Online Appendix
What to Observe When Assuming Selection on Observables

Kevin M. Quinn[*] Guoer Liu[†] Lee Epstein[‡] Andrew D. Martin[§]

December 3, 2024

[*]Emory University
[†]University of California, San Diego
[‡]Washington University in St. Louis
[§]Washington University in St. Louis

Contents

A Preliminaries

In Section 2.1 of the manuscript we noted that ATE and ATT can be expressed as Weighted Average Treatment Effects (WATEs) (Li et al., 2018). The Hájek-type estimator of WATE that we consider in the main body of the manuscript is:

$$\hat{\tau}_h = \frac{\sum_{i=1}^n w_h^{(1)}(x_i) Z_i Y_i}{\sum_{i=1}^n w_h^{(1)}(x_i) Z_i} - \frac{\sum_{i=1}^n w_h^{(0)}(x_i)(1 - Z_i) Y_i}{\sum_{i=1}^n w_h^{(0)}(x_i)(1 - Z_i)}. \tag{A.1}$$

B Regression Estimator Weights

In this section, we consider the MRI estimators (Chattopadhyay and Zubizarreta, 2023) for ATE and ATT:

$$\hat{\tau}_{ATE}^{\text{reg}} = \frac{1}{n} \sum_{i=1}^n (\hat{m}_1(x_i) - \hat{m}_0(x_i)) \tag{B.1}$$

$$\hat{\tau}_{ATT}^{\text{reg}} = \frac{1}{n^{(1)}} \sum_{i:Z_i=1} (Y_i - \hat{m}_0(x_i)) \tag{B.2}$$

where $\hat{m}_1(x)$ is the ordinary least squares (OLS) regression estimator of $\mathbb{E}[Y|X = x, Z = 1]$ constructed by subsetting the data to the $Z = 1$ units and fitting an OLS regression to those data and $\hat{m}_0(x)$ is the OLS regression estimator of $\mathbb{E}[Y|X = x, Z = 0]$ constructed by subsetting the data to the $Z = 0$ units and fitting an OLS regression to those data.

We will also consider the OLS estimator of τ in the regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}\tau + \boldsymbol{\epsilon}. \tag{B.3}$$

Following Chattopadhyay and Zubizarreta (2023), we refer to this latter regression estimator as the URI estimator.

B.1 Regression Estimators are WATE Estimators

Chattopadhyay and Zubizarreta (2023) proved that the MRI estimators $\hat{\tau}_{ATE}^{\text{reg}}$ and $\hat{\tau}_{ATT}^{\text{reg}}$ are Hájek-type estimators in the form of Equation A.1 and, in doing so, they derived expressions for the corresponding weights.

We build on this result with a slightly different proof. The resulting weights are identical to the Chattopadhyay and Zubizarreta (2023) weights up to a different normalizing constant. The fact that these two slightly different versions of the MRI weights sum to different constants does not affect Equation A.1 because of the normalization of the weights in the denominators.

We begin by showing how to write the estimator $\hat{\tau}_{ATE}^{\text{reg}}$ as a Hájek-type estimator in the form of Equation A.1.

Let $\mathbf{X}$ denote the $n \times (k + 1)$ matrix of covariates including a constant. Similarly, let $\mathbf{X}_1$ and $\mathbf{X}_0$ denote the submatrices that correspond to the portions of $\mathbf{X}$ from the treated and control units respectively. To make indexing easier, we assume that the data have been sorted so that the first $n^{(1)}$ rows of $\mathbf{X}$ correspond to the treated units and the last $n^{(0)} = n - n^{(1)}$

rows correspond to the control units. The treatment indicator, Z, is not included in $\mathbf{X}$, $\mathbf{X}_0$, or $\mathbf{X}_1$. Relatedly, let $\mathbf{y}$ denote the $n \times 1$ vector of outcomes for the full sample and $\mathbf{y}_1$ and $\mathbf{y}_0$ be the observed outcomes for the treated and control units respectively.

Proposition B.1 (The MRI Estimator of ATE is a Hájek-Type Estimator (Chattopadhyay and Zubizarreta, 2023))**.**

$$\hat{\tau}^{reg}_{ATE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_1(x_i) - \hat{m}_0(x_i)\right) = \frac{\sum_{i=1}^{n} w_i Z_i Y_i}{\sum_{i=1}^{n} w_i Z_i} - \frac{\sum_{i=1}^{n} w_i (1-Z_i) Y_i}{\sum_{i=1}^{n} w_i (1-Z_i)}$$

where $\mathbf{w} = (\mathbf{w}^{(1)'}, \mathbf{w}^{(0)'})'$, $\mathbf{w}^{(1)'} = \mathbf{1}_n'\mathbf{X}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'$ *and* $\mathbf{w}^{(0)'} = \mathbf{1}_n'\mathbf{X}\left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'$. *In words, the MRI estimator of ATE in Equation B.1 can be written as the Hájek-Type estimator in Equation A.1.*

Proof. Let $\hat{\mathbf{m}}_1(\mathbf{X})$ be the n-vector formed by stacking $\hat{m}_1(x_i)$ for $i = 1, \ldots, n$ with $\hat{\mathbf{m}}_0(\mathbf{X})$ defined similarly. Note that $\hat{\mathbf{m}}_1(\mathbf{X}) = \mathbf{P}_1\mathbf{y}_1$ where $\mathbf{P}_1 = \mathbf{X}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'$ and $\hat{\mathbf{m}}_0(\mathbf{X}) = \mathbf{P}_0\mathbf{y}_0$ where $\mathbf{P}_0 = \mathbf{X}\left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'$.

Define $\mathbf{w}^{(1)'} = \mathbf{1}_n'\mathbf{P}_1$ and $\mathbf{w}^{(0)'} = \mathbf{1}_n'\mathbf{P}_0$ where $\mathbf{1}_n$ is an n-vector of ones. Further, let $\mathbf{w} = (\mathbf{w}^{(1)'}, \mathbf{w}^{(0)'})'$ be the n-vector formed by concatenating $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$. w_i is the ith element of $\mathbf{w}$. We can now write Equation B.1 as

$$\begin{aligned}
\hat{\tau}^{\text{reg}}_{ATE} &= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_1(x_i) - \hat{m}_0(x_i)\right) \\
&= \frac{1}{n}\left(\mathbf{1}_n'\mathbf{P}_1\mathbf{y}_1 - \mathbf{1}_n'\mathbf{P}_0\mathbf{y}_0\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i Z_i Y_i - \frac{1}{n}\sum_{i=1}^{n} w_i (1-Z_i) Y_i \\
&= \frac{\sum_{i=1}^{n} w_i Z_i Y_i}{\sum_{i=1}^{n} w_i Z_i} - \frac{\sum_{i=1}^{n} w_i (1-Z_i) Y_i}{\sum_{i=1}^{n} w_i (1-Z_i)}
\end{aligned}$$

where the last line follows from the fact that $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ both sum to n by construction (see Proposition B.4). Thus the MRI estimator of ATE in Equation B.1 can be written as the Hájek-type estimator in Equation A.1 where the weights are $\mathbf{w}^{(1)'} = \mathbf{1}_n'\mathbf{P}_1$ and $\mathbf{w}^{(0)'} = \mathbf{1}_n'\mathbf{P}_0$. Note that these weights can be negative. $\square$

Proposition B.2 (The MRI Estimator of ATT is a WATE Estimator (Chattopadhyay and Zubizarreta, 2023))**.**

$$\hat{\tau}^{reg}_{ATT} = \frac{1}{n^{(1)}} \sum_{i:Z_i=1} (Y_i - \hat{m}_0(x_i)) = \frac{\sum_{i=1}^n w_i Z_i Y_i}{\sum_{i=1}^n w_i Z_i} - \frac{\sum_{i=1}^n w_i (1-Z_i) Y_i}{\sum_{i=1}^n w_i (1-Z_i)}$$

where $\mathbf{w} = (\mathbf{w}^{(1)\prime}, \mathbf{w}^{(0)\prime})'$, $\mathbf{w}^{(1)} = \mathbf{1}_{n^{(1)}}$, *and* $\mathbf{w}^{(0)\prime} = \mathbf{1}'_{n^{(1)}} \mathbf{X}_1 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0'$. *In words, the MRI estimator of ATT in Equation B.2 can be written as the Hájek-type estimator in Equation A.1.*

