Online Appendix for
"Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights"

Tymon Słoczyński

May 18, 2020

Appendices

Appendix A Proof of Theorem 1

First, consider equation (2) in the main text, $\mathrm{L}(y \mid 1, d, X) = \alpha + \tau d + X\beta$. By the Frisch–Waugh theorem, $\tau = \tau_a$, where τ_a is defined by

$$\mathrm{L}\left[y \mid 1, d, p(X)\right] = \alpha_a + \tau_a d + \gamma_a \cdot p(X). \tag{A1}$$

Second, note that (A1) is a linear projection of y on two variables: one binary, d, and one arbitrarily discrete or continuous, $p(X)$. Thus, we can use the following result from Elder et al. (2010).

Lemma A1 (Elder et al., 2010)**.** *Let* $\mathrm{L}(y \mid 1, d, x) = \alpha_e + \tau_e d + \beta_e x$ *denote the linear projection of* y *on* d *(a binary variable) and* x *(a single, possibly continuous variable). Then, assuming all objects are well defined,*

$$\begin{aligned}
\tau_e &= \frac{\rho \cdot \mathrm{V}(x \mid d=1)}{\rho \cdot \mathrm{V}(x \mid d=1) + (1-\rho) \cdot \mathrm{V}(x \mid d=0)} \cdot \theta_1 \\
&+ \frac{(1-\rho) \cdot \mathrm{V}(x \mid d=0)}{\rho \cdot \mathrm{V}(x \mid d=1) + (1-\rho) \cdot \mathrm{V}(x \mid d=0)} \cdot \theta_0,
\end{aligned}$$

where

$$\theta_1 = \frac{\mathrm{Cov}(d, y)}{\mathrm{V}(d)} - \frac{\mathrm{Cov}(d, x)}{\mathrm{V}(d)} \cdot \frac{\mathrm{Cov}(x, y \mid d=1)}{\mathrm{V}(x \mid d=1)}$$

and

$$\theta_0 = \frac{\mathrm{Cov}(d, y)}{\mathrm{V}(d)} - \frac{\mathrm{Cov}(d, x)}{\mathrm{V}(d)} \cdot \frac{\mathrm{Cov}(x, y \mid d=0)}{\mathrm{V}(x \mid d=0)}.$$

Combining the two pieces gives

$$\begin{aligned}
\tau &= \frac{\rho \cdot \mathrm{V}\left[p(X) \mid d=1\right]}{\rho \cdot \mathrm{V}\left[p(X) \mid d=1\right] + (1-\rho) \cdot \mathrm{V}\left[p(X) \mid d=0\right]} \cdot \theta_1^* \\
&+ \frac{(1-\rho) \cdot \mathrm{V}\left[p(X) \mid d=0\right]}{\rho \cdot \mathrm{V}\left[p(X) \mid d=1\right] + (1-\rho) \cdot \mathrm{V}\left[p(X) \mid d=0\right]} \cdot \theta_0^*,
\end{aligned} \tag{A2}$$

where

$$\theta_1^* = \frac{\mathrm{Cov}(d, y)}{\mathrm{V}(d)} - \frac{\mathrm{Cov}\left[d, p(X)\right]}{\mathrm{V}(d)} \cdot \frac{\mathrm{Cov}\left[p(X), y \mid d=1\right]}{\mathrm{V}\left[p(X) \mid d=1\right]} \tag{A3}$$

and

$$\theta_0^* = \frac{\mathrm{Cov}(d, y)}{\mathrm{V}(d)} - \frac{\mathrm{Cov}\left[d, p(X)\right]}{\mathrm{V}(d)} \cdot \frac{\mathrm{Cov}\left[p(X), y \mid d=0\right]}{\mathrm{V}\left[p(X) \mid d=0\right]}. \tag{A4}$$

Third, notice that $\theta_1^* = \tau_{APLE,0}$ and $\theta_0^* = \tau_{APLE,1}$, as defined in equation (8) in the main text. Indeed,

$$\frac{\mathrm{Cov}\,(d, y)}{\mathrm{V}\,(d)} = \mathrm{E}\,(y \mid d = 1) - \mathrm{E}\,(y \mid d = 0) \tag{A5}$$

and also

$$\frac{\mathrm{Cov}\,[d, p\,(X)]}{\mathrm{V}\,(d)} = \mathrm{E}\,[p\,(X) \mid d = 1] - \mathrm{E}\,[p\,(X) \mid d = 0]\,. \tag{A6}$$

Moreover, for $j = 0, 1$,

$$\frac{\mathrm{Cov}\,[p\,(X)\,, y \mid d = j]}{\mathrm{V}\,[p\,(X) \mid d = j]} = \gamma_j, \tag{A7}$$

where γ_1 and γ_0 are defined in equations (5) and (6) in the main text, respectively. Because

$$\begin{aligned} \mathrm{E}\,(y \mid d = 1) - \mathrm{E}\,(y \mid d = 0) &= \{\mathrm{E}\,[p\,(X) \mid d = 1] - \mathrm{E}\,[p\,(X) \mid d = 0]\} \cdot \gamma_1 \\ &+ (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \mathrm{E}\,[p\,(X) \mid d = 0] \end{aligned} \tag{A8}$$

and also

$$\begin{aligned} \mathrm{E}\,(y \mid d = 1) - \mathrm{E}\,(y \mid d = 0) &= \{\mathrm{E}\,[p\,(X) \mid d = 1] - \mathrm{E}\,[p\,(X) \mid d = 0]\} \cdot \gamma_0 \\ &+ (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \mathrm{E}\,[p\,(X) \mid d = 1]\,, \end{aligned} \tag{A9}$$

where again α_1 and α_0 are defined in equations (5) and (6) in the main text, we get the result that $\theta_1^* = \tau_{APLE,0}$ and $\theta_0^* = \tau_{APLE,1}$. Note that equations (A8) and (A9) correspond to special cases of the Oaxaca–Blinder decomposition (Blinder, 1973; Oaxaca, 1973; Fortin et al., 2011), which is also the focus of Elder et al. (2010). Finally, combining the three pieces gives

$$\begin{aligned} \tau &= \frac{\rho \cdot \mathrm{V}\,[p\,(X) \mid d = 1]}{\rho \cdot \mathrm{V}\,[p\,(X) \mid d = 1] + (1 - \rho) \cdot \mathrm{V}\,[p\,(X) \mid d = 0]} \cdot \tau_{APLE,0} \\ &+ \frac{(1 - \rho) \cdot \mathrm{V}\,[p\,(X) \mid d = 0]}{\rho \cdot \mathrm{V}\,[p\,(X) \mid d = 1] + (1 - \rho) \cdot \mathrm{V}\,[p\,(X) \mid d = 0]} \cdot \tau_{APLE,1}, \end{aligned} \tag{A10}$$

which completes the proof.

Appendix B Extensions

B1 Proportion of Treated Units and OLS Weights

To show formally that w_1 is decreasing in ρ and that w_0 is increasing in ρ, it is convenient to additionally assume that $\mathrm{E}(d \mid X)$ is linear in X.

Assumption B1.1. $\mathrm{E}(d \mid X) = p(X) = \alpha_p + X\beta_p$.

This restriction is satisfied automatically in saturated models, as studied by Angrist (1998) and Humphreys (2009). It is also used by Aronow and Samii (2016) and Abadie et al. (2020). In the present context there are two reasons why Assumption B1.1 is useful. First, it allows us to rewrite w_0 and w_1 solely in terms of unconditional expectations of $p(X)$ and its powers. Second, it simplifies the calculation of the derivatives of w_0 and w_1 with respect to the intercept of the propensity score model. Imposing a shift on this intercept is equivalent to changing ρ by a small amount. It turns out that Theorem 1 and Assumption B1.1 imply the following result.

