since the data include two predictors only correlated with X but not correlated with Y (given X), increasing ω assigns higher probabilities to include them in the outcome model, which yields larger standard errors. On the other hand, the bias decreases as ω increases. Since the data include U_1 , a confounder strongly correlated with X but weakly correlated with

Table S1

Comparison of estimates of β from BAC and TBAC using odds priors. BIAS is the difference between the mean of estimates of β and the true value, SEE is the mean of standard error estimates, SSE is the standard error of the estimates of β , MSE is the mean square error, and CP is the coverage probability of the 95% confidence interval or credible interval.

		BIAS	SEE	SSE	MSE	CP
True model		-0.002	0.040	0.040	0.002	0.94
BAC	$\omega = \infty$	0.001	0.072	0.074	0.006	0.93
	$\omega = 10$	0.019	0.055	0.054	0.003	0.93
	$\omega = 4$	0.025	0.050	0.051	0.003	0.90
	$\omega = 2$	0.029	0.047	0.050	0.003	0.88
TBAC	$\omega = \infty$	0.001	0.073	0.074	0.006	0.94
	$\omega = 10$	0.018	0.055	0.054	0.003	0.94
	$\omega = 4$	0.025	0.051	0.051	0.003	0.91
	$\omega = 2$	0.029	0.049	0.049	0.003	0.89

Y , increasing ω assigns higher probability to include this predictor in the outcome model, which yields less bias. In terms of MSE, which balances bias and variation, having ω less than infinity yields smaller MSEs. But in terms of coverage probability of 95% CIs, small ω , such as $\omega = 2$, provides lower coverage probability than desired. Interestingly, $\omega = 10$ has the same coverage probability as $\omega = \infty$ but much smaller MSE, and therefore is a better choice under this simulation scenario.

Web Appendix C: Simulation Results when Exposure Model is Misspecified

Consider the same true outcome model as in the first simulation scenario in the paper:

$Y_i = \beta X_i + \delta_1^Y U_{1i} + \delta_2^Y U_{2i} + \epsilon_i^Y$, where $i = 1, \dots, 1000$, and ϵ_i^Y are independent $N(0, 1)$. The

Table S2

Comparison of estimates of β from BAC and TBAC when exposure model is misspecified. The dependence parameter ω was set to ∞ . BIAS is the difference between the mean of estimates of β and the true value, SEE is the mean of standard error estimates, SSE is the standard error of the estimates of β , MSE is the mean square error, and CP is the coverage probability of the 95% confidence interval or credible interval.

	BIAS	SEE	SSE	MSE	CP
True model	0.000	0.017	0.018	0.0003	0.95
BAC	-0.001	0.017	0.018	0.0003	0.95
TBAC	-0.001	0.017	0.018	0.0003	0.95

set of potential confounders $\mathbf{U}$ includes U_1, U_2 as well as 49 additional random variables. All these potential confounders follow independent $N(0, 1)$. In this scenario, the exposure X is modeled as a non-linear function of U_1 : $X_i = \delta_1^X U_{1i}^3 + \epsilon_i^X$, where ϵ_i^X are independent $N(0, 0.5)$. In our simulation, we set $\beta = \delta_1^Y = \delta_2^Y = 0.1$, and $\delta_1^X = 0.7$. We generated 500 data sets and applied BAC and TBAC with $\omega = \infty$. The results are summarized in Table S2.

Estimates from BAC and TBAC are very similar to each other, both close to the results obtained from the true model. In this simulation scenario, both methods are robust to the misspecification of the exposure model.

Web Appendix D: Simulation Results for Comparing TBAC vs. BAC

Consider the true model: $Y_i = \beta X_i + \delta_1^Y U_{1i} + \delta_2^Y U_{2i} + \delta_3^Y U_{3i} + \delta_4^Y U_{4i} + \epsilon_i^Y$, where $i = 1, \dots, 100$, and ϵ_i^Y are independent $N(0, 1)$. The set of potential confounders $\mathbf{U}$ consists of $U_{1i}, \dots, U_{4i}$, which are independent $N(0, 1)$ random variables. X is modeled by $X_i = \delta_1^X U_{1i} + \delta_2^X U_{2i} + \delta_3^X U_{3i} + \delta_4^X U_{4i} + \epsilon_i^X$, where ϵ_i^X are independent $N(0, 0.5^2)$. In our simulation,

Table S3

Comparison of marginal posterior distributions of α^X . The dependence parameter ω was set to ∞ .

model	$p(\alpha^X D)$ from BAC	$p(\alpha^X D)$ from TBAC
(0,0,1,1)	0.62	0.43
(0,1,1,1)	0.27	0.23
(1,0,1,1)	0.07	0.23
(1,1,1,1)	0.03	0.11

we set $\beta = \delta_1^X = \delta_2^X = \delta_1^Y = \delta_3^Y = 0.1$, $\delta_3^X = \delta_4^X = 0.6$ and $\delta_2^Y = \delta_4^Y = 2$. Five hundred independent data sets were generated and BAC and TBAC with $\omega = \infty$ were applied.

