Supplementary Materials for Transporting Experimental Results using Entropy Balancing

Kevin P. Josey, Seth A. Berkowitz, Debashis Ghosh, Sridharan Raghavan

April 2, 2021

S.1 Transporting Inferences using Entropy Balancing

Recall that another consistent estimator for τ_{TATE} as

$$\hat{\tau}_{\text{DR}} = \sum_{\{i:S_i=1\}} \frac{\hat{\gamma}_i^{\text{PS}} Z_i[Y_i - \hat{\mu}_1(\mathbf{X}_i)]}{\sum_{\{i:S_i=1\}} \hat{\gamma}_i^{\text{PS}} Z_i} - \sum_{\{i:S_i=1\}} \frac{\hat{\gamma}_i^{\text{PS}} (1 - Z_i)[Y_i - \hat{\mu}_0(\mathbf{X}_i)]}{\sum_{\{i:S_i=1\}} \hat{\gamma}_i^{\text{PS}} (1 - Z_i)} + \frac{1}{n_0} \sum_{\{i:S_i=0\}} [\hat{\mu}_1(\mathbf{X}_i) - \hat{\mu}_0(\mathbf{X}_i)] \,.$$

Suppose Assumption 4 holds with $\hat{\mu}_0(\mathbf{X}_i) = \mathbf{c}(\mathbf{X}_i)^T \hat{\boldsymbol{\beta}} \to \mu_0^*(\mathbf{X}_i)$, $\hat{\mu}_1(\mathbf{X}_i) = \mathbf{c}(\mathbf{X}_i)^T \hat{\boldsymbol{\alpha}} \to \mu_1^*(\mathbf{X}_i)$, and $\mu_0^*(\mathbf{X}_i), \mu_1^*(\mathbf{X}_i)$ denote the true means of the potential outcomes. Both $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ can be any estimator of the coefficients in Assumption 4 that satisfies the above consistency. To show that $\hat{\tau}_{\text{CAL}}$ is a consistent estimator for τ_{TATE}, we show the algebraic equivalency of $\hat{\tau}_{\text{CAL}}$ and $\hat{\tau}_{\text{DR}}$ when we substitute $\hat{\gamma}^{\text{PS}}$ with $\hat{\gamma}^{\text{CAL}}$ into $\hat{\tau}_{DR}$. In other words, the calibration estimator of the target population average treatment effect is the same as the augmented estimator if the estimated odds of treatment are found with with the calibration weights (even when the estimated inverse odds are misspecified).

Observe that if $c_1(\mathbf{X}_i) = 1$ for all $i \in \{i : S_i = 1\}$ and $\hat{\theta}_{01} = 1$ then $\sum_{\{i:S_i=1\}} \hat{\gamma}_i^{\text{CAL}} Z_i = \sum_{\{i:S_i=1\}} \hat{\gamma}_i^{\text{CAL}} (1 - Z_i) = n_1$. Given Assumption 4, we can expand the difference between $\hat{\tau}_{\text{CAL}}$ and

$\hat{\tau}_{\text{DR}}$, substituting $\hat{\boldsymbol{\gamma}}^{\text{PS}}$ with $\hat{\boldsymbol{\gamma}}^{\text{CAL}}$, to get