Proof.

$$\begin{aligned}
\hat{\tau}^{\text{reg}}_{ATT} &= \frac{1}{n^{(1)}} \sum_{i:Z_i=1} (Y_i - \hat{m}_0(x_i)) \\
&= \frac{1}{n^{(1)}} \mathbf{z}'\mathbf{y} - \frac{1}{n^{(1)}} \mathbf{1}'_{n^{(1)}} \mathbf{X}_1 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0' \mathbf{y}_0 \\
&= \frac{\sum_{i=1}^n w_i Z_i Y_i}{\sum_{i=1}^n w_i Z_i} - \frac{\sum_{i=1}^n w_i (1-Z_i) Y_i}{\sum_{i=1}^n w_i (1-Z_i)}
\end{aligned}$$

where $\mathbf{w} = (\mathbf{w}^{(1)\prime}, \mathbf{w}^{(0)\prime})'$, $\mathbf{w}^{(1)} = \mathbf{1}_{n^{(1)}}$, and $\mathbf{w}^{(0)\prime} = \mathbf{1}'_{n^{(1)}} \mathbf{X}_1 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0'$. The last line follows from the fact that $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ both sum to $n^{(1)}$ by construction (see Corollary B.4). □

More generally, we conjecture that any regression estimator in the form of Equation B.1 that constructs the estimates $\hat{m}_1(\cdot)$ and $\hat{m}_0(\cdot)$ with linear smoothers[1] can be rewritten in the form of Equation A.1.

Chattopadhyay and Zubizarreta (2023) also showed that the OLS estimator of τ in Equation B.3 (the URI estimator) is equivalent to the Hájek-type estimator in Equation A.1. Again, we build on that work with a slightly different proof. The resulting weights are identical to the Chattopadhyay and Zubizarreta (2023) URI weights up to a different normalizing constant.

[1] For instance, see the discussion on p. 159 of Bishop (2006).

Begin by defining $\mathbf{U} = (\mathbf{X}, \mathbf{z})$ to be the $n \times (k+2)$ matrix formed by concatenating the vector of treatment indicators $\mathbf{z}$ to the right of the matrix of pre-treatment covariates $\mathbf{X}$. $\mathbf{U}_1$ and $\mathbf{U}_0$ are similarly defined to be $\mathbf{U}_1 = (\mathbf{X}, \mathbf{1}_n)$ and $\mathbf{U}_0 = (\mathbf{X}, \mathbf{0}_n)$ where $\mathbf{0}_n$ is an n-vector of 0s.

Proposition B.3 (The URI Estimator is a WATE Estimator (Chattopadhyay and Zubizarreta))**.** *The OLS estimator, $\hat{\tau}$, of τ in*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}\tau + \boldsymbol{\epsilon}$$

is:

$$\hat{\tau} = \frac{\sum_{i=1}^{n} w_i Z_i Y_i}{\sum_{i=1}^{n} w_i Z_i} - \frac{\sum_{i=1}^{n} w_i (1 - Z_i) Y_i}{\sum_{i=1}^{n} w_i (1 - Z_i)}$$

where the n-vector of weights is:

$$\mathbf{w} = \mathbf{w}^{(1)} \circ \mathbf{z} \;+\; \mathbf{w}^{(0)} \circ (\mathbf{1}_n - \mathbf{z})$$

where $\circ$ denotes the Hadamard (element-by-element) product,

$$\mathbf{w}^{(1)\prime} = \mathbf{1}_n' \mathbf{U}_1 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}' - \mathbf{1}_n' \mathbf{U}_0 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}',$$

and

$$\mathbf{w}^{(0)\prime} = \mathbf{1}_n' \mathbf{U}_0 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}' - \mathbf{1}_n' \mathbf{U}_1 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'.$$

Proof. Define $\hat{\mathbf{m}}^{(1)} = \mathbf{U}_1 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'\mathbf{y}$ and $\hat{\mathbf{m}}^{(0)} = \mathbf{U}_0 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'\mathbf{y}$ and note that

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{m}_i^{(1)} - \hat{m}_i^{(0)} \right).$$

Then write:

$$
\begin{aligned}
\hat{\tau} &= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_i^{(1)} - \hat{m}_i^{(0)}\right) \\
&= \frac{1}{n}\left(\mathbf{1}_n'\mathbf{U}_1(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{y} - \mathbf{1}_n'\mathbf{U}_0(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{y}\right) \\
&= \frac{1}{n}\left(\widetilde{\mathbf{w}}^{(1)'}\mathbf{y} - \widetilde{\mathbf{w}}^{(0)'}\mathbf{y}\right) \\
&= \frac{1}{n}\left(\left(\widetilde{\mathbf{w}}^{(1)} - \widetilde{\mathbf{w}}^{(0)}\right)'\mathbf{y}\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}\left(\widetilde{w}_i^{(1)} - \widetilde{w}_i^{(0)}\right) Z_iY_i + \sum_{i=1}^{n}\left(\widetilde{w}_i^{(1)} - \widetilde{w}_i^{(0)}\right)(1-Z_i)Y_i\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}\left(\widetilde{w}_i^{(1)} - \widetilde{w}_i^{(0)}\right) Z_iY_i - \sum_{i=1}^{n}\left(\widetilde{w}_i^{(0)} - \widetilde{w}_i^{(1)}\right)(1-Z_i)Y_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} w_iZ_iY_i - \frac{1}{n}\sum_{i=1}^{n} w_i(1-Z_i)Y_i \\
&= \frac{\sum_{i=1}^{n} w_iZ_iY_i}{\sum_{i=1}^{n} w_iZ_i} - \frac{\sum_{i=1}^{n} w_i(1-Z_i)Y_i}{\sum_{i=1}^{n} w_i(1-Z_i)}
\end{aligned}
$$

where

$$
w_i = \begin{cases} \left(\widetilde{w}_i^{(1)} - \widetilde{w}_i^{(0)}\right) & \text{if } Z_i = 1 \\ \left(\widetilde{w}_i^{(0)} - \widetilde{w}_i^{(1)}\right) & \text{if } Z_i = 0. \end{cases}
$$

Note that the last line in the derivation follows because, as we prove in Proposition B.4, $n = \sum_{i=1}^{n} w_iZ_i$ and $n = \sum_{i=1}^{n} w_i(1-Z_i)$.

In other words, the n-vector of weights is:

$$
\mathbf{w} = \mathbf{w}^{(1)} \circ \mathbf{z} \ + \ \mathbf{w}^{(0)} \circ (\mathbf{1}_n - \mathbf{z})
$$

where $\circ$ denotes the Hadamard (element-by-element) product,

$$
\mathbf{w}^{(1)'} = \mathbf{1}_n'\mathbf{U}_1(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}' - \mathbf{1}_n'\mathbf{U}_0(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}',
$$

and

$$
\mathbf{w}^{(0)'} = \mathbf{1}_n'\mathbf{U}_0(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}' - \mathbf{1}_n'\mathbf{U}_1(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'.
$$

□

B.2 Properties of Regression WATE Weights

Chattopadhyay and Zubizarreta (2023) proved a number properties of the MRI and URI weights. We exactly restate some of those results (the results regarding the perfect mean covariate balance produced by these weights), restate some results in a very slightly modified form (the results related to the sums of the weights), and present new results for our measures of extrapolation.