We first compared the marginal posterior distributions of α^X (Table S3). For both BAC and TBAC, all the posterior weights are assigned to models containing U_3 and U_4 since these two predictors are strongly correlated with X . TBAC assigns equal weights to $\alpha^X = (0, 1, 1, 1)$ and $\alpha^X = (1, 0, 1, 1)$ since U_1 and U_2 have the same correlation coefficient with X . In contrast, BAC assigns much higher weight to $\alpha^X = (0, 1, 1, 1)$ than to $\alpha^X = (1, 0, 1, 1)$. This is the result of feedback effect since U_2 is strongly correlated with Y while U_1 is weakly correlated with Y .

We next compared the marginal posterior distribution of α^Y (Table S4). For both BAC and TBAC, the posterior weights are concentrated on models containing U_2 , U_3 and U_4 since these three predictors are highly correlated with either X or Y or both. Large weights are assigned to $\alpha^Y = (0, 1, 1, 1)$, the model not containing U_1 , since U_1 is only weakly correlated with both X and Y . Compared to TBAC, BAC assigns more weight to $\alpha^Y = (0, 1, 1, 1)$. By considering the feedback effect and joint modeling the exposure and outcome models, BAC tends to assign higher weights to more parsimonious models.

Finally, we compared the estimation of exposure effect, β . As shown in Table S5, the estimates from two methods are very similar.

Table S4

Comparison of marginal posterior distributions of α^Y . The dependence parameter ω was set to ∞ .

model	$p(\alpha^Y D)$ from BAC	$p(\alpha^Y D)$ from TBAC
(0,1,1,1)	0.80	0.61
(1,1,1,1)	0.20	0.39

Table S5

Comparison of estimates of β from BAC and TBAC. The dependence parameter ω was set to ∞ . BIAS is the difference between the mean of estimates of β and the true value, SEE is the mean of standard error estimates, SSE is the standard error of the estimates of β , MSE is the mean square error, and CP is the coverage probability of the 95% confidence interval or credible interval.

	BIAS	SEE	SSE	MSE	CP
MLE from model (0,1,1,1)	0.040	0.203	0.200	0.042	0.94
MLE from model (1,1,1,1)	0.001	0.207	0.204	0.042	0.94
BAC	0.032	0.206	0.202	0.042	0.95
TBAC	0.025	0.207	0.201	0.041	0.95

Web Appendix E: Simulation Results for Comparing the MSEs from BAC and TBAC vs. BMA

We used the same models as in the two simulation scenarios in the paper, but considered a smaller sample size of 100. For each simulation scenario, we generated 500 replications. The dependence parameter ω in BAC and TBAC was set to ∞ . The estimation results are summarized in Tables S6.

When sample size is 100, MSE from BMA is lower than those from BAC and TBAC in simulation scenario one but is higher in scenario two. When sample size increases to 1000, as shown in the paper, MSEs of BAC and TBAC are lower in both scenarios.

Table S6

Comparison of MSEs from BAC, TBAC and BMA. The data were generated from the same models as in the two simulation scenarios in the paper, but with sample size 100. The dependence parameter ω was set to ∞ . BIAS is the difference between the mean of estimates of β and the true value, SEE is the mean of standard error estimates, SSE is the standard error of the estimates of β , MSE is the mean square error, and CP is the coverage probability of the 95% confidence interval or credible interval.

		BIAS	SEE	SSE	MSE	CP
Scenario one	BAC	−0.006	0.142	0.152	0.023	0.93
	TBAC	−0.005	0.149	0.155	0.024	0.94
	BMA	0.043	0.116	0.124	0.017	0.93
Scenario two	BAC	0.059	0.162	0.170	0.032	0.92
	TBAC	0.041	0.175	0.178	0.033	0.94
	BMA	0.153	0.129	0.127	0.040	0.76

References

- Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. John Wiley & Sons, England.
- Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review* **63**, 215–232.
- Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**, 179–191.