Proposition B1.1. *Under Assumptions 1, 2, and B1.1,*

$$\frac{dw_1}{d\alpha_p} < 0 \qquad \textit{and} \qquad \frac{dw_0}{d\alpha_p} > 0.$$

Proof. For simplicity, we first focus on a_0 and a_1, which we define as $a_0 = \rho \cdot \mathrm{V}[p(X) \mid d = 1]$ and $a_1 = (1-\rho) \cdot \mathrm{V}[p(X) \mid d = 0]$. As a result, $w_0 = \frac{a_0}{a_0+a_1}$ and $w_1 = \frac{a_1}{a_0+a_1}$. It turns out that we can rewrite a_0 as

$$\begin{aligned}
a_0 &= \mathrm{E}(d) \cdot \mathrm{E}\left(\{p(X) - \mathrm{E}[p(X) \mid d = 1]\}^2 \mid d = 1\right) \\
&= \mathrm{E}(d) \cdot \left(\mathrm{E}\left[p(X)^2 \mid d = 1\right] - \{\mathrm{E}[p(X) \mid d = 1]\}^2\right) \\
&= \mathrm{E}(d) \cdot \left(\frac{\mathrm{E}\left[p(X)^2 d\right]}{\mathrm{E}(d)} - \left\{\frac{\mathrm{E}[p(X) d]}{\mathrm{E}(d)}\right\}^2\right) \\
&= \mathrm{E}\left[p(X)^2 d\right] - \frac{\{\mathrm{E}[p(X) d]\}^2}{\mathrm{E}(d)} \\
&= \mathrm{E}\left[p(X)^2 \mathrm{E}(d \mid X)\right] - \frac{\{\mathrm{E}[p(X) \mathrm{E}(d \mid X)]\}^2}{\mathrm{E}[\mathrm{E}(d \mid X)]} \\
&= \mathrm{E}\left[p(X)^3\right] - \frac{\left\{\mathrm{E}\left[p(X)^2\right]\right\}^2}{\mathrm{E}[p(X)]}.
\end{aligned} \tag{B1.1}$$

Then, taking the derivative of a_0 with respect to α_p gives

$$\begin{aligned}
\frac{da_0}{d\alpha_p} &= 3\mathrm{E}\left[p(X)^2\right] - \frac{4\mathrm{E}\left[p(X)^2\right]\mathrm{E}[p(X)]}{\mathrm{E}[p(X)]} + \frac{\left\{\mathrm{E}\left[p(X)^2\right]\right\}^2}{\mathrm{E}[p(X)]^2} \\
&= -\mathrm{E}\left[p(X)^2\right] + \frac{\left\{\mathrm{E}\left[p(X)^2\right]\right\}^2}{\mathrm{E}[p(X)]^2} \\
&= \frac{\left\{\mathrm{E}\left[p(X)^2\right]\right\}^2 - \mathrm{E}\left[p(X)^2\right]\mathrm{E}[p(X)]^2}{\mathrm{E}[p(X)]^2} \\
&= \frac{\mathrm{E}\left[p(X)^2\right]\left\{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}[p(X)]^2\right\}}{\mathrm{E}[p(X)]^2} \\
&= \frac{\mathrm{E}\left[p(X)^2\right]\mathrm{V}[p(X)]}{\mathrm{E}[p(X)]^2} > 0. \qquad \text{(B1.2)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
a_1 &= [1-\mathrm{E}(d)]\cdot\mathrm{E}\left(\{p(X)-\mathrm{E}[p(X)\mid d=0]\}^2 \mid d=0\right) \\
&= [1-\mathrm{E}(d)]\cdot\left(\mathrm{E}\left[p(X)^2 \mid d=0\right] - \{\mathrm{E}[p(X)\mid d=0]\}^2\right) \\
&= [1-\mathrm{E}(d)]\cdot\left(\frac{\mathrm{E}\left[p(X)^2\right]-\mathrm{E}\left[p(X)^2 d\right]}{1-\mathrm{E}(d)} - \left\{\frac{\mathrm{E}[p(X)]-\mathrm{E}[p(X)d]}{1-\mathrm{E}(d)}\right\}^2\right) \\
&= \mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^2 d\right] - \frac{\{\mathrm{E}[p(X)]-\mathrm{E}[p(X)d]\}^2}{1-\mathrm{E}(d)} \\
&= \mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^2\,\mathrm{E}(d\mid X)\right] - \frac{\{\mathrm{E}[p(X)]-\mathrm{E}[p(X)\,\mathrm{E}(d\mid X)]\}^2}{1-\mathrm{E}[\mathrm{E}(d\mid X)]} \\
&= \mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right] - \frac{\left\{\mathrm{E}[p(X)]-\mathrm{E}\left[p(X)^2\right]\right\}^2}{1-\mathrm{E}[p(X)]} \qquad \text{(B1.3)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{da_1}{d\alpha_p} &= 2\mathrm{E}[p(X)] - 3\mathrm{E}\left[p(X)^2\right] - \frac{\left\{\mathrm{E}[p(X)]-\mathrm{E}\left[p(X)^2\right]\right\}^2}{\{1-\mathrm{E}[p(X)]\}^2} \\
&- \frac{2\cdot\{1-\mathrm{E}[p(X)]\}\cdot\{1-2\mathrm{E}[p(X)]\}\cdot\left\{\mathrm{E}[p(X)]-\mathrm{E}\left[p(X)^2\right]\right\}}{\{1-\mathrm{E}[p(X)]\}^2} \\
&= \frac{\mathrm{E}[p(X)]^2-\mathrm{E}\left[p(X)^2\right]}{\{1-\mathrm{E}[p(X)]\}^2}
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{2\mathrm{E}[p(X)]\,\mathrm{E}\left[p(X)^2\right] - 2\mathrm{E}[p(X)]^3}{\{1-\mathrm{E}[p(X)]\}^2} \\
&+ \quad \frac{\mathrm{E}\left[p(X)^2\right]\mathrm{E}[p(X)]^2 - \left\{\mathrm{E}\left[p(X)^2\right]\right\}^2}{\{1-\mathrm{E}[p(X)]\}^2} \\
&= \quad \frac{-\mathrm{V}[p(X)]\cdot\left\{1-2\mathrm{E}[p(X)]+\mathrm{E}\left[p(X)^2\right]\right\}}{\{1-\mathrm{E}[p(X)]\}^2} \\
&= \quad \frac{-\mathrm{V}[p(X)]\cdot\mathrm{E}\left\{[1-p(X)]^2\right\}}{\{1-\mathrm{E}[p(X)]\}^2} < 0. \qquad \text{(B1.4)}
\end{aligned}
$$

Finally, it follows that

$$
\frac{dw_1}{d\alpha_p} < 0 \quad \text{and} \quad \frac{dw_0}{d\alpha_p} > 0, \qquad \text{(B1.5)}
$$

since $w_0 = \frac{a_0}{a_0+a_1}$, $w_1 = \frac{a_1}{a_0+a_1}$, $a_0 > 0$, $a_1 > 0$, $\frac{da_0}{d\alpha_p} > 0$, and $\frac{da_1}{d\alpha_p} < 0$. □

B2 Further Intuition for Theorem 1

We begin by noting that because the linear projection passes through the point of means of all variables, which implies, for example, that $\mathrm{E}(y \mid d=1) = \alpha_1 + \gamma_1 \cdot \mathrm{E}[p(X) \mid d=1]$ and $\mathrm{E}(y \mid d=0) = \alpha_0 + \gamma_0 \cdot \mathrm{E}[p(X) \mid d=0]$, the average partial linear effects of d on both groups of interest can also be expressed as

$$\tau_{APLE,1} = \mathrm{E}(y \mid d=1) - \{\alpha_0 + \gamma_0 \cdot \mathrm{E}[p(X) \mid d=1]\} \tag{B2.1}$$

and

$$\tau_{APLE,0} = \{\alpha_1 + \gamma_1 \cdot \mathrm{E}[p(X) \mid d=0]\} - \mathrm{E}(y \mid d=0). \tag{B2.2}$$

In other words, we only need the linear projection of y on $p(X)$ in group zero, and not in group one, to define $\tau_{APLE,1}$. Similarly, we need the linear projection of y on $p(X)$ in group one, but not in group zero, to define $\tau_{APLE,0}$. When all objects are well defined, $\tau_{APLE,j}$ is also equivalent to the coefficient on d in the linear projection of y on d, $p(X)$, and $d \cdot \{p(X) - \mathrm{E}[p(X) \mid d=j]\}$.