We begin by looking at the weights for the MRI estimator of ATE in Equation B.1—$\hat{\tau}^{\text{reg}}_{ATE}$. We assume that $\mathbf{X}$, $\mathbf{X}_1$, and $\mathbf{X}_0$ are as defined above. In addition, we define:

$$\text{EXTRAP}^{(0)} = \frac{\sum_{i=1}^{n} |w_i^{(0)}| \mathbb{I}(w_i^{(0)} < 0)(1 - z_i)}{\sum_{i=1}^{n} w_i^{(0)} \mathbb{I}(w_i^{(0)} \geq 0)(1 - z_i)}$$

and

$$\text{EXTRAP}^{(1)} = \frac{\sum_{i=1}^{n} |w_i^{(1)}| \mathbb{I}(w_i^{(1)} < 0) z_i}{\sum_{i=1}^{n} w_i^{(1)} \mathbb{I}(w_i^{(1)} \geq 0) z_i}$$

as in the body of the paper.

Proposition B.4 (Properties of the Regression Weights for ATE). *The weights,* $\mathbf{w} = (\mathbf{w}^{(1)\prime}, \mathbf{w}^{(0)\prime})'$ *in the following estimator:*

$$\hat{\tau}^{reg}_{ATE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{m}_1(x_i) - \hat{m}_0(x_i)) = \frac{\sum_{i=1}^{n} w_i Z_i Y_i}{\sum_{i=1}^{n} w_i Z_i} - \frac{\sum_{i=1}^{n} w_i (1 - Z_i) Y_i}{\sum_{i=1}^{n} w_i (1 - Z_i)}$$

where $\mathbf{w}^{(1)\prime} = \mathbf{1}_n' \mathbf{X} \left(\mathbf{X}_1' \mathbf{X}_1\right)^{-1} \mathbf{X}_1'$ *and* $\mathbf{w}^{(0)\prime} = \mathbf{1}_n' \mathbf{X} \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0'$*, have the following properties:*

1. $\sum_{i=1}^{n^{(0)}} w_i^{(0)} = n$
2. $\sum_{i=1}^{n^{(1)}} w_i^{(1)} = n$
3. $0 \leq \text{EXTRAP}^{(0)} < 1$
4. $0 \leq \text{EXTRAP}^{(1)} < 1$
5. *ASMD*$(x) = 0$ *for all* x *variables included in* $\mathbf{X}$
6. *TASMD*$(x^{(0)}) = 0$ *for all* x *variables included in* $\mathbf{X}$
7. *TASMD*$(x^{(1)}) = 0$ *for all* x *variables included in* $\mathbf{X}$

Proof. Recall that for this estimator, $\mathbf{P}_1 = \mathbf{X}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'$, $\mathbf{P}_0 = \mathbf{X}\left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'$, $\mathbf{w}^{(1)'} = \mathbf{1}_n'\mathbf{P}_1$ and $\mathbf{w}^{(0)'} = \mathbf{1}_n'\mathbf{P}_0$ where $\mathbf{1}_n$ is an n-vector of ones.

Looking at $\mathbf{w}^{(1)'}$ (the weights applied to the treated units) we have:

$$\begin{aligned}\mathbf{w}^{(1)'} &= \mathbf{1}_n'\mathbf{P}_1\\ &= \mathbf{1}_n'(\mathbf{X}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1').\end{aligned}$$

Transposing both sides we get:

$$\begin{aligned}\mathbf{w}^{(1)} &= (\mathbf{X}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1')'\mathbf{1}_n\\ &= \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}'\mathbf{1}_n.\end{aligned}$$

Premultiplying both sides by $\mathbf{X}_1'$ we get:

$$\begin{aligned}\mathbf{X}_1'\mathbf{w}^{(1)} &= \mathbf{X}_1'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}'\mathbf{1}_n\\ \mathbf{X}_1'\mathbf{w}^{(1)} &= \mathbf{X}'\mathbf{1}_n = \mathbf{s}\end{aligned}$$

Analogous arguments give us $\mathbf{X}_0'\mathbf{w}^{(0)} = \mathbf{X}'\mathbf{1}_n = \mathbf{s}$. We thus have:

$$\mathbf{X}_1'\mathbf{w}^{(1)} = \mathbf{X}_0'\mathbf{w}^{(0)} = \mathbf{X}'\mathbf{1}_n = \mathbf{s} \tag{B.4}$$

Note that $\mathbf{s}$ is the $(k+1)\times 1$ vector holding the column sums of $\mathbf{X}$. Two facts follow from Equation B.4.

First, since the first columns of $\mathbf{X}$, $\mathbf{X}_1$, and $\mathbf{X}_0$ are all vectors of ones, $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ must both sum to n. This proves the first two properties. Chattopadhyay and Zubizarreta (2023) prove an analogous result for MRI weights that are proportional to our $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ but that sum to 1.

The fact that $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ both sum to $n > 0$ implies that the denominators of EXTRAP$^{(0)}$ and EXTRAP$^{(1)}$ are positive and that the numerators are less than the respective denominators. This, together with the absolute value function in the numerators ensures that properties 3 and 4 are true.

Finally, Chattopadhyay and Zubizarreta (2023) have proved versions of properties 5, 6, and 7 for their MRI weights. Those results obviously carry over to our weights which are proportional to theirs.

One can also demonstrate this directly with our notation. With $\mathbf{w} = (\mathbf{w}^{(1)'}, \mathbf{w}^{(0)'})'$ we have

$$\frac{\sum_{i=1}^n w_i z_i x_{ik}}{\sum_{i=1}^n w_i z_i} = \frac{\sum_{i=1}^n w_i(1-z_i)x_{ik}}{\sum_{i=1}^n w_i(1-z_i)} = \frac{1}{n}\sum_{i=1}^n x_{ik}$$

where w_i is the ith element of $\mathbf{w}$ and x_{ik} is the value of the kth column of $\mathbf{X}$ for the ith unit. In words, the weighted mean of any column of $\mathbf{X}_1$ is equal to the weighted mean of the same column of $\mathbf{X}_0$ (with weights $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ respectively); and both are equal to the simple sample mean of the same column of $\mathbf{X}$. In other words, the regression estimator, by construction, will always produce perfect weighted mean balance on all measured covariates included in $\mathbf{X}$—both in terms of ASMD and TASMD. This proves properties 5, 6, and 7. $\square$

We next look at the weights for the MRI estimator of ATE in Equation B.1— $\hat{\tau}^{\text{reg}}_{ATE}$.

Corollary B.4 (Properties of the MRI Weights for ATT)**.** *The weights,* $\mathbf{w} = (\mathbf{w}^{(1)\prime}, \mathbf{w}^{(0)\prime})'$ *in the following estimator:*

$$\hat{\tau}^{reg}_{ATT} = \frac{1}{n^{(1)}} \sum_{i:Z_i=1} (Y_i - \hat{m}_0(x_i)) = \frac{\sum_{i=1}^n w_i Z_i Y_i}{\sum_{i=1}^n w_i Z_i} - \frac{\sum_{i=1}^n w_i (1-Z_i) Y_i}{\sum_{i=1}^n w_i (1-Z_i)}$$

where $\mathbf{w}^{(1)} = \mathbf{1}_{n^{(1)}}$, *and* $\mathbf{w}^{(0)\prime} = \mathbf{1}'_{n^{(1)}} \mathbf{X}_1 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0'$. *have the following properties:*

1. $\sum_{i=1}^{n^{(0)}} w_i^{(0)} = n^{(1)}$
2. $\sum_{i=1}^{n^{(1)}} w_i^{(1)} = n^{(1)}$
3. $0 \leq \text{EXTRAP}^{(0)} < 1$
4. $0 \leq \text{EXTRAP}^{(1)} < 1$
5. *ASMD*$(x) = 0$ *for all x variables included in* $\mathbf{X}$
6. *TASMD*$(x^{(0)}) = 0$ *for all x variables included in* $\mathbf{X}$
7. *TASMD*$(x^{(1)}) = 0$ *for all x variables included in* $\mathbf{X}$

Proof. Again, Chattopadhyay and Zubizarreta (2023) have proven essential the same results with the exception of 3 and 4 which reference our measure of extrapolation.

It is also straightforward to deduce these results using logic analogous to that in the proof of Proposition B.4.