$$
\begin{aligned}
\hat{\tau}_{\text{CAL}} - \hat{\tau}_{\text{DR}} &= \frac{1}{n_1} \sum_{\{i:S_i=1\}} \left[\hat{\gamma}_i^{\text{CAL}} Z_i \hat{\mu}_1(\mathbf{X}_i) - \hat{\gamma}_i^{\text{CAL}} (1-Z_i)\hat{\mu}_0(\mathbf{X}_i)\right] \\
&\quad - \frac{1}{n_0} \sum_{\{i:S_i=0\}} \left[\hat{\mu}_1(\mathbf{X}_i) - \hat{\mu}_0(\mathbf{X}_i)\right] \\
&= \frac{1}{n_1} \sum_{\{i:S_i=1\}} \left[\hat{\gamma}_i^{\text{CAL}} Z_i \mathbf{c}(\mathbf{X}_i)^T \hat{\boldsymbol{\alpha}} - \hat{\gamma}_i^{\text{CAL}} (1-Z_i)\mathbf{c}(\mathbf{X}_i)^T \hat{\boldsymbol{\beta}}\right] - \hat{\boldsymbol{\theta}}_0^T \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\theta}}_0^T \hat{\boldsymbol{\beta}} \\
&= \frac{1}{n_1} \sum_{\{i:S_i=1\}} \left[\hat{\gamma}_i^{\text{CAL}} Z_i \mathbf{c}(\mathbf{X}_i)^T \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\theta}}_0^T \hat{\boldsymbol{\alpha}} - \hat{\gamma}_i^{\text{CAL}} (1-Z_i)\mathbf{c}(\mathbf{X}_i)^T \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\theta}}_0^T \hat{\boldsymbol{\beta}}\right] \\
&= \frac{1}{n_1} \sum_{\{i:S_i=1\}} \left\{\left[\hat{\gamma}_i^{\text{CAL}} Z_i \mathbf{c}(\mathbf{X}_i) - \hat{\boldsymbol{\theta}}_0\right]^T \hat{\boldsymbol{\alpha}} - \left[\hat{\gamma}_i^{\text{CAL}} (1-Z_i)\mathbf{c}(\mathbf{X}_i) - \hat{\boldsymbol{\theta}}_0\right]^T \hat{\boldsymbol{\beta}}\right\} \\
&= \hat{\boldsymbol{\alpha}}^T \left\{\frac{1}{n_1} \sum_{\{i:S_i=1\}} \left[\hat{\gamma}_i^{\text{CAL}} Z_i \mathbf{c}(\mathbf{X}_i) - \hat{\boldsymbol{\theta}}_0\right]\right\} \\
&\quad - \hat{\boldsymbol{\beta}}^T \left\{\frac{1}{n_1} \sum_{\{i:S_i=1\}} \left[\hat{\gamma}_i^{\text{CAL}} (1-Z_i)\mathbf{c}(\mathbf{X}_i) - \hat{\boldsymbol{\theta}}_0\right]\right\} = 0
\end{aligned} \tag{1}
$$

Now let's suppose that $\text{logit}[\rho^*(\mathbf{X})] = \text{logit}[\rho^*(\mathbf{X}; \boldsymbol{\lambda}^*)] = \mathbf{c}(\mathbf{X})^T \boldsymbol{\lambda}^*$ for some $\boldsymbol{\lambda}^* \in \Re^m$, then the entropy balancing approach to transporting experimental results as implemented with equations (8)–(10) in the main text is consistent for τ_{TATE}. The proof of this is via a standard application of M-estimation theory.Stefanski and Boos (2002) Let $\hat{\boldsymbol{\lambda}}_0$ and $\hat{\boldsymbol{\lambda}}_1$ be determined by (8) and $\hat{\tau}$ by (10), both of which are found in the main text. In order to estimate the variance of estimators for τ_{TATE}, we require the trivial estimating equation

$$\boldsymbol{\delta}(S, \mathbf{X}; \boldsymbol{\theta}_0) = (1-S)[\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0].$$

Observe that $\sum_{i=1}^n (1-S_i)(\mathbf{c}(\mathbf{X}_i) - \hat{\boldsymbol{\theta}}_0) = 0$. In the case of estimating τ_{TSATE}, we would ignore $\boldsymbol{\zeta}(\cdot)$ and treat $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0$ as though it were fixed and known. Next, we define the following estimating equations for $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ as

$$
\begin{aligned}
&\boldsymbol{\zeta}_0(S, \mathbf{X}, Z; \boldsymbol{\lambda}_0, \boldsymbol{\theta}_0) = S(1-Z)\exp\left[-\mathbf{c}(\mathbf{X})^T \boldsymbol{\lambda}_0\right]\left[\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0\right] \quad \text{and} \\
&\boldsymbol{\zeta}_1(S, \mathbf{X}, Z; \boldsymbol{\lambda}_1, \boldsymbol{\theta}_0) = SZ\exp\left[-\mathbf{c}(\mathbf{X})^T \boldsymbol{\lambda}_1\right]\left[\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0\right].
\end{aligned}
$$