Then, an alternative intuition for the OLS weights in Theorem 1 follows from partial residualization that is implicit in least squares estimation. The first thing to note is that τ, the OLS estimand, is equal to the coefficient on d in the linear projection of $y - \gamma_a \cdot p(X)$ on d, where γ_a is defined in equation (A1). An implication of Deaton (1997) and Solon et al. (2015) is that γ_a is also a convex combination of γ_1 and γ_0, where the weight on γ_1 is *increasing* in ρ. It follows that τ is a weighted average as well; it combines the coefficients on d in the linear projections of $y - \gamma_1 \cdot p(X)$ and $y - \gamma_0 \cdot p(X)$ on d in group zero and one, respectively. While the weight on the former (latter) is increasing (decreasing) in ρ, this parameter corresponds to $\tau_{APLE,0}$ ($\tau_{APLE,1}$), as can be seen from equations (B2.1) and (B2.2). Indeed, as noted above, it is γ_1 (and not γ_0) that is necessary to define $\tau_{APLE,0}$. The bottom line is that when there are more treated than untreated units, γ_1 is likely to be better estimated than γ_0 and OLS gives more weight to the contrast of $y - \gamma_1 \cdot p(X)$, which in turn corresponds to $\tau_{APLE,0}$. Interestingly, this parallels the intuition in Angrist (1998) and Angrist and Pischke (2009) that OLS gives more weight to treatment effects that are better estimated in finite samples. Also, this discussion leads to an alternative proof of Theorem 1.

Proof. As in online appendix A, consider equation (2) in the main text, $\mathrm{L}(y \mid 1, d, X) = \alpha + \tau d + X\beta$, and note that $\tau = \tau_a$, where τ_a is defined by $\mathrm{L}[y \mid 1, d, p(X)] = \alpha_a + \tau_a d + \gamma_a \cdot p(X)$. We can write this linear projection in error form as

$$y = \alpha_a + \tau_a d + \gamma_a \cdot p(X) + \upsilon. \tag{B2.3}$$

As in the main text, we also consider separate linear projections for $d=1$ and $d=0$, namely

$$\mathrm{L}[y \mid 1, p(X), d=1] = \alpha_1 + \gamma_1 \cdot p(X) \tag{B2.4}$$

and

$$\mathrm{L}\left[y \mid 1, p\left(X\right), d=0\right] = \alpha_0 + \gamma_0 \cdot p\left(X\right). \tag{B2.5}$$

Henceforth, to simplify notation I will use $l_1(X)$ to denote $\alpha_1 + \gamma_1 \cdot p\left(X\right)$ and $l_0(X)$ to denote $\alpha_0 + \gamma_0 \cdot p\left(X\right)$. To understand the relationship between γ_a, γ_1, and γ_0, we can use the following result from Deaton (1997) and Solon et al. (2015).

Lemma B2.1 (Deaton, 1997; Solon et al., 2015)**.** *Let* $\mathrm{L}\left(y \mid 1, d, x\right) = \alpha_e + \tau_e d + \beta_e x$ *denote the linear projection of y on d (a binary variable) and x (a single, possibly continuous variable). Then, assuming all objects are well defined,*

$$\begin{aligned} \beta_e &= \frac{\rho \cdot \mathrm{V}\left(x \mid d=1\right)}{\rho \cdot \mathrm{V}\left(x \mid d=1\right) + \left(1-\rho\right) \cdot \mathrm{V}\left(x \mid d=0\right)} \cdot \beta_{1,e} \\ &+ \frac{\left(1-\rho\right) \cdot \mathrm{V}\left(x \mid d=0\right)}{\rho \cdot \mathrm{V}\left(x \mid d=1\right) + \left(1-\rho\right) \cdot \mathrm{V}\left(x \mid d=0\right)} \cdot \beta_{0,e}, \end{aligned}$$

where $\beta_{1,e}$ *and* $\beta_{0,e}$ *are defined by*

$$\mathrm{L}\left(y \mid 1, x, d=1\right) = \alpha_{1,e} + \beta_{1,e} x$$

and

$$\mathrm{L}\left(y \mid 1, x, d=0\right) = \alpha_{0,e} + \beta_{0,e} x.$$

An implication of Lemma B2.1 is that

$$\gamma_a = w_0 \cdot \gamma_1 + w_1 \cdot \gamma_0. \tag{B2.6}$$

Next, we can rewrite equation (B2.3) as

$$\begin{aligned} y - w_0 \cdot \gamma_1 \cdot p\left(X\right) - w_1 \cdot \gamma_0 \cdot p\left(X\right) &= \alpha_a + \tau_a d + \upsilon \\ &= \mathrm{E}\left(y\right) - \tau_a \cdot \mathrm{E}\left(d\right) - \gamma_a \cdot \mathrm{E}\left[p\left(X\right)\right] + \tau_a d + \upsilon. \end{aligned} \tag{B2.7}$$

Moreover, it turns out that

$$\alpha_1 = \mathrm{E}\left(y \mid d=1\right) - \gamma_1 \cdot \mathrm{E}\left[p\left(X\right) \mid d=1\right] \tag{B2.8}$$

and also

$$\alpha_0 = \mathrm{E}\left(y \mid d=0\right) - \gamma_0 \cdot \mathrm{E}\left[p\left(X\right) \mid d=0\right]. \tag{B2.9}$$

It follows that

$$
\begin{aligned}
y - w_0 \cdot l_1(X) - w_1 \cdot l_0(X) &= \mathrm{E}\,(y) - w_0 \cdot \mathrm{E}\,(y \mid d = 1) - w_1 \cdot \mathrm{E}\,(y \mid d = 0) \\
&+ w_0 \cdot \gamma_1 \cdot \{\mathrm{E}\,[p\,(X) \mid d = 1] - \mathrm{E}\,[p\,(X)]\} \\
&+ w_1 \cdot \gamma_0 \cdot \{\mathrm{E}\,[p\,(X) \mid d = 0] - \mathrm{E}\,[p\,(X)]\} \\
&- \tau_a \cdot \mathrm{E}\,(d) + \tau_a d + \upsilon. \qquad \text{(B2.10)}
\end{aligned}
$$

In other words, in a linear projection of $y - w_0 \cdot l_1(X) - w_1 \cdot l_0(X)$ on d, the coefficient on d is equal to τ_a and the intercept is equal to $\mathrm{E}\,(y) - w_0 \cdot \mathrm{E}\,(y \mid d = 1) - w_1 \cdot \mathrm{E}\,(y \mid d = 0) + w_0 \cdot \gamma_1 \cdot \{\mathrm{E}\,[p\,(X) \mid d = 1] - \mathrm{E}\,[p\,(X)]\} + w_1 \cdot \gamma_0 \cdot \{\mathrm{E}\,[p\,(X) \mid d = 0] - \mathrm{E}\,[p\,(X)]\} - \tau_a \cdot \mathrm{E}\,(d)$. However, τ_a must also be equal to the difference in expected values of the dependent variable for $d = 1$ and $d = 0$. Using equations (B2.1) and (B2.2), we can write these expected values as

$$
\mathrm{E}\,[y - w_0 \cdot l_1(X) - w_1 \cdot l_0(X) \mid d = 1] = w_1 \cdot \tau_{APLE,1} \qquad \text{(B2.11)}
$$

and

$$
\mathrm{E}\,[y - w_0 \cdot l_1(X) - w_1 \cdot l_0(X) \mid d = 0] = -w_0 \cdot \tau_{APLE,0}. \qquad \text{(B2.12)}
$$

Thus,

$$
\tau = \tau_a = w_1 \cdot \tau_{APLE,1} + w_0 \cdot \tau_{APLE,0}, \qquad \text{(B2.13)}
$$

which completes the proof. □

B3 A Weighted Least Squares Correction

Suppose we use weighted least squares (WLS) to estimate the model with d and $p(X)$ as the only independent variables. In this case we would like to obtain a set of weights, w, such that τ_w in

$$\mathrm{L}\left(\sqrt{w} \cdot y \mid \sqrt{w}, \sqrt{w} \cdot d, \sqrt{w} \cdot p(X)\right) = \alpha_w \sqrt{w} + \tau_w \sqrt{w} \cdot d + \gamma_w \sqrt{w} \cdot p(X) \tag{B3.1}$$

has a useful interpretation. An appropriate set of weights is provided in Proposition B3.1.