Begin with $\mathbf{w}^{(0)\prime}$.

$$\mathbf{w}^{(0)\prime} = \mathbf{1}'_{n^{(1)}} \mathbf{X}_1 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0'$$

Transposing both sides we get:

$$\begin{aligned} \mathbf{w}^{(0)} &= (\mathbf{X}_1 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_0')' \mathbf{1}_{n^{(1)}} \\ &= \mathbf{X}_0 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_1' \mathbf{1}_{n^{(1)}}. \end{aligned}$$

Premultiplying both sides by $\mathbf{X}_0'$ we get:

$$\begin{aligned} \mathbf{X}_0' \mathbf{w}^{(0)} &= \mathbf{X}_0' \mathbf{X}_0 \left(\mathbf{X}_0' \mathbf{X}_0\right)^{-1} \mathbf{X}_1' \mathbf{1}_{n^{(1)}} \\ \mathbf{X}_0' \mathbf{w}^{(0)} &= \mathbf{X}_1' \mathbf{1}_{n^{(1)}} = \mathbf{s}_1 \end{aligned}$$

In words, the weighted column sums of $\mathbf{X}_0$ (with weights equal to $\mathbf{w}^{(0)}$) equal the weighted column sums of X_1 (with weights equal to $\mathbf{w}^{(1)} = \mathbf{1}_{n^{(1)}}$) which are the same as the unweighted column sums of X_1. Logic similar to that in the proof of Proposition B.4 completes the proof. □

Finally, Chattopadhyay and Zubizarreta (2023) have proved a number of related results for the URI estimator. We restate those results[2] in our notation and also provide additional results related to our measure of extrapolation. In doing so, we use the same notation as used earlier in this section and write the URI regression estimator in the form of the Hájek-type estimator from Proposition B.3.

Proposition B.5 (Properties of the URI Estimator). *The weights,* $\mathbf{w}$*, in the following estimator:*

$$\hat{\tau} = \frac{\sum_{i=1}^{n} w_i Z_i Y_i}{\sum_{i=1}^{n} w_i Z_i} - \frac{\sum_{i=1}^{n} w_i (1 - Z_i) Y_i}{\sum_{i=1}^{n} w_i (1 - Z_i)}$$

where the n*-vector of weights is:*

$$\mathbf{w} = \mathbf{w}^{(1)} \circ \mathbf{z} \ + \ \mathbf{w}^{(0)} \circ (\mathbf{1}_n - \mathbf{z})$$

where $\circ$ *denotes the Hadamard (element-by-element) product,*

$$\mathbf{w}^{(1)'} = \mathbf{1}_n' \mathbf{U}_1 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}' - \mathbf{1}_n' \mathbf{U}_0 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}',$$

and

$$\mathbf{w}^{(0)'} = \mathbf{1}_n' \mathbf{U}_0 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}' - \mathbf{1}_n' \mathbf{U}_1 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'.$$

have the following properties:

1. $\sum_{i=1}^{n^{(0)}} w_i^{(0)} = n$
2. $\sum_{i=1}^{n^{(1)}} w_i^{(1)} = n$
3. $0 \leq \text{EXTRAP}^{(0)} < 1$
4. $0 \leq \text{EXTRAP}^{(1)} < 1$
5. *ASMD*$(x) = 0$ *for all* x *variables included in* $\mathbf{X}$
6. *If the estimand is ATE, TASMD*$(x^{(0)}) = 0$ *for all* x *variables included in* $\mathbf{X}$
7. *If the estimand is ATE, TASMD*$(x^{(1)}) = 0$ *for all* x *variables included in* $\mathbf{X}$

Proof. Recall from the proof of Proposition B.3 that

$$\widetilde{\mathbf{w}}^{(1)'} = \mathbf{1}_n' \mathbf{U}_1 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'$$

[2] Results 1 and 2 about the sum of the weights look different, but are effectively the same as the Chattopadhyay and Zubizarreta (2023) results once one realizes that the weights are differently normalized in their paper and our paper.

Transposing both sides we get:

$$\begin{aligned}\widetilde{\mathbf{w}}^{(1)} &= (\mathbf{U}_1 \left(\mathbf{U}'\mathbf{U}\right)^{-1} \mathbf{U}')'\mathbf{1}_n \\ &= \mathbf{U} \left(\mathbf{U}'\mathbf{U}\right)^{-1} \mathbf{U}_1'\mathbf{1}_n.\end{aligned}$$

Premultiplying both sides by $\mathbf{U}'$ we get:

$$\begin{aligned}\mathbf{U}'\widetilde{\mathbf{w}}^{(1)} &= \mathbf{U}'\mathbf{U} \left(\mathbf{U}'\mathbf{U}\right)^{-1} \mathbf{U}_1'\mathbf{1}_n \\ \mathbf{U}'\widetilde{\mathbf{w}}^{(1)} &= \mathbf{U}_1'\mathbf{1}_n = \mathbf{s}_1\end{aligned}$$

An analogous argument gives

$$\mathbf{U}'\widetilde{\mathbf{w}}^{(0)} = \mathbf{U}_0'\mathbf{1}_n = \mathbf{s}_0$$

Since the last column of $\mathbf{U}$ is $\mathbf{z}$, the last column of $\mathbf{U}_1$ is $\mathbf{1}_n$, and the last column of $\mathbf{U}_0$ is $\mathbf{0}_n$, these expressions tell us that $\mathbf{z}'\widetilde{\mathbf{w}}^{(1)} = n$ and $\mathbf{z}'\widetilde{\mathbf{w}}^{(0)} = 0$. Thus $\mathbf{z}'\mathbf{w} = \mathbf{z}'(\widetilde{\mathbf{w}}^{(1)} - \widetilde{\mathbf{w}}^{(0)}) = n$. In words, $\mathbf{w}^{(1)}$ sums to n. An analogous argument can be used to show that $(\mathbf{1}_n - \mathbf{z})'\mathbf{w} = (\mathbf{1}_n - \mathbf{z})'(\widetilde{\mathbf{w}}^{(0)} - \widetilde{\mathbf{w}}^{(1)}) = n$. In words, $\mathbf{w}^{(0)}$ sums to n. This proves properties 1 and 2.

The fact that $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ both sum to $n > 0$ implies that the denominators of EXTRAP$^{(0)}$ and EXTRAP$^{(1)}$ are positive and that the numerators are less than the respective denominators. This, together with the absolute value function in the numerators ensures that properties 3 and 4 are true.

Finally, note that since $\mathbf{U}$, $\mathbf{U}_1$ and $\mathbf{U}_0$ only differ on their last column (which corresponds to the observed or hypothetical value of z) and because $\mathbf{z}'\widetilde{\mathbf{w}}^{(0)} = 0$ and $(\mathbf{1}_n - \mathbf{z})'\widetilde{\mathbf{w}}^{(1)} = 0$ it will be the case that

$$\mathbf{x}_k'(\mathbf{w} \circ \mathbf{z}) = \mathbf{x}_k'(\widetilde{\mathbf{w}}^{(1)} \circ \mathbf{z}) - \mathbf{x}_k'(\widetilde{\mathbf{w}}^{(0)} \circ \mathbf{z}) = \mathbf{x}_k'(\widetilde{\mathbf{w}}^{(1)} \circ \mathbf{z}) = \mathbf{x}_k'\mathbf{1}_n$$

and

$$\mathbf{x}_k'(\mathbf{w} \circ (\mathbf{1}_n - \mathbf{z})) = \mathbf{x}_k'(\widetilde{\mathbf{w}}^{(0)} \circ (\mathbf{1}_n - \mathbf{z})) - \mathbf{x}_k'(\widetilde{\mathbf{w}}^{(1)} \circ (\mathbf{1}_n - \mathbf{z})) = \mathbf{x}_k'(\widetilde{\mathbf{w}}^{(0)} \circ (\mathbf{1}_n - \mathbf{z})) = \mathbf{x}_k'\mathbf{1}_n$$

where $\mathbf{x}_k$ is an arbitrary column of $\mathbf{X}$. This, along with the fact that $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ each sum to n, implies:

$$\frac{\sum_{i=1}^n w_i z_i x_{ik}}{\sum_{i=1}^n w_i z_i} = \frac{\sum_{i=1}^n w_i (1 - z_i) x_{ik}}{\sum_{i=1}^n w_i (1 - z_i)} = \frac{1}{n}\sum_{i=1}^n x_{ik}$$

where w_i is the ith element of $\mathbf{w}$ and x_{ik} is the value of the kth column of $\mathbf{X}$ for the ith unit. In words, the weighted mean of any column of $\mathbf{X}_1$ is equal to the weighted mean of the same column of $\mathbf{X}_0$ (with weights $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(0)}$ respectively); and both are equal to the simple sample mean of the same column of $\mathbf{X}$. In other words, the constant effects regression estimator, by construction, will always produce perfect weighted mean balance on all measured covariates included in $\mathbf{X}$ in terms of ASMD. If the estimand is ATE, TASMD($x^{(1)}$) and TASMD($x^{(0)}$) will also be equal to 0 for all x variables included in $\mathbf{X}$. This proves properties 5, 6, and 7. $\square$