These equations correspond to the first order conditions derived from (8) in the main text. Notice that

$$
\begin{aligned}
&\sum_{i=1}^n \boldsymbol{\zeta}_0(S_i, \mathbf{X}_i, Z_i; \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0) = \mathbf{0}_m \quad \text{and} \\
&\sum_{i=1}^n \boldsymbol{\zeta}_1(S_i, \mathbf{X}_i, Z_i; \hat{\boldsymbol{\lambda}}_1, \hat{\boldsymbol{\theta}}_0) = \mathbf{0}_m.
\end{aligned}
$$

Finally, we require the estimating equation for τ which is defined as

$$\psi(S, \mathbf{X}, Y, Z; \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \tau) = S\left\{Z \exp\left[-\mathbf{c}(\mathbf{X})^T \boldsymbol{\lambda}_1\right][Y(1) - \tau] - (1 - Z)\exp\left[-\mathbf{c}(\mathbf{X})^T \boldsymbol{\lambda}_0\right] Y(0)\right\}.$$

Again, notice that

$$\sum_{i=1}^{n} \psi(S_i, \mathbf{X}_i, Y_i, Z_i; \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\lambda}}_1, \hat{\tau}_{\text{CAL}}) = 0.$$

We also write $\boldsymbol{\eta} \equiv \left(\boldsymbol{\theta}_0^T, \boldsymbol{\lambda}_0^T, \boldsymbol{\lambda}_1^T\right)^T$ and

$$\boldsymbol{\xi}(S, \mathbf{X}, Z; \boldsymbol{\eta}) \equiv \left[\boldsymbol{\delta}(S, \mathbf{X}; \boldsymbol{\theta}_0)^T, \boldsymbol{\zeta}_0(S, \mathbf{X}, Z; \boldsymbol{\lambda}_0, \boldsymbol{\theta}_0)^T, \boldsymbol{\zeta}_1(\mathbf{X}, Z; \boldsymbol{\lambda}_1, \boldsymbol{\theta}_0)^T\right]^T.$$

For the sake of compactness, we sometimes omit the parameters that characterize these estimating equations while using this function notation.

Remark 1. *The estimating equations for $\hat{\tau}_{MOM}$ simplifies the problem that we defined above. Instead of estimating $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ we only need to estimate $\boldsymbol{\lambda} \in \Re^m$ using the estimating equation*

$$\boldsymbol{\zeta}(S, \mathbf{X}; \boldsymbol{\lambda}, \boldsymbol{\theta}_0) = S \exp\left[-\mathbf{c}(\mathbf{X})^T \boldsymbol{\lambda}\right][\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0].$$

Otherwise the problem remains the same throughout the rest of this Appendix while reducing $(\boldsymbol{\lambda}_0^T, \boldsymbol{\lambda}_1^T)^T$ into $\boldsymbol{\lambda}$ and $\left[\boldsymbol{\zeta}_0(\cdot)^T, \boldsymbol{\zeta}_1(\cdot)^T\right]^T$ into $\boldsymbol{\zeta}(\cdot)$ wherever they appear.

Remark 2. *If the desired estimand is τ_{TSATE}, we would exclude $\boldsymbol{\delta}(\cdot)$ from $\boldsymbol{\xi}(\cdot)$ and $\boldsymbol{\theta}_0$ from $\boldsymbol{\eta}$ since this change would amount to assuming $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0$ is fixed and known. That is, we would substitute*

$$\boldsymbol{\xi}'(S, \mathbf{X}, Z; \boldsymbol{\eta}) \equiv \left[\boldsymbol{\zeta}_0(S, \mathbf{X}, Z; \boldsymbol{\lambda}_0, \hat{\boldsymbol{\theta}}_0)^T, \boldsymbol{\zeta}_1(\mathbf{X}, Z; \boldsymbol{\lambda}_1, \hat{\boldsymbol{\theta}}_0)^T\right]^T$$

and $\boldsymbol{\eta}' \equiv \left(\boldsymbol{\lambda}_0^T, \boldsymbol{\lambda}_1^T\right)^T$. We then substitute $\boldsymbol{\xi}'(\cdot)$ and $\boldsymbol{\eta}'$ throughout.