Proposition B3.1 (Weighted Least Squares Correction)**.** *Suppose that Assumptions 1 and 2 are satisfied. Also, $w = \frac{1-\rho}{w_0} \cdot d + \frac{\rho}{w_1} \cdot (1-d)$. Then,*

$$\tau_w = \tau_{APLE}.$$

Suppose that Assumptions 1, 2, 3, and 4 are satisfied. Also, $w = \frac{1-\rho}{w_0} \cdot d + \frac{\rho}{w_1} \cdot (1-d)$. Then,

$$\tau_w = \tau_{ATE}.$$

The proof of Proposition B3.1 follows directly from the proofs of Theorem 1 and Corollary 1, and is omitted. Proposition B3.1 establishes that the average effect of d can be recovered from a weighted least squares procedure, with weights of $\frac{1-\rho}{w_0}$ for units with $d = 1$ and weights of $\frac{\rho}{w_1}$ for units with $d = 0$. These weights consist of two parts: either $\frac{1}{w_1}$ or $\frac{1}{w_0}$; and either ρ or $1-\rho$. The role of the first part is always to undo the OLS weights (w_1 and w_0 in Theorem 1); the role of the second part is to impose the correct weights of ρ on the average effect of d on group one and $1-\rho$ on the average effect of d on group zero. Finally, it is useful to note that there is no similar procedure to recover the average effects of d on group zero and one; both of these objects, however, are easily obtained from equation (8) in the main text.

Interestingly, the structure of the weighted least squares procedure in Proposition B3.1 resembles the "tyranny of the minority" estimator in Lin (2013). This method uses weights of $\frac{1-\rho}{\rho}$ for units with $d = 1$ and weights of $\frac{\rho}{1-\rho}$ for units with $d = 0$; it also controls for X instead of $p(X)$. It is important to note, however, that this method is designed to solve a different problem than Proposition B3.1. In particular, Freedman (2008b,a) demonstrates that regression adjustments to experimental data can lead to a loss in precision. On the other hand, Lin (2013) shows that this is no longer possible if we additionally interact d with X (see also Negi and Wooldridge, 2019). Then, Lin (2013) derives the "tyranny of the minority" estimator as an alternative procedure, based on a single conditional mean, which does not suffer from this loss in precision. In the context of observational data, however, the weights in Lin (2013) guarantee that $\tau_w = \tau_{APLE}$ only in a special case, namely under Assumption 5, $\mathrm{V}[p(X) \mid d = 1] = \mathrm{V}[p(X) \mid d = 0]$.

B4 Comparison with Angrist (1998) and Aronow and Samii (2016)

The result in Angrist (1998) states that if $\mathrm{L}(y \mid d, X) = \tau_n d + \sum_{s=1}^{S} \beta_{n,s} x_s$, where $X = (x_1, \ldots, x_S)$ is a vector of exhaustive and mutually exclusive "stratum" indicators, then

$$\tau_n = \sum_{s=1}^{S} \frac{\mathrm{P}(x_s = 1) \cdot \mathrm{P}(d = 1 \mid x_s = 1) \cdot \mathrm{P}(d = 0 \mid x_s = 1)}{\sum_{t=1}^{S} \mathrm{P}(x_t = 1) \cdot \mathrm{P}(d = 1 \mid x_t = 1) \cdot \mathrm{P}(d = 0 \mid x_t = 1)} \cdot \tau_s, \tag{B4.1}$$

where $\tau_s = \mathrm{E}(y \mid d = 1, x_s = 1) - \mathrm{E}(y \mid d = 0, x_s = 1)$. Further, under standard assumptions, $\tau_s = \mathrm{E}[y(1) - y(0) \mid X]$. In this appendix I show that equation (B4.1) follows from Corollary 1 when the model for y is saturated.

The starting point is to note that, because the model for y is saturated, $\mathrm{E}(d \mid X) = p(X) = \sum_{s=1}^{S} \beta_{p,s} x_s$. Additionally, Assumptions 3 and 4 allow us to write $\mathrm{E}[y(1) - y(0) \mid X] = (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot p(X)$. It follows that equation (B4.1) can alternatively be expressed as

$$\begin{aligned} \tau_n &= \frac{\mathrm{E}\{p(X) \cdot [1 - p(X)] \cdot [(\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot p(X)]\}}{\mathrm{E}\{p(X) \cdot [1 - p(X)]\}} \\ &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right]}{\mathrm{E}[p(X)] - \mathrm{E}\left[p(X)^2\right]}. \end{aligned} \tag{B4.2}$$

The same representation of the OLS estimand under Assumptions 3 and 4 follows from Aronow and Samii (2016), who generalize the result in Angrist (1998) to any model, saturated or not, where $\mathrm{E}(d \mid X)$ is linear in X.

To demonstrate that the results in Angrist (1998) and Aronow and Samii (2016) follow from Corollary 1, we need to show that equation (B4.2) can be obtained by rearranging the expression in Corollary 1. To see this note that, under Assumptions 3 and 4, $\tau_{ATT} = (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \mathrm{E}[p(X) \mid d = 1]$ and $\tau_{ATU} = (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \mathrm{E}[p(X) \mid d = 0]$. Upon rearrangement,

$$\begin{aligned} \tau_{ATT} &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\mathrm{E}[p(X)\, d]}{\mathrm{E}(d)} \\ &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\mathrm{E}\left[p(X)^2\right]}{\mathrm{E}[p(X)]} \end{aligned} \tag{B4.3}$$

and

$$\begin{aligned} \tau_{ATU} &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\mathrm{E}[p(X) \cdot (1 - d)]}{1 - \mathrm{E}(d)} \\ &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\mathrm{E}[p(X)] - \mathrm{E}\left[p(X)^2\right]}{1 - \mathrm{E}[p(X)]}. \end{aligned} \tag{B4.4}$$

Also, because $\mathrm{E}(d \mid X)$ is linear in X and hence equal to $p(X)$, we can use the results from online appendix B1, which state that $\rho \cdot \mathrm{V}[p(X) \mid d=1] = \mathrm{E}\left[p(X)^3\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]}$ and $(1-\rho) \cdot \mathrm{V}[p(X) \mid d=0] = \mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right] - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}$. It follows that

$$\begin{aligned} w_0 &= \frac{\rho \cdot \mathrm{V}[p(X) \mid d=1]}{\rho \cdot \mathrm{V}[p(X) \mid d=1] + (1-\rho) \cdot \mathrm{V}[p(X) \mid d=0]} \\ &= \frac{\mathrm{E}\left[p(X)^3\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]}}{\mathrm{E}\left[p(X)^2\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]} - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}} \end{aligned} \tag{B4.5}$$

and

$$\begin{aligned} w_1 &= \frac{(1-\rho) \cdot \mathrm{V}[p(X) \mid d=0]}{\rho \cdot \mathrm{V}[p(X) \mid d=1] + (1-\rho) \cdot \mathrm{V}[p(X) \mid d=0]} \\ &= \frac{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right] - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}}{\mathrm{E}\left[p(X)^2\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]} - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}}. \end{aligned} \tag{B4.6}$$

Consequently, an implication of Corollary 1 is that

$$\begin{aligned} \tau_n &= w_1 \cdot \tau_{ATT} + w_0 \cdot \tau_{ATU} \\ &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\left\{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right] - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}\right\} \cdot \mathrm{E}\left[p(X)^2\right]}{\left\{\mathrm{E}\left[p(X)^2\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]} - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}\right\} \cdot \mathrm{E}[p(X)]} \\ &+ (\gamma_1 - \gamma_0) \cdot \frac{\left\{\mathrm{E}\left[p(X)^3\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]}\right\} \cdot \left\{\mathrm{E}[p(X)] - \mathrm{E}\left[p(X)^2\right]\right\}}{\left\{\mathrm{E}\left[p(X)^2\right] - \frac{\{\mathrm{E}[p(X)^2]\}^2}{\mathrm{E}[p(X)]} - \frac{\{\mathrm{E}[p(X)] - \mathrm{E}[p(X)^2]\}^2}{1-\mathrm{E}[p(X)]}\right\} \cdot \{1 - \mathrm{E}[p(X)]\}} \end{aligned} \tag{B4.7}$$

or, equivalently,

$$\tau_n = (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\lambda_n}{\lambda_d}, \tag{B4.8}$$

where

$$\lambda_d = \left\{\mathrm{E}\left[p(X)^2\right] - \frac{\left\{\mathrm{E}\left[p(X)^2\right]\right\}^2}{\mathrm{E}[p(X)]} - \frac{\left\{\mathrm{E}[p(X)] - \mathrm{E}\left[p(X)^2\right]\right\}^2}{1-\mathrm{E}[p(X)]}\right\} \cdot \mathrm{E}[p(X)] \cdot \{1 - \mathrm{E}[p(X)]\} \tag{B4.9}$$

and

$$
\begin{aligned}
\lambda_n &= \left\{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right] - \frac{\left\{\mathrm{E}\left[p(X)\right] - \mathrm{E}\left[p(X)^2\right]\right\}^2}{1 - \mathrm{E}\left[p(X)\right]}\right\} \cdot \mathrm{E}\left[p(X)^2\right] \cdot \{1 - \mathrm{E}\left[p(X)\right]\} \\
&+ \left\{\mathrm{E}\left[p(X)^3\right] - \frac{\left\{\mathrm{E}\left[p(X)^2\right]\right\}^2}{\mathrm{E}\left[p(X)\right]}\right\} \cdot \left\{\mathrm{E}\left[p(X)\right] - \mathrm{E}\left[p(X)^2\right]\right\} \cdot \mathrm{E}\left[p(X)\right]. \qquad \text{(B4.10)}
\end{aligned}
$$