C WATE Weights for 1:M Matching with Replacement

In this appendix, we derive the Hájek-type estimator weights for 1:M matching with replacement. The algorithm used to create the matches does not affect the results below.

Following Abadie and Imbens (2006) we define

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } Z_i = 0 \\ \frac{1}{M} \sum_{j \in \mathcal{J}_i} Y_j & \text{if } Z_i = 1 \end{cases}$$

and

$$\hat{Y}_i(1) = \begin{cases} Y_i & \text{if } Z_i = 1 \\ \frac{1}{M} \sum_{j \in \mathcal{J}_i} Y_j & \text{if } Z_i = 0 \end{cases}$$

where $\mathcal{J}_i$ is the set of M unit indices of the units matched to unit i. Define

$$K_i = \sum_{l=1}^{n} \mathbb{I}(i \in \mathcal{J}_l)$$

to be the number of times that unit i is used as a match. Since we are considering matching with replacement, K_i can be greater than 1.

Following Abadie and Imbens (2006, p. 241), the 1:M matching estimator of ATE can be written as:

$$\begin{aligned}
\hat{\tau}^M_{ATE} &= \frac{1}{n} \sum_{i=1}^{n} \left(\hat{Y}_i(1) - \hat{Y}_i(0) \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} (2Z_i - 1) \left(1 + \frac{K_i}{M} \right) Y_i \\
&= \frac{1}{n} \sum_{i:Z_i=1} \left(1 + \frac{K_i}{M} \right) Y_i - \frac{1}{n} \sum_{i:Z_i=0} \left(1 + \frac{K_i}{M} \right) Y_i \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(1 + \frac{K_i}{M} \right) Z_i Y_i - \frac{1}{n} \sum_{i=1}^{n} \left(1 + \frac{K_i}{M} \right) (1 - Z_i) Y_i \\
&= \frac{\sum_{i=1}^{n} w_i^{(1)} Z_i Y_i}{\sum_{i=1}^{n} w_i^{(1)} Z_i} - \frac{\sum_{i=1}^{n} w_i^{(0)} (1 - Z_i) Y_i}{\sum_{i=1}^{n} w_i^{(0)} (1 - Z_i)}
\end{aligned}$$

where $w_i^{(1)} = w_i^{(0)} = \left(1 + \frac{K_i}{M}\right)$ and the last line follows from the fact that $\sum_{i=1}^{n} \left(1 + \frac{K_i}{M}\right) Z_i = \sum_{i=1}^{n} \left(1 + \frac{K_i}{M}\right) (1 - Z_i) = n$. Thus, this matching estimator can be rewritten as the Hájek-type estimator in Equation A.1.

Relatedly, the 1:M matching estimator of ATT can be written as:

$$
\begin{aligned}
\hat{\tau}^{M}_{ATT} &= \frac{1}{n^{(1)}} \sum_{i:Z_i=1} \left(Y_i - \hat{Y}_i(0) \right) \\
&= \frac{1}{n^{(1)}} \sum_{i=1}^{n} \left(Z_i - (1-Z_i)\frac{K_i}{M} \right) Y_i \\
&= \frac{1}{n^{(1)}} \sum_{i:Z_i=1} Y_i - \frac{1}{n^{(1)}} \sum_{i:Z_i=0} \frac{K_i}{M} Y_i \\
&= \frac{1}{n^{(1)}} \sum_{i=1}^{n} Z_i Y_i - \frac{1}{n^{(1)}} \sum_{i=1}^{n} \frac{K_i}{M} (1-Z_i) Y_i \\
&= \frac{\sum_{i=1}^{n} w_i^{(1)} Z_i Y_i}{\sum_{i=1}^{n} w_i^{(1)} Z_i} - \frac{\sum_{i=1}^{n} w_i^{(0)} (1-Z_i) Y_i}{\sum_{i=1}^{n} w_i^{(0)} (1-Z_i)}
\end{aligned}
$$

where $w_i^{(1)} = 1$ if $Z_i = 1$ and $w_i^{(0)} = \frac{K_i}{M}$ if $Z_i = 0$ and the last line follows from the fact that $\sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} \left(\frac{K_i}{M}\right)(1 - Z_i) = n^{(1)}$. Thus, this matching estimator can also be rewritten as the Hájek-type estimator in Equation A.1.

D Effective Sample Size Derivations

Rewrite Equation A.1 as:

$$\hat{\tau}_h = \frac{\sum_{i:Z_i=1} w_{hi}^{(1)} Y_i}{\sum_{i:Z_i=1} w_{hi}^{(1)}} - \frac{\sum_{i:Z_i=0} w_{hi}^{(0)} Y_i}{\sum_{i:Z_i=0} w_{hi}^{(0)}}.$$

Assume that

$$\mathbb{V}[Y|Z=1] = \mathbb{V}[Y|Z=0] = \mathbb{V}[Y] = \sigma^2.$$

The variance of $\hat{\tau}_h$ given the weights is:

$$\begin{aligned}
\mathbb{V}\left[\hat{\tau}_h\right] &= \frac{\sum\limits_{i:Z_i=1} (w_{hi}^{(1)})^2 \mathbb{V}\left[Y_i\right]}{\left(\sum\limits_{i:Z_i=1} w_{hi}^{(1)}\right)^2} + \frac{\sum\limits_{i:Z_i=0} (w_{hi}^{(0)})^2 \mathbb{V}\left[Y_i\right]}{\left(\sum\limits_{i:Z_i=0} w_{hi}^{(0)}\right)^2} \\
&= \frac{\sigma^2 \sum\limits_{i:Z_i=1} (w_{hi}^{(1)})^2}{\left(\sum\limits_{i:Z_i=1} w_{hi}^{(1)}\right)^2} + \frac{\sigma^2 \sum\limits_{i:Z_i=0} (w_{hi}^{(0)})^2}{\left(\sum\limits_{i:Z_i=0} w_{hi}^{(0)}\right)^2} \\
&= \frac{\sigma^2}{n_{\text{eff}}^{(1)}} + \frac{\sigma^2}{n_{\text{eff}}^{(0)}} \\
&= \frac{\sigma^2 n_{\text{eff}}^{(0)} + \sigma^2 n_{\text{eff}}^{(1)}}{n_{\text{eff}}^{(1)} n_{\text{eff}}^{(0)}} \\
&= \sigma^2 \frac{n_{\text{eff}}^{(0)} + n_{\text{eff}}^{(1)}}{n_{\text{eff}}^{(1)} n_{\text{eff}}^{(0)}}
\end{aligned}$$

Note that the validity of this expression does not depend on the weights being positive.