Remark 3. *When the target sample is drawn from the target population with a known weighting scheme, lets say q_i for all $i \in \{i : S_i = 0\}$, we would instead use the target marginal $\hat{\boldsymbol{\theta}}_0 \equiv \left(\sum_{\{i:S_i=0\}} q_i\right)^{-1} \sum_{\{i:S_i=0\}} q_i \mathbf{c}(\mathbf{X}_i)$. We also modify the estimating equation for $\boldsymbol{\theta}_0$ to be*

$$\boldsymbol{\delta}(S, \mathbf{X}; \boldsymbol{\theta}_0) = (1 - S)[q\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0].$$

The efficient influence function for τ_{TATE} can be written as

$$\begin{aligned}\phi(S, \mathbf{X}, Y, Z; \boldsymbol{\eta}, \tau) &\equiv \psi(S, \mathbf{X}, Y, Z; \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \tau) \\ &\quad - \mathbb{E}\left[\frac{\partial \psi(S, \mathbf{X}, Y, Z; \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \tau)}{\partial \boldsymbol{\eta}}\right]^T \left\{\mathbb{E}\left[\frac{\partial \boldsymbol{\xi}(S, \mathbf{X}, Z; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]\right\}^{-1} \boldsymbol{\xi}(S, \mathbf{X}, Z; \boldsymbol{\eta})\end{aligned}$$

which is used to show

$$\hat{\tau}_{\text{CAL}} - \tau_{\text{TATE}} = \frac{1}{n_1} \sum_{\{i:S_i=1\}} \phi(S_i, \mathbf{X}_i, Y_i, Z_i; \boldsymbol{\eta}^*, \tau_{\text{TATE}}) + o_p\left(n_1^{-1/2}\right) \tag{2}$$

where the true parameter values are $\boldsymbol{\eta}^* = \left[(\boldsymbol{\theta}_0^*)^T, (\boldsymbol{\lambda}_0^*)^T, (\boldsymbol{\lambda}_1^*)^T\right]^T$.Kennedy (2016) Thus, we only need to show that the expected value of $\boldsymbol{\delta}(\cdot)$, $\boldsymbol{\zeta}_0(\cdot)$, $\boldsymbol{\zeta}_1(\cdot)$, and $\psi(\cdot)$ equals zero in order to prove consistency of $\hat{\tau}_{\text{CAL}}$. It is trivial to show that $\mathbb{E}[\boldsymbol{\delta}(S, \mathbf{X}; \boldsymbol{\theta}_0^*)] = \mathbb{E}\left[\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0^* | S = 0\right] = \mathbf{0}_m$. For $\boldsymbol{\zeta}_0(\cdot)$ and $\boldsymbol{\zeta}_1(\cdot)$, note that given $Z \perp\!\!\!\perp \mathbf{X}$, as is typical with an RCT, we have

$$\mathbb{E}[g(\mathbf{X})|S = 0, Z = 0] = \mathbb{E}[g(\mathbf{X})|S = 0, Z = 1] = \mathbb{E}[g(\mathbf{X})|S = 0]$$

for some function $g : \Re^m \to \Re$. Furthermore, we know that $\mathbb{E}\{[1 - \rho(\mathbf{X})]g(\mathbf{X})\} \propto \mathbb{E}\left[g(\mathbf{X})|S = 0\right]$. Thus, we have