Upon further rearrangement,

$$
\begin{aligned}
\lambda_d &= \mathrm{E}\left[p(X)^2\right] \cdot \mathrm{E}\left[p(X)\right] + \mathrm{E}\left[p(X)^2\right] \cdot \{\mathrm{E}\left[p(X)\right]\}^2 - \left\{\mathrm{E}\left[p(X)^2\right]\right\}^2 - \{\mathrm{E}\left[p(X)\right]\}^3 \\
&= \left\{\mathrm{E}\left[p(X)\right] - \mathrm{E}\left[p(X)^2\right]\right\} \cdot \left\{\mathrm{E}\left[p(X)^2\right] - \{\mathrm{E}\left[p(X)\right]\}^2\right\} \qquad \text{(B4.11)}
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda_n &= \left\{\mathrm{E}\left[p(X)^2\right]\right\}^2 + \mathrm{E}\left[p(X)^3\right] \cdot \{\mathrm{E}\left[p(X)\right]\}^2 - \mathrm{E}\left[p(X)^3\right] \cdot \mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^2\right] \cdot \{\mathrm{E}\left[p(X)\right]\}^2 \\
&= \left\{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right]\right\} \cdot \left\{\mathrm{E}\left[p(X)^2\right] - \{\mathrm{E}\left[p(X)\right]\}^2\right\}. \qquad \text{(B4.12)}
\end{aligned}
$$

Finally, plugging equations (B4.11) and (B4.12) into equation (B4.8) gives

$$
\begin{aligned}
\tau_n &= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\left\{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right]\right\} \cdot \left\{\mathrm{E}\left[p(X)^2\right] - \{\mathrm{E}\left[p(X)\right]\}^2\right\}}{\left\{\mathrm{E}\left[p(X)\right] - \mathrm{E}\left[p(X)^2\right]\right\} \cdot \left\{\mathrm{E}\left[p(X)^2\right] - \{\mathrm{E}\left[p(X)\right]\}^2\right\}} \\
&= (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0) \cdot \frac{\mathrm{E}\left[p(X)^2\right] - \mathrm{E}\left[p(X)^3\right]}{\mathrm{E}\left[p(X)\right] - \mathrm{E}\left[p(X)^2\right]}. \qquad \text{(B4.13)}
\end{aligned}
$$

The equivalence between equations (B4.2) and (B4.13) confirms that the result in Angrist (1998) follows from Corollary 1 when the model for y is saturated. Similarly, the result in Aronow and Samii (2016) follows from Corollary 1 when $\mathrm{E}(d \mid X)$ is linear in X.

Appendix C Implementation in Stata

This appendix discusses the implementation of my theoretical results using the Stata package `hettreatreg`. In particular, I show how to apply this package to obtain the estimates in column 4 of Table 1 in the main text. To download this package and the NSW–CPS data from SSC, type

```
. ssc install hettreatreg, all
```

in the Command window. Then, type

```
. use nswcps, clear
```

to open the NSW–CPS data set. Then, the standard way to obtain the OLS estimate in column 4 of Table 1 in the main text would be to type

```
. regress re78 treated age-re75, vce(robust)

Linear regression                               Number of obs     =     16,177
                                                F(10, 16166)      =    1718.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4762
                                                Root MSE          =     7001.7

------------------------------------------------------------------------------
             |               Robust
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |    793.587   618.6092     1.28   0.200    -418.9555     2006.13
         age |  -233.6775    40.7162    -5.74   0.000    -313.4857   -153.8692
        age2 |   1.814371   .5581946     3.25   0.001     .7202474    2.908494
        educ |   166.8492   28.70683     5.81   0.000     110.5807    223.1178
       black |  -790.6086   197.8149    -4.00   0.000    -1178.348   -402.8694
    hispanic |  -175.9751   218.3033    -0.81   0.420    -603.8738    251.9235
     married |    224.266   152.4363     1.47   0.141    -74.52594    523.0579
    nodegree |   311.8445    176.414     1.77   0.077     -33.9464    657.6355
        re74 |   .2953363   .0152084    19.42   0.000     .2655261    .3251466
        re75 |   .4706353   .0153101    30.74   0.000     .4406259    .5006447
       _cons |   7634.344   737.8143    10.35   0.000     6188.146    9080.542
------------------------------------------------------------------------------
```

It is also possible, however, to obtain the same output and several additional estimates—including those of my diagnostics and those of implicit estimates of ATE, ATT, and ATU—by typing

```
. hettreatreg age-re75, o(re78) t(treated) noisily vce(robust)

Linear regression                               Number of obs     =     16,177
                                                F(10, 16166)      =    1718.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4762
                                                Root MSE          =     7001.7

------------------------------------------------------------------------------
             |               Robust
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |    793.587   618.6092     1.28   0.200    -418.9555     2006.13
         age |  -233.6775    40.7162    -5.74   0.000    -313.4857   -153.8692
        age2 |   1.814371   .5581946     3.25   0.001     .7202474    2.908494
        educ |   166.8492   28.70683     5.81   0.000     110.5807    223.1178
       black |  -790.6086   197.8149    -4.00   0.000    -1178.348   -402.8694
    hispanic |  -175.9751   218.3033    -0.81   0.420    -603.8738    251.9235
     married |    224.266   152.4363     1.47   0.141    -74.52594    523.0579
    nodegree |   311.8445    176.414     1.77   0.077     -33.9464    657.6355
        re74 |   .2953363   .0152084    19.42   0.000     .2655261    .3251466
        re75 |   .4706353   .0153101    30.74   0.000     .4406259    .5006447
       _cons |   7634.344   737.8143    10.35   0.000     6188.146    9080.542
------------------------------------------------------------------------------

"OLS" is the estimated regression coefficient on treated.

   OLS  =  793.6

P(d=1)  =  .011
P(d=0)  =  .989

    w1  =  .983
    w0  =  .017
 delta  =  -.971

   ATE  =  -6751
   ATT  =  928.4
```

```
   ATU  =  -6840

OLS = w1*ATT + w0*ATU = 793.6
```

Alternatively, we may restrict our attention to this additional output by typing

```
. hettreatreg age-re75, o(re78) t(treated)

"OLS" is the estimated regression coefficient on treated.

   OLS  =  793.6

P(d=1)  =  .011
P(d=0)  =  .989

    w1  =  .983
    w0  =  .017
 delta  =  -.971

   ATE  =  -6751
   ATT  =  928.4
   ATU  =  -6840

OLS = w1*ATT + w0*ATU = 793.6
```

In any case, `OLS` is the estimated regression coefficient on the variable designated as treatment. `P(d=1)` and `P(d=0)` correspond to $\hat{\rho}$ and $1 - \hat{\rho}$, respectively. `w1`, `w0`, and `delta` correspond to $\hat{w}_1$, $\hat{w}_0$, and $\hat{\delta}$, respectively. Finally, `ATE`, `ATT`, and `ATU` correspond to $\hat{\tau}_{APLE}$, $\hat{\tau}_{APLE,1}$, and $\hat{\tau}_{APLE,0}$, respectively. `hettreatreg` stores all these estimates in `e()`. Type

```
. help hettreatreg
```

for more information and additional examples.