By analogy to $\mathbb{V}[\bar{Y}] = \mathbb{V}[Y]/n$, where $\bar{Y}$ is the sample mean of Y (see Kish (1965)), we say that the effective sample size for this weighted estimator is

$$n_{\text{eff}_w} = \frac{n_{\text{eff}}^{(1)} n_{\text{eff}}^{(0)}}{n_{\text{eff}}^{(0)} + n_{\text{eff}}^{(1)}} \tag{D.1}$$

where

$$n_{\text{eff}}^{(0)} = \frac{\left(\sum\limits_{i:Z_i=0} w_{hi}^{(0)}\right)^2}{\sum\limits_{i:Z_i=0} (w_{hi}^{(0)})^2}$$

and

$$n_{\text{eff}}^{(1)} = \frac{\left(\sum_{i:Z_i=1} w_{hi}^{(1)}\right)^2}{\sum_{i:Z_i=1} (w_{hi}^{(1)})^2}.$$

Note that $n_{\text{eff}}^{(0)}$, $n_{\text{eff}}^{(1)}$, and thus n_{eff_w} do not change if the weights are rescaled by a constant term.[3]

A special case of the above occurs when the weights are constant within the treated and control groups. This situation is equivalent to the difference of conditional means estimator

$$\hat{\tau} = \frac{1}{n^{(1)}} \sum_{i:Z_i=1} Y_i - \frac{1}{n^{(0)}} \sum_{i:Z_i=0} Y_i$$

where $n^{(1)} = \sum_{i=1}^{n} Z_i$ and $n^{(0)} = \sum_{i=1}^{n}(1 - Z_i)$ are the number of treated and control units respectively. Again assuming

$$\mathbb{V}[Y|Z = 1] = \mathbb{V}[Y|Z = 0] = \mathbb{V}[Y] = \sigma^2.$$

we calculate the variance of this estimator as

$$\begin{aligned}
\mathbb{V}\left[\hat{\tau}\right] &= \frac{\sigma^2}{n^{(1)}} + \frac{\sigma^2}{n^{(0)}} \\
&= \frac{\sigma^2 n^{(0)} + \sigma^2 n^{(1)}}{n^{(1)} n^{(0)}} \\
&= \sigma^2 \frac{n^{(0)} + n^{(1)}}{n^{(1)} n^{(0)}}
\end{aligned}$$

By analogy to the variance of the sample mean ($\mathbb{V}[\bar{Y}] = \mathbb{V}[Y]/n$), we say that the effective sample size for this estimator is

$$\frac{n^{(1)} n^{(0)}}{n^{(0)} + n^{(1)}}.$$

Note the similarity to the expression for n_{eff_w} in Equation D.1.

A special case of this occurs when there are equal numbers of treated and control units: $n^{(1)} = n^{(0)} = n/2$. Substituting and simplifying, we get get an effective sample size of $n/4$.

We next prove the claim made in the main body of the manuscript that:

$$0 < n_{\text{eff}}^{(0)} \leq n^{(0)}$$

$$0 < n_{\text{eff}}^{(1)} \leq n^{(1)}$$

[3]Removing unnecessary superscripts and subscripts, we have $\left(\sum_i cw_i\right)^2 / \left(\sum_i (cw_i)^2\right) = \left(c \sum_i w_i\right)^2 / \left(\sum_i c^2 w_i^2\right) = \left(c^2 (\sum_i w_i)^2\right) / \left(c^2 \sum_i w_i^2\right) = \left(\sum_i w_i\right)^2 / \left(\sum_i w_i^2\right)$ for an arbitrary constant c.

and

$$0 < n_{\mathrm{eff}_w} \leq \frac{n^{(1)}n^{(0)}}{n^{(0)} + n^{(1)}} \leq n/4.$$

We begin with a lemma.

Lemma D.1. *Let n be an integer greater than or equal to 1 and let $w_i \in \mathbb{R}, i = 1, \ldots, n$ be a sequence of weights with at least one $w_i \neq 0$. Then*

$$n^* = \frac{\left(\sum_{i=1}^n w_i\right)^2}{\sum_{i=1}^n w_i^2} \leq n.$$

Proof. Let $\mathbf{w} = (w_1, \ldots, w_n)'$ and let $\mathbf{1}_n$ be an n-vector of 1s. Then write

$$n^* = \frac{\left(\mathbf{w}'\mathbf{1}_n\right)^2}{\mathbf{w}'\mathbf{w}}$$

By the Cauchy-Schwarz inequality we know

$$\left(\mathbf{w}'\mathbf{1}_n\right)^2 \leq \left(\mathbf{w}'\mathbf{w}\right)\left(\mathbf{1}_n'\mathbf{1}_n\right).$$

Thus

$$n^* = \frac{\left(\mathbf{w}'\mathbf{1}_n\right)^2}{\mathbf{w}'\mathbf{w}} \leq \mathbf{1}_n'\mathbf{1}_n = n.$$

which completes the proof. □

Proposition D.2 (Bounds on the Effective Sample Size)**.** *Let n be an integer greater than or equal to 2, $n^{(1)}$ be a positive integer less than n, and $n^{(0)} = n - n^{(1)}$. Further, let $w_i \in \mathbb{R}$ such that $\sum_{i=1}^{n^{(1)}} w_i \neq 0$ and $\sum_{i=(n^{(1)}+1)}^{n} w_i \neq 0$. Define*

$$n_{eff_w} = \frac{n_{eff}^{(1)} n_{eff}^{(0)}}{n_{eff}^{(0)} + n_{eff}^{(1)}}$$

where

$$n_{eff}^{(0)} = \frac{\left(\sum_{i=(n^{(1)}+1)}^{n} w_i\right)^2}{\sum_{i=(n^{(1)}+1)}^{n} w_i^2}$$

and

$$n_{eff}^{(1)} = \frac{\left(\sum_{i=1}^{n^{(1)}} w_i\right)^2}{\sum_{i=1}^{n^{(1)}} w_i^2}.$$

Then:

$$0 < n_{eff}^{(0)} \leq n^{(0)} \tag{D.2}$$

$$0 < n_{eff}^{(1)} \leq n^{(1)} \tag{D.3}$$

$$0 < n_{eff_w} \leq \frac{n^{(1)} n^{(0)}}{n^{(0)} + n^{(1)}} \leq n/4. \tag{D.4}$$

Proof. Inequalities D.2 and D.3 follows directly from Lemma D.1.

For inequality D.3, begin by considering how to choose $n^{(0)}$ and $n^{(1)}$ to maximize

$$\frac{n^{(1)} n^{(0)}}{n^{(0)} + n^{(1)}} = \frac{(n - n^{(0)}) n^{(0)}}{n}.$$

If we treat n as fixed and $n^{(0)}$ as continuous, elementary calculus reveals that this expression is maximized when $n^{(0)} = (1/2)n = n^{(1)}$. From Lemma D.1 we know that $n_{\text{eff}}^{(0)} \leq n^{(0)}$ and $n_{\text{eff}}^{(1)} \leq n^{(1)}$. Thus the maximum value of n_{eff_w} is $n/4$ when $n_{\text{eff}}^{(0)} = n^{(0)} = (1/2)n$ and $n_{\text{eff}}^{(1)} = n^{(1)} = (1/2)n$.