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\zeta}_0(S, \mathbf{X}, Z; \boldsymbol{\lambda}_0^*, \boldsymbol{\theta}_0^*)\right] &= \mathbb{E}\left(\mathbb{E}\left\{S(1-Z)\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}_0^*\right]\left[\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0^*\right]\big|\mathbf{X}\right\}\right) \\
&= (1-\pi)\mathbb{E}\left\{\frac{\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}_0^*\right]}{1 + \exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}\left[\mathbf{c}(\mathbf{X}) - \boldsymbol{\theta}_0^*\right]\right\} \\
&= (1-\pi)\mathbb{E}\left(\frac{\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}_0^*\right]}{1 + \exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}\left\{\mathbf{c}(\mathbf{X}) - \mathbb{E}[\mathbf{c}(\mathbf{X})|S = 0]\right\}\right),
\end{aligned}$$

which can only equal zero if $\exp[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}_0^*] = r_0 \exp[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*]$ since

$$\mathbb{E}\left(\frac{\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}{1 + \exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}\left\{\mathbf{c}(\mathbf{X}) - \mathbb{E}[\mathbf{c}(\mathbf{X})|S = 0]\right\}\right) \propto \mathbb{E}\left\{\mathbf{c}(\mathbf{X}) - \mathbb{E}[\mathbf{c}(\mathbf{X})|S = 0]|S = 0\right\} = 0.$$

A similar result can be derived for $\mathbb{E}\left[\boldsymbol{\zeta}_1(S, \mathbf{X}, Z; \boldsymbol{\lambda}_1^*, \boldsymbol{\theta}_0^*)\right]$ where we once again recognize that it is necessary for $\exp[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}_1^*] = r_1 \exp[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*]$. We can also take the expectation of $\psi(\cdot)$ to get

$$\begin{aligned}
\mathbb{E}\left[\psi(S, \mathbf{X}, Y, Z; \boldsymbol{\lambda}_0^*, \boldsymbol{\lambda}_1^*, \tau_{\text{TATE}})\right] &= \mathbb{E}\left(\mathbb{E}\left\{SZ\exp\left[-\mathbf{c}(\mathbf{X}_i)^T\boldsymbol{\lambda}_1^*\right]\left[Y(1) - \tau_{\text{TATE}}\right]\big|\mathbf{X}\right\}\right) \\
&\quad - \mathbb{E}\left(\mathbb{E}\left\{S(1-Z)\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}_0^*\right]Y(0)\big|\mathbf{X}\right\}\right) \\
&= r_1\pi\mathbb{E}\left\{\frac{\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}{1 + \exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}\left[\mu_1(\mathbf{X}) - \tau_{\text{TATE}}\right]\right\} \\
&\quad - r_0(1-\pi)\mathbb{E}\left\{\frac{\exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}{1 + \exp\left[-\mathbf{c}(\mathbf{X})^T\boldsymbol{\lambda}^*\right]}\mu_0(\mathbf{X})\right\} \\
&\propto r_1\pi\mathbb{E}\left[\mu_1(\mathbf{X}) - \tau_{\text{TATE}}|S = 0\right] - r_0(1-\pi)\mathbb{E}\left[\mu_0(\mathbf{X})|S = 0\right].
\end{aligned} \tag{3}$$

For (3) to equate to zero requires $r_1 = \pi^{-1}$ and $r_0 = (1-\pi)^{-1}$. By applying the weak law of large numbers to (2), we conclude $\hat{\tau}_{\text{CAL}} \to_p \tau_{\text{TATE}}$.

With M-estimation, we know

$$n_1^{1/2}(\hat{\tau}_{\text{CAL}} - \tau_{\text{TATE}}) \to_d \mathcal{N}(0, \Sigma^*)$$

where

$$\Sigma^* = \mathbb{E}\left[\phi\left(S, \mathbf{X}, Y, Z; \boldsymbol{\eta}^*, \tau_{\text{TATE}}\right)^2\right]$$

under the weak law of large numbers. Therefore, the robust variance estimator we use is

$$\hat{\Sigma} = \frac{1}{n_1} \sum_{\{i:S_i=1\}} \phi\left(S, \mathbf{X}, Y, Z; \hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\lambda}}_1, \hat{\tau}_{\text{CAL}}\right)^2.$$