Appendix D Implementation in R

Similar to online appendix C, I now discuss the implementation of my theoretical results using the R package `hettreatreg`. As before, I show how to obtain the estimates reported in column 4 of Table 1 in the main text. To download this package and the NSW–CPS data from CRAN, type

```
> install.packages("hettreatreg")
```

in the R/R Studio console. Next, type

```
> library(hettreatreg)
```

to load `hettreatreg` and

```
> data("nswcps")
```

to open the NSW–CPS data set. Then, the standard way to obtain the OLS estimate in column 4 of Table 1 in the main text would be to type

```
> lm(re78 ~ treated + age + age2 + educ + black + hispanic + married + nodegree
 + re74 + re75, data = nswcps)

Call:
lm(formula = re78 ~ treated + age + age2 + educ + black + hispanic +
    married + nodegree + re74 + re75, data = nswcps)

Coefficients:
(Intercept)      treated          age         age2         educ        black
  7634.3441     793.5870    -233.6775       1.8144     166.8492    -790.6086
   hispanic      married     nodegree         re74         re75
  -175.9751     224.2660     311.8445       0.2953       0.4706
```

Using `hettreatreg`, it is possible to obtain several additional estimates, including those of my diagnostics and those of implicit estimates of ATE, ATT, and ATU. Before doing so, it is useful to designate an outcome variable, a treatment variable, and a list of control variables. To do this, type

```
> outcome <- nswcps$re78
> treated <- nswcps$treated
> our_vars <- c("age", "age2", "educ", "black", "hispanic", "married", "nodegree",
 "re74", "re75")
> covariates <- subset(nswcps, select = our_vars)
```

Then, type

```
> hettreatreg(outcome, treated, covariates, verbose = TRUE)

"OLS" is the estimated regression coefficient on treated.

   OLS = 793.6

P(d=1) = 0.011
P(d=0) = 0.989

    w1 = 0.983
    w0 = 0.017
 delta = -0.971

   ATE = -6751
   ATT = 928.4
   ATU = -6840

OLS = w1*ATT + w0*ATU = 793.6
```

To interpret these estimates, note that `OLS` is the estimated regression coefficient on the variable designated as treatment. `P(d=1)` and `P(d=0)` correspond to $\hat{\rho}$ and $1 - \hat{\rho}$, respectively. `w1`, `w0`, and `delta` correspond to $\hat{w}_1$, $\hat{w}_0$, and $\hat{\delta}$, respectively. Finally, `ATE`, `ATT`, and `ATU` correspond to $\hat{\tau}_{APLE}$, $\hat{\tau}_{APLE,1}$, and $\hat{\tau}_{APLE,0}$, respectively. Type

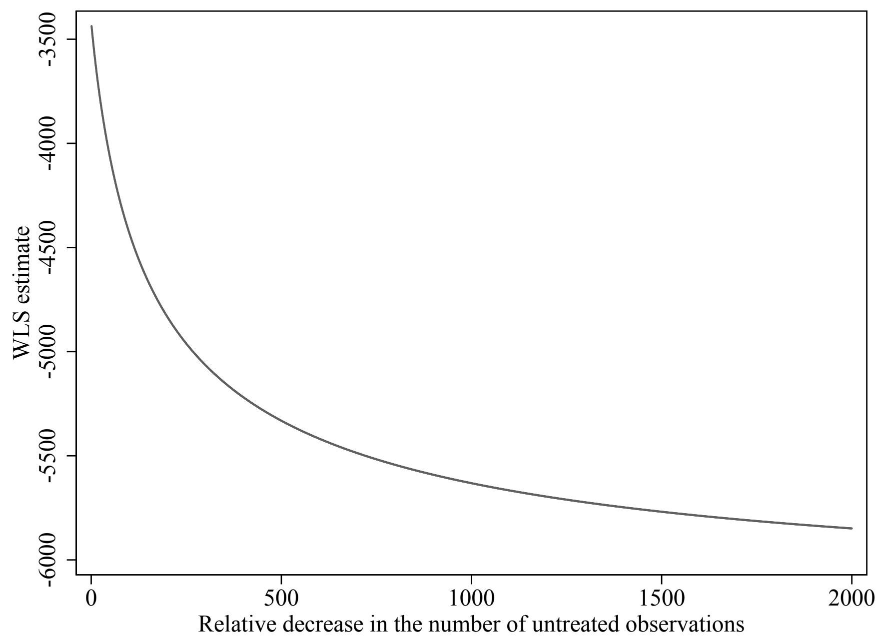
```
> ?hettreatreg
```

for more information and an additional example. Further information is also available from CRAN at https://CRAN.R-project.org/package=hettreatreg.

Appendix E   Robustness Checks

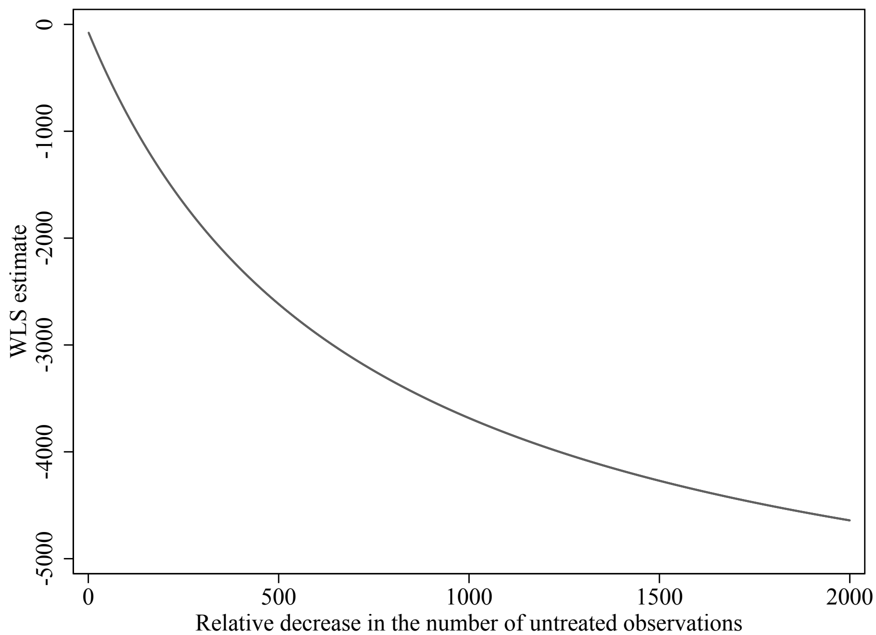
E1   The Effects of a Training Program on Earnings

Figure E1.1: WLS Estimates of the Effects of a Training Program on Earnings



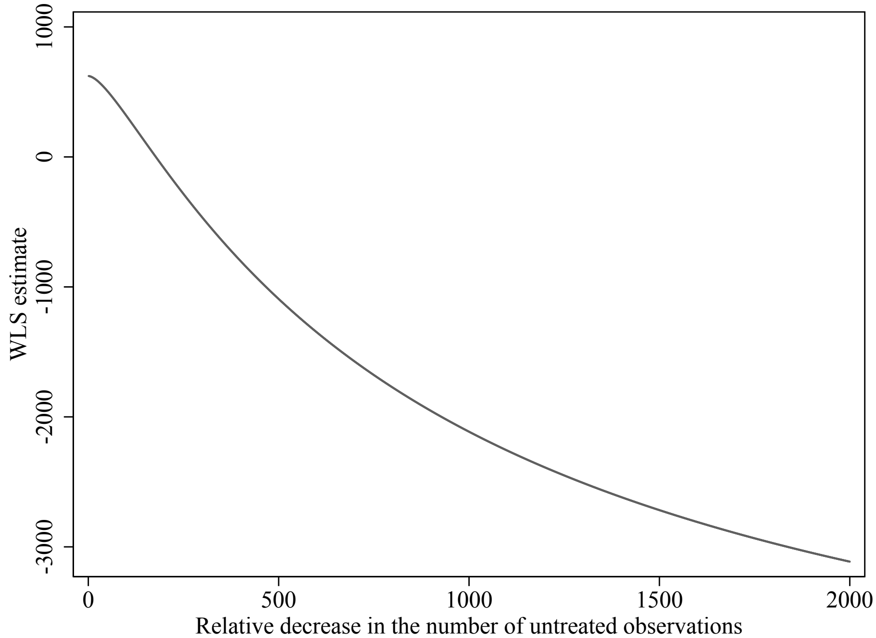
Notes: The vertical axis represents WLS estimates of the effect of NSW program on earnings in 1978 using the model in equation (1) in the main text and the specification in column 1 of Table 1, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