To show that $\frac{n^{(1)} n^{(0)}}{n^{(0)} + n^{(1)}} \geq n_{\text{eff}_w}$ for values of $n^{(1)}$ and $n^{(0)}$ other than $(1/2)n$, consider $n_{\text{eff}}^{(1)}$ fixed and write

$$f(n_{\text{eff}}^{(0)}) = \frac{n_{\text{eff}}^{(1)} n_{\text{eff}}^{(0)}}{n_{\text{eff}}^{(0)} + n_{\text{eff}}^{(1)}}.$$

After some simplification, the first derivative can be written as:

$$f'(n_{\text{eff}}^{(0)}) = \frac{n_{\text{eff}}^{(1)}}{n_{\text{eff}}^{(1)} + n_{\text{eff}}^{(0)}} - \frac{n_{\text{eff}}^{(1)}}{n_{\text{eff}}^{(1)} + n_{\text{eff}}^{(0)}} \times \frac{n_{\text{eff}}^{(0)}}{n_{\text{eff}}^{(1)} + n_{\text{eff}}^{(0)}}$$

This will be strictly greater than 0 as long as $n_{\text{eff}}^{(1)}$ and $n_{\text{eff}}^{(0)}$ are positive which is guaranteed by the assumption that $\sum_{i=1}^{n^{(1)}} w_i \neq 0$ and $\sum_{i=(n^{(1)}+1)}^{n} w_i \neq 0$. Thus, when $n_{\text{eff}}^{(1)}$ is fixed, the maximum value of n_{eff_w} is obtained by setting $n_{\text{eff}}^{(0)}$ equal to its maximum possible value of $n^{(0)}$. By symmetry, when $n_{\text{eff}}^{(0)}$ is fixed, the maximum value of n_{eff_w} is obtained by setting $n_{\text{eff}}^{(1)}$ equal to its maximum possible value of $n^{(1)}$. Thus, for given values of $n^{(0)}$ and $n^{(1)}$ we have $n_{\text{eff}_w} \leq \frac{n^{(1)}n^{(0)}}{n^{(0)}+n^{(1)}}$.

Finally, because $n_{\text{eff}}^{(1)}$ and $n_{\text{eff}}^{(0)}$ are positive (again, guaranteed by the assumption that $\sum_{i=1}^{n^{(1)}} w_i \neq 0$ and $\sum_{i=(n^{(1)}+1)}^{n} w_i \neq 0$), we know that $0 < n_{\text{eff}_w}$ □

E Calculating DFBETA for Hájek-Type Estimators

The following is based on Li and Valliant (2011).

Recall from the main body of the paper that the Hájek-type estimator of $\hat{\tau}_h$ in Equation A.1 is equivalent to the weighted least squares (WLS) estimator of τ_h in the following bivariate regression of the observed outcomes on the treatment indicator (Imbens, 2004):

$$Y_i = \beta_0 + Z_i \tau_h + \epsilon_i, \qquad i = 1, \ldots, n \tag{E.1}$$

with the ith regression weight equal to:

$$\omega_i = Z_i w_h^{(1)}(x_i) + (1 - Z_i) w_h^{(0)}(x_i).$$

Define $\mathbf{\Omega}$ to be the $n \times n$ diagonal matrix with

$$\omega_{ii} = \begin{cases} w_{hi}^{(1)} & \text{if } Z_i = 1 \\ w_{hi}^{(0)} & \text{if } Z_i = 0 \end{cases}$$

and let $\tilde{\mathbf{X}}$ be the $n \times 2$ matrix with 1s in the first column and Z in the second column. Then

$$\hat{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}'\mathbf{\Omega}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'\mathbf{\Omega}\mathbf{y}$$

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}$$

$$\mathbf{A} = \tilde{\mathbf{X}}'\mathbf{\Omega}\tilde{\mathbf{X}}$$

$$\text{DFBETA}_i = \frac{\mathbf{A}^{-1}\mathbf{x}_i \hat{\epsilon}_i \omega_{ii}}{1 - h_i}$$

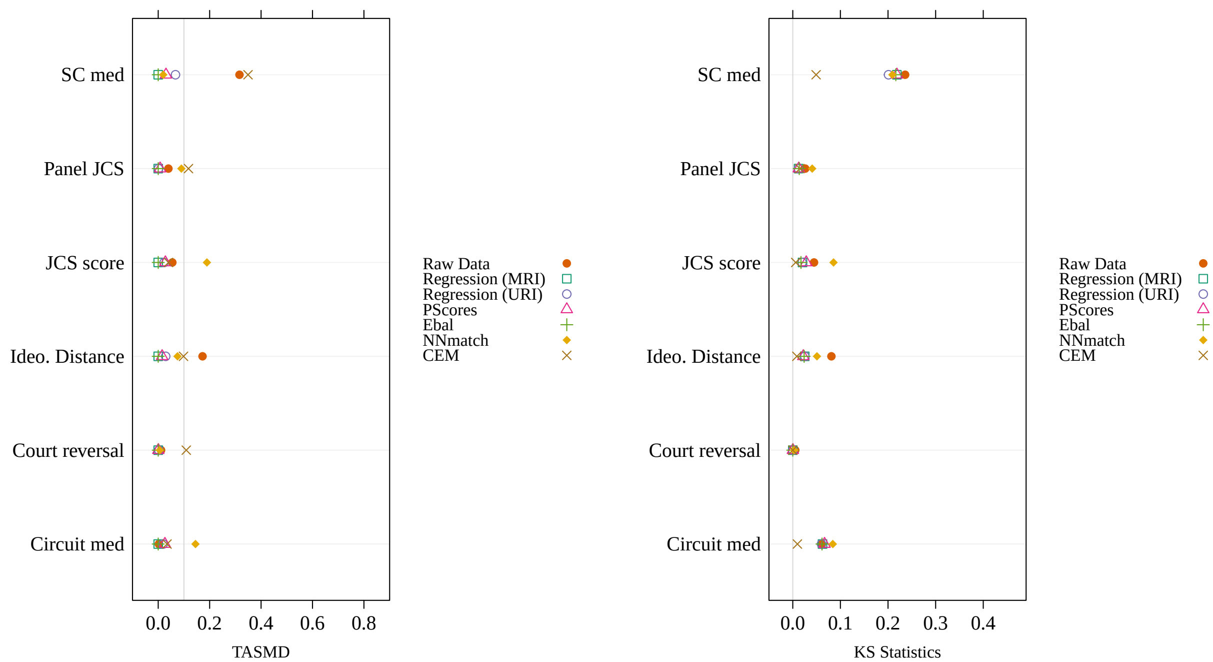
where

$$h_i = \omega_{ii}\left(\tilde{\mathbf{x}}_i' \mathbf{A}^{-1} \tilde{\mathbf{x}}_i\right)$$

and ω_{ii} is the ith diagonal element of $\mathbf{\Omega}$ (Li and Valliant, 2011).

F Application I: Promotion-Seeking Judges

In this section, we present additional results from our re-analysis of Black and Owens (2016). First, following Figure 1 in the main text, Figure A1 shows the diagnostic plots for non-contender judges.



(a) Non-contender Judges: TASMD

(b) Non-contender Judges: KS

Figure A1: *Weighted Mean Covariate Balance (Assessed by TASMD) and KS Test Statistic for Control Units Using Multiple ATT Estimation Methods in the Re-analysis of Black and Owens (2016) for Non-contender Judges.* In TASMD plots, each symbol shows the standardized difference between the weighted mean of the control and treated data for each covariate and method. The gray vertical line marks where the TASMD value is 0.1. In KS test statistic plots, each symbol shows the maximum absolute difference in the empirical cumulative distribution functions (ECDFs) of the control and treated groups, using both raw and weighted control data. The gray vertical line marks KS statistics at 0. `SC med` is the JCS score of the median Supreme Court justice; `Panel JCS` is the ideological distance between a judge and the remaining panelists; `JCS score` is each judge's JCS score; `Ideo. Distance` is the ideological distance between the judge and the president; `Court reversal` is whether the circuit court reversed the lower court; `Circuit med` is the JCS score of the median judge on the circuit.

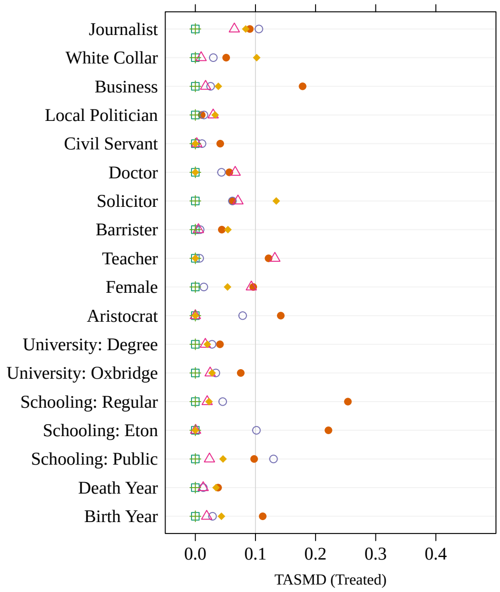
Second, Table A1 presents estimates of ATE and diagnostics from several estimation methos. In each case, the results do not support the authors' hypotheses about promotion seeking judges.