Under the conditions of Theorem 3 in Chan *et al.*, $\Sigma^* = \Sigma_{\text{semi}}$.Chan et al. (2016)

S.2 Additional Tables from Simulation

n_0	n_1	Scenario	τ_{TATE}	IOSW	OM	DR	TMLE	MOM	CAL
500	200	baseline	-0.1	1.05 (1.42)	0.64 (0.40)	0.66 (0.44)	0.66 (0.45)	0.93 (1.11)	0.63 (0.44)
500	200	missing	-0.1	1.03 (1.36)	0.72 (0.55)	0.75 (0.61)	0.75 (0.62)	0.94 (1.10)	0.73 (0.62)
500	200	outcome	8.9	0.91 (1.25)	0.90 (0.92)	1.14 (1.69)	1.18 (1.09)	0.80 (0.92)	0.73 (0.83)
500	200	positivity	-0.3	1.09 (1.94)	0.71 (0.51)	0.83 (0.76)	0.85 (2.14)	0.94 (1.83)	0.63 (0.82)
500	200	sampling	3.5	0.84 (1.42)	0.60 (0.36)	0.63 (0.38)	0.89 (1.57)	0.66 (0.46)	0.57 (0.39)
500	200	sparse	-0.1	1.16 (2.09)	0.68 (0.48)	0.75 (0.60)	0.76 (0.95)	0.96 (1.59)	0.62 (0.64)
500	1000	baseline	-0.1	0.79 (0.69)	0.56 (0.31)	0.57 (0.32)	0.57 (0.32)	0.73 (0.57)	0.56 (0.32)
500	1000	missing	-0.1	0.77 (0.65)	0.59 (0.35)	0.60 (0.37)	0.60 (0.37)	0.73 (0.57)	0.60 (0.37)
500	1000	outcome	8.9	0.71 (0.64)	0.63 (0.40)	0.89 (0.84)	0.94 (0.51)	0.64 (0.49)	0.61 (0.44)
500	1000	positivity	-0.3	0.93 (1.20)	0.54 (0.29)	0.64 (0.43)	0.65 (0.62)	0.82 (0.90)	0.59 (0.44)
500	1000	sampling	4.5	0.77 (1.79)	0.50 (0.19)	0.54 (0.22)	0.61 (0.77)	0.53 (0.24)	0.50 (0.19)
500	1000	sparse	-0.1	0.92 (1.08)	0.57 (0.34)	0.62 (0.40)	0.62 (0.45)	0.80 (0.79)	0.59 (0.40)
1000	200	baseline	-0.1	1.05 (1.42)	0.60 (0.37)	0.62 (0.43)	0.62 (0.43)	0.91 (1.08)	0.59 (0.43)
1000	200	missing	-0.1	1.04 (1.36)	0.70 (0.49)	0.73 (0.54)	0.73 (0.54)	0.93 (1.02)	0.70 (0.54)
1000	200	outcome	8.9	0.91 (1.29)	0.90 (0.88)	1.15 (1.68)	1.17 (1.06)	0.78 (0.92)	0.72 (0.83)
1000	200	positivity	-0.3	1.08 (2.00)	0.69 (0.49)	0.82 (0.70)	0.81 (1.76)	0.92 (1.78)	0.60 (0.78)
1000	200	sampling	3.2	0.83 (1.58)	0.58 (0.32)	0.62 (0.37)	1.32 (1.25)	0.63 (0.44)	0.55 (0.36)
1000	200	sparse	-0.1	1.17 (2.01)	0.64 (0.45)	0.73 (0.61)	0.73 (0.92)	0.95 (1.64)	0.57 (0.64)
1000	1000	baseline	-0.1	0.78 (0.69)	0.49 (0.24)	0.50 (0.26)	0.50 (0.26)	0.70 (0.54)	0.50 (0.26)
1000	1000	missing	-0.1	0.77 (0.64)	0.53 (0.27)	0.55 (0.30)	0.55 (0.30)	0.70 (0.52)	0.54 (0.30)
1000	1000	outcome	8.91	0.71 (0.66)	0.63 (0.39)	0.88 (0.86)	0.91 (0.48)	0.63 (0.48)	0.59 (0.44)
1000	1000	positivity	-0.3	0.93 (1.31)	0.50 (0.25)	0.62 (0.40)	0.63 (0.55)	0.81 (0.90)	0.56 (0.40)
1000	1000	sampling	4.0	0.77 (1.90)	0.44 (0.16)	0.49 (0.20)	0.56 (1.37)	0.48 (0.20)	0.44 (0.16)
1000	1000	sparse	-0.1	0.91 (1.01)	0.51 (0.26)	0.57 (0.34)	0.57 (0.37)	0.78 (0.76)	0.53 (0.32)