Figure E1.2: WLS Estimates of the Effects of a Training Program on Earnings



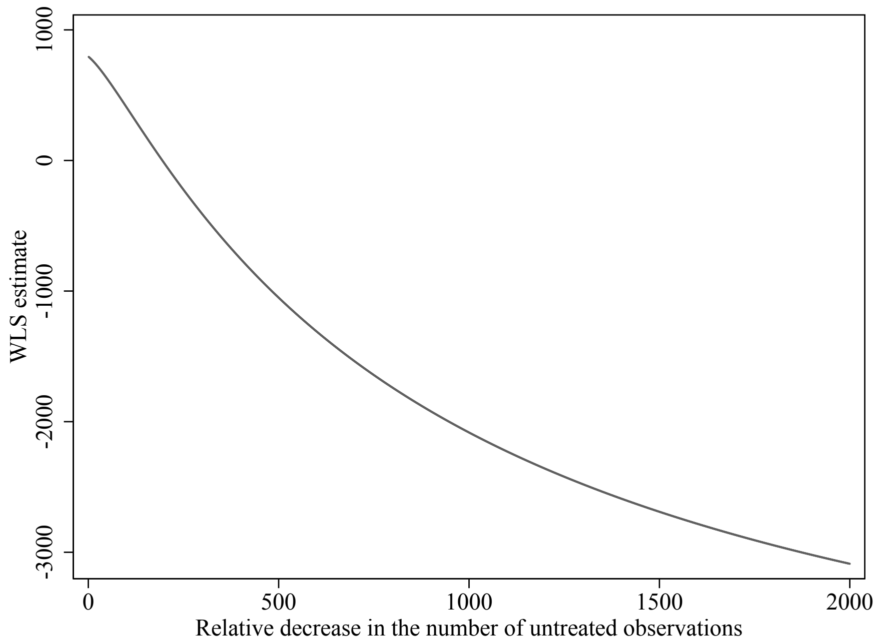
Notes: The vertical axis represents WLS estimates of the effect of NSW program on earnings in 1978 using the model in equation (1) in the main text and the specification in column 2 of Table 1, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

Figure E1.3: WLS Estimates of the Effects of a Training Program on Earnings



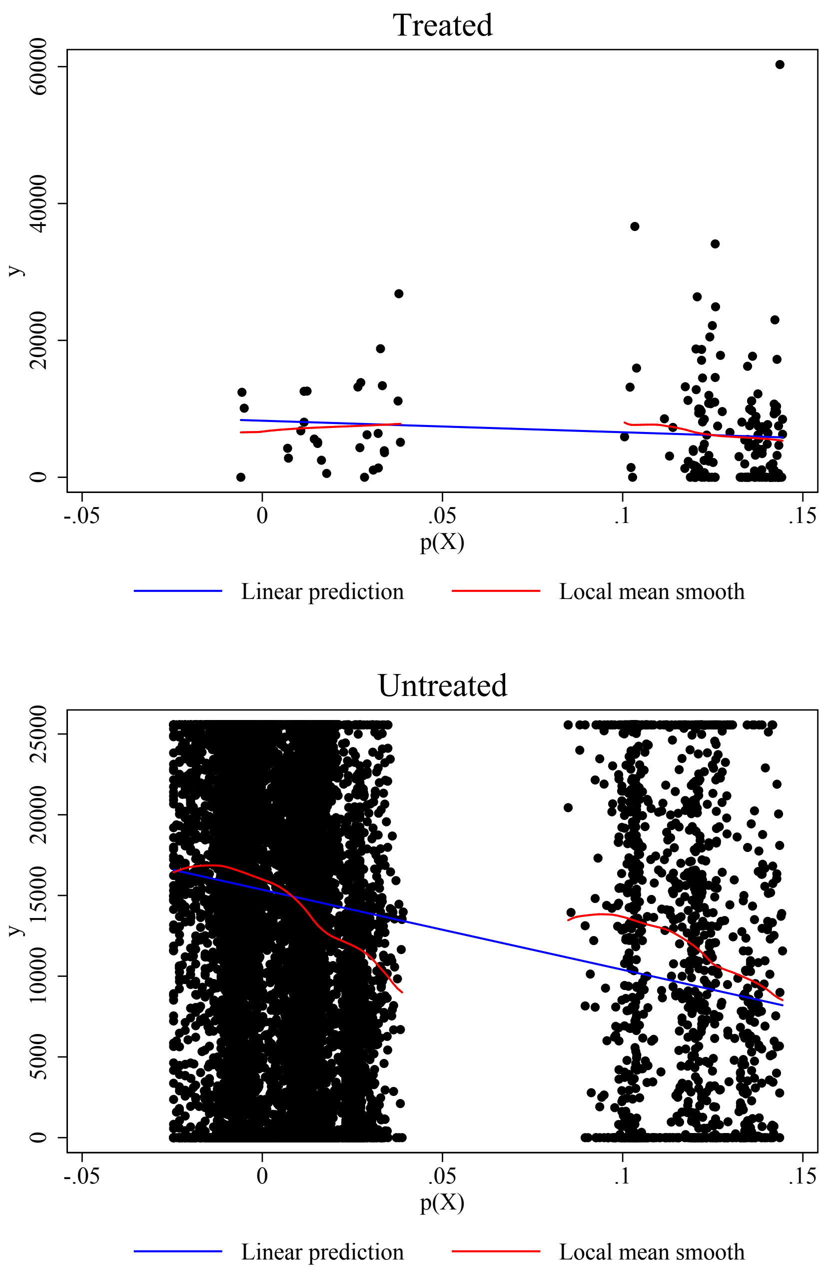
Notes: The vertical axis represents WLS estimates of the effect of NSW program on earnings in 1978 using the model in equation (1) in the main text and the specification in column 3 of Table 1, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

Figure E1.4: WLS Estimates of the Effects of a Training Program on Earnings



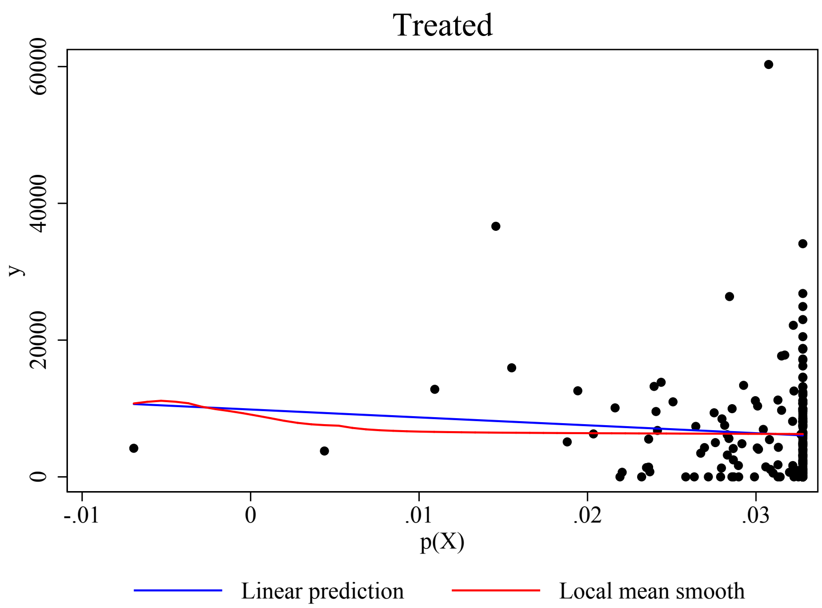
Notes: The vertical axis represents WLS estimates of the effect of NSW program on earnings in 1978 using the model in equation (1) in the main text and the specification in column 4 of Table 1, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