	ATE Estimate	SE	Effective Sample Size	Effective Sample Size Ratio	DFBETA Minimum Value	DFBETA Minimum Judge	DFBETA Maximum Value	DFBETA Maximum Judge	Extrapolation Control Units	Extrapolation Treatment Units
The Contender Judges										
Regression (MRI)	0.098	0.012	1527.0	0.609	-0.001	Cornelia G. Kennedy	0.000	Amalya Lyle Kearse	0.002	0.026
Regression (URI)	0.068	0.011	1823.6	0.727	-0.000	Cornelia G. Kennedy	0.000	Merrick Garland	0.027	0.000
Propensity Score Weighting	0.066	0.013	1252.8	0.500	-0.002	Cornelia G. Kennedy	0.002	Cornelia G. Kennedy		
Entropy Balancing	0.095	0.010	1356.0	0.541	-0.002	Harrie B. Chase	0.001	Amalya Lyle Kearse		
Nearest-Neighbor Matching	0.051	0.007	521.3	0.208	-0.007	Edith Clement	0.006	Cornelia G. Kennedy		
The Non-Contender Judges										
Regression (MRI)	0.029	0.009	3223.8	0.928	-0.000	NA	0.000	NA	0.000	0.000
Regression (URI)	0.040	0.009	3413.8	0.983	-0.000	NA	0.000	NA	0.000	0.000
Propensity Score Weighting	0.032	0.009	3114.6	0.897	-0.000	NA	0.000	NA		
Entropy Balancing	0.031	0.008	3217.2	0.926	-0.000	NA	0.000	NA		
Nearest-Neighbor Matching	0.026	0.004	2306.1	0.664	-0.001	NA	0.002	NA		

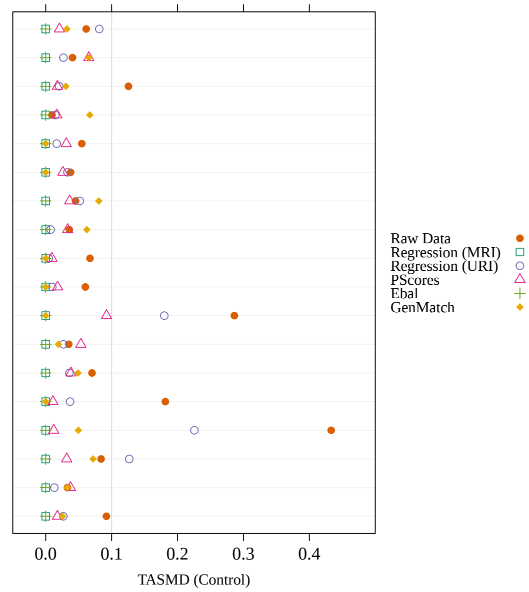
Table A1: *Re-analysis of Black and Owens (2016) with Multiple Approaches to Estimating ATE.* The *Regression (MRI)* rows correspond to the MRI estimator of ATE. The *Regression (URI)* rows correspond to the URI estimator. The *Propensity Score Weighting* rows correspond to the propensity score weighting estimator of ATE in which the propensity scores are estimated via logistic regression. The *Entropy Balancing* rows correspond to the entropy balancing estimator of ATE as implemented in the `WeightIt` `R` package. The *Nearest-Neighbor Matching* rows correspond to 1:1 nearest neighbor propensity score matching using the `Match` function in the `Matching` `R` package and the same estimated propensity scores as above. The minimum and maximum DFBETA values actually correspond to votes by particular judges on particular cases (the units in the study). The listed judge names are the names of the judges from those influential judge-case combinations. Judge names were not reported in the non-contender dataset.

G Application II: Wealth-Maximizing Politicians

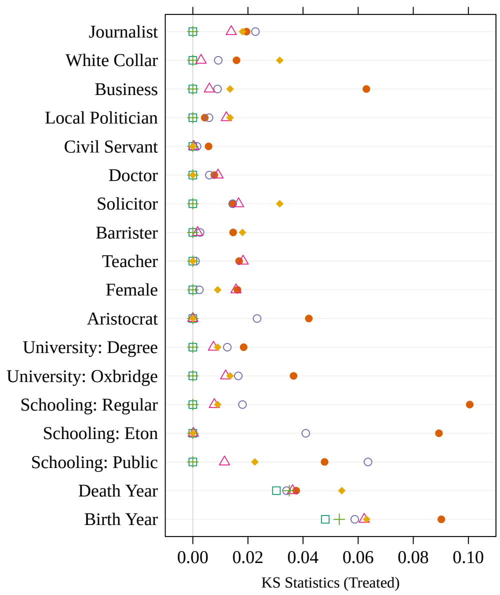
Following Figure 3 in the main text, Figure A2 shows the TASMD and KS statistics diagnostic plots for the treated and control units separately using multiple ATE estimation methods.



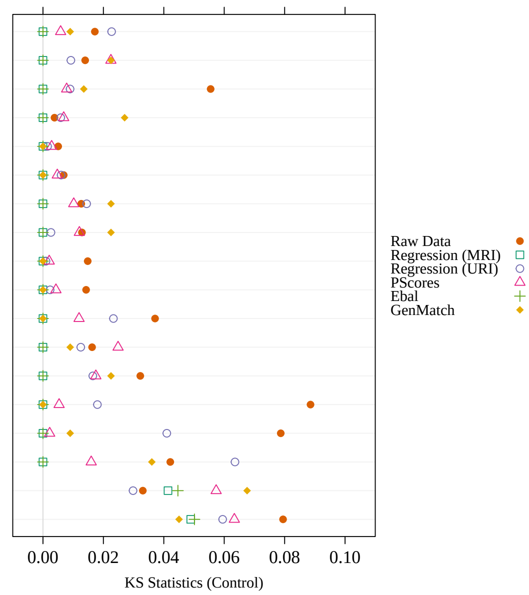
(a) Conservative Winners (Treated): TASMD



(b) Conservative Losers (Control): TASMD



(c) Conservative Winners (Treated): KS



(d) Conservative Losers (Control): KS

Figure A2: *Weighted Mean Covariate Balance (Assessed by TASMD) and KS Test Statistic All Units Using Multiple ATE Estimation Methods in the Re-analysis of Eggers and Hainmueller (2009).* In TASMD plots, each symbol represents the standardized difference between the weighted mean of the control units and the sample mean, and weighted mean of the treated units and the sample mean for each covariate and method. The gray vertical line marks where the TASMD value is 0.1. In KS test statistic plots, each symbol represents the maximum absolute difference in the empirical cumulative distribution functions (ECDFs) between the control and the sample mean, and between the treated group and the sample mean. The gray vertical line marks KS statistics at 0.

References

Abadie, A., and G. W. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74(1):235–267.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Black, R. C., and R. J. Owens. 2016. "Courting the President: How Circuit Court Judges Alter Their Behavior for Promotion to the Supreme Court." *American Journal of Political Science* 60(1):30–43.

Chattopadhyay, A., and J. R. Zubizarreta. 2023. "On the Implied Weights of Linear Regression for Causal Inference." *Biometrika* 110(3):615–629.

Eggers, A. C., and J. Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103(4):513–533.

Imbens, G. W. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *The Review of Economics and Statistics* 86:4–29.

Kish, L. 1965. *Survey Sampling*. New York: Wiley.

Li, F., K. L. Morgan, and A. M. Zaslavsky. 2018. "Balancing Covariates via Propensity Score Weighting." *Journal of the American Statistical Association* 113(521):390–400.

Li, J., and R. Valliant. 2011. "Linear Regression Influence Diagnostics for Unclustered Survey Data." *Journal of Official Statistics* 27(1):99–119.