Table S1: Model standard errors and (empirical standard errors) of τ_{TATE} estimates.

n_0	n_1	Scenario	τ_{TATE}	IOSW	OM	DR	TMLE	MOM	CAL
500	200	baseline	-0.1	0.742	1.592	1.503	1.483	0.836	1.430
500	200	missing	-0.1	0.761	1.312	1.221	1.211	0.855	1.168
500	200	outcome	8.9	0.734	0.978	0.677	1.079	0.869	0.877
500	200	positivity	-0.3	0.562	1.385	1.098	0.398	0.513	0.760
500	200	sampling	3.5	0.590	1.690	1.662	0.565	1.420	1.480
500	200	sparse	-0.1	0.552	1.399	1.242	0.800	0.603	0.980
500	1000	baseline	-0.1	1.145	1.809	1.760	1.770	1.270	1.757
500	1000	missing	-0.1	1.180	1.688	1.630	1.624	1.288	1.616
500	1000	outcome	8.9	1.120	1.568	1.053	1.833	1.313	1.378
500	1000	positivity	-0.3	0.770	1.875	1.486	1.052	0.910	1.342
500	1000	sampling	4.5	0.428	2.712	2.424	0.789	2.253	2.590
500	1000	sparse	-0.1	0.854	1.699	1.525	1.386	1.012	1.476
1000	200	baseline	-0.1	0.738	1.599	1.458	1.467	0.840	1.375
1000	200	missing	-0.1	0.766	1.428	1.344	1.346	0.917	1.284
1000	200	outcome	8.9	0.707	1.027	0.682	1.110	0.850	0.861
1000	200	positivity	-0.3	0.540	1.398	1.162	0.461	0.515	0.770
1000	200	sampling	3.2	0.524	1.784	1.689	1.058	1.429	1.528
1000	200	sparse	-0.1	0.580	1.439	1.192	0.792	0.581	0.893
1000	1000	baseline	-0.1	1.129	2.040	1.948	1.961	1.301	1.938
1000	1000	missing	-0.1	1.193	1.931	1.830	1.836	1.360	1.814
1000	1000	outcome	8.91	1.077	1.629	1.024	1.919	1.320	1.351
1000	1000	positivity	-0.3	0.713	1.970	1.548	1.129	0.894	1.397
1000	1000	sampling	4	0.404	2.846	2.405	0.411	2.418	2.692
1000	1000	sparse	-0.1	0.901	1.954	1.668	1.524	1.033	1.627

Table S2: Ratio between model standard errors/empirical standard errors found for the estimates of τ_{TATE} which are detailed further in Table S1.

References

Chan, K. C. G., Yam, S. C. P., and Zhang, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3):673–700.

Kennedy, E. H. (2016). Semiparametric Theory and Empirical Processes in Causal Inference. In He, H., Wu, P., and Chen, D.-G. D., editors, *Statistical Causal Inferences and Their Applications in Public Health Research*, ICSA Book Series in Statistics, pages 141–167. Springer International Publishing, Cham.

Stefanski, L. A. and Boos, D. D. (2002). The Calculus of M-Estimation. *The American Statistician*, 56(1):29–38.