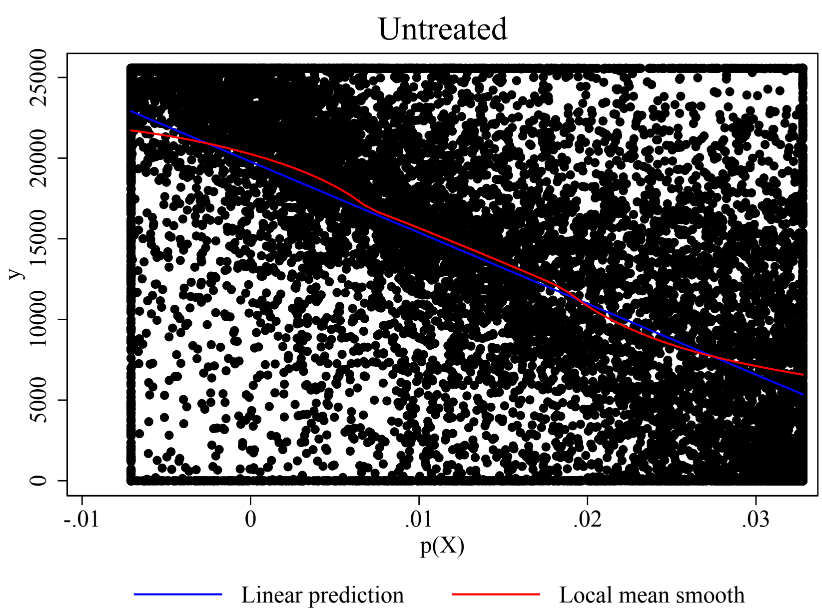
Figure E1.5: Relationship Between Earnings and $p\,(X)$



Notes: The vertical axis represents earnings in 1978. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 1 of Table 1. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

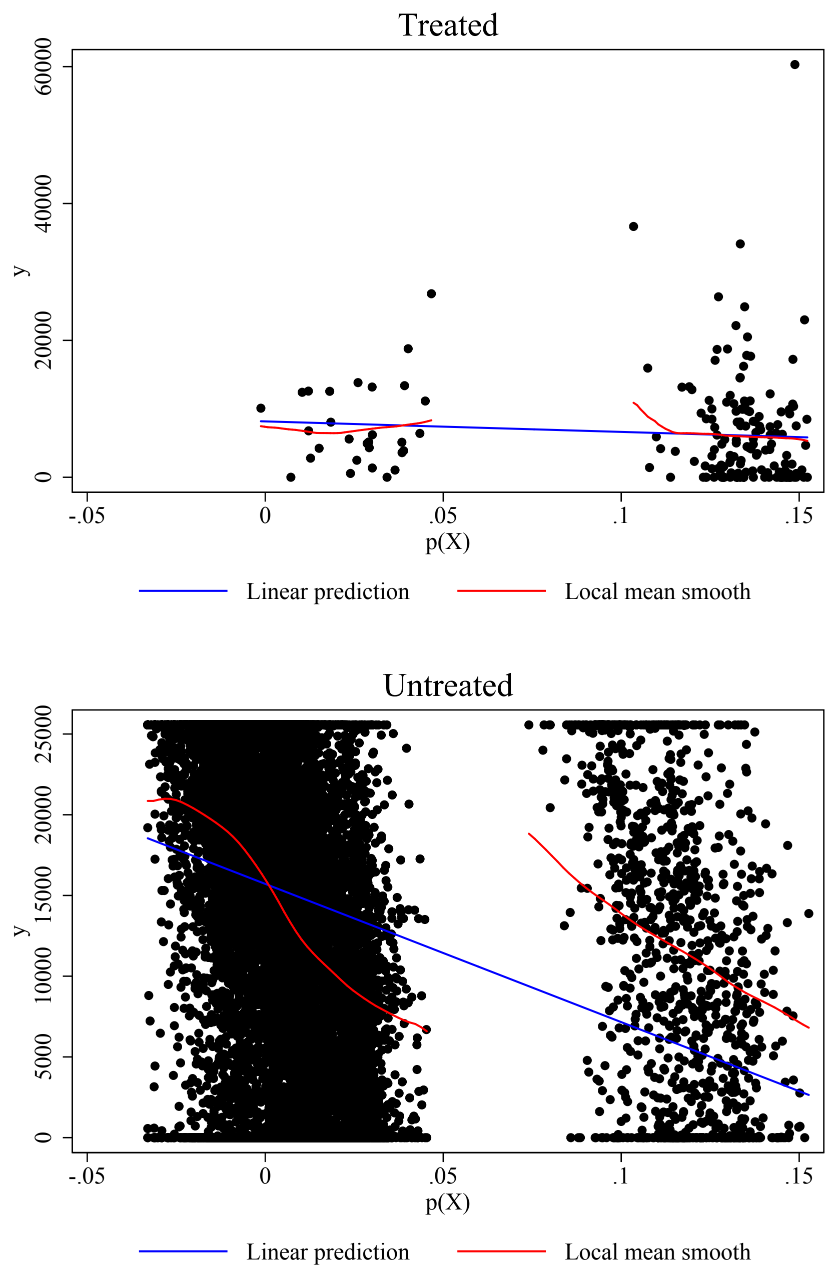
Figure E1.6: Relationship Between Earnings and $p(X)$





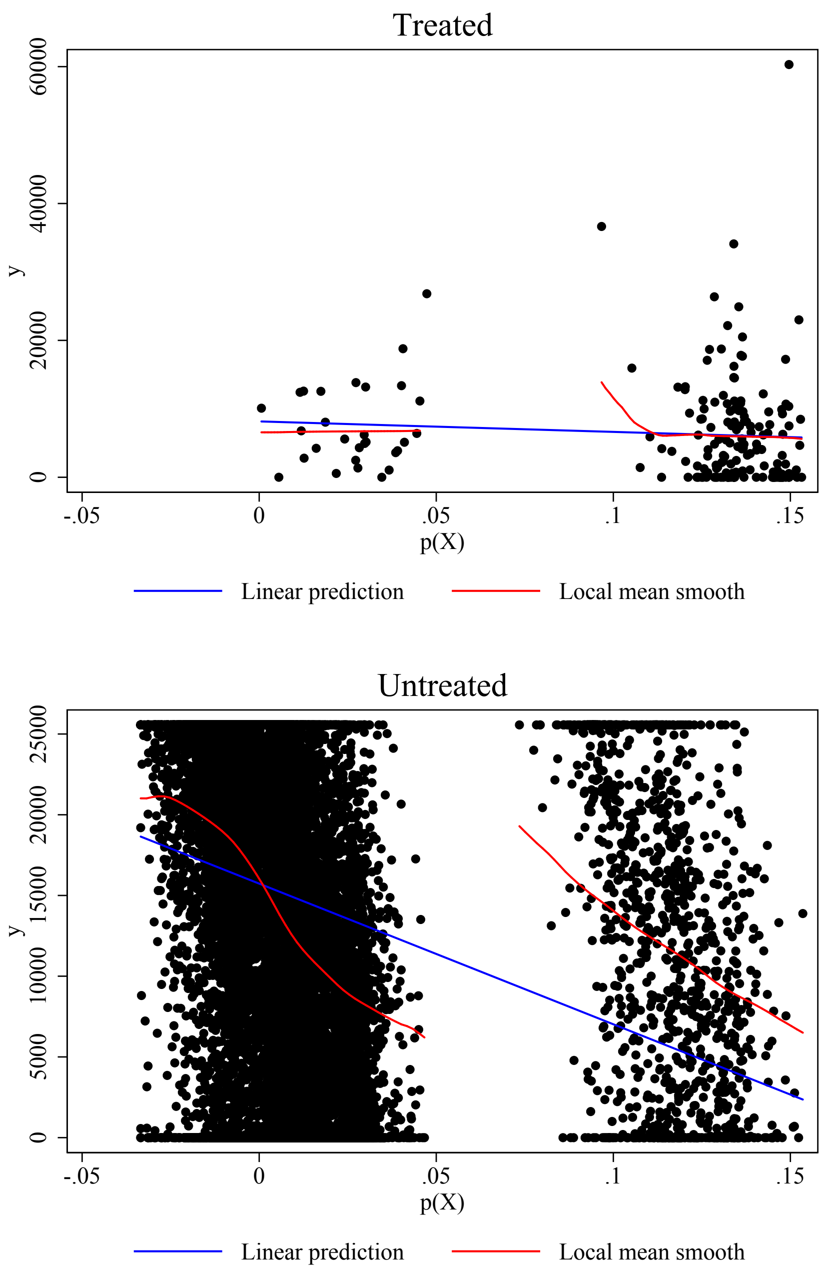
Notes: The vertical axis represents earnings in 1978. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 2 of Table 1. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

Figure E1.7: Relationship Between Earnings and $p(X)$



Notes: The vertical axis represents earnings in 1978. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 3 of Table 1. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

Figure E1.8: Relationship Between Earnings and $p\,(X)$



Notes: The vertical axis represents earnings in 1978. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 4 of Table 1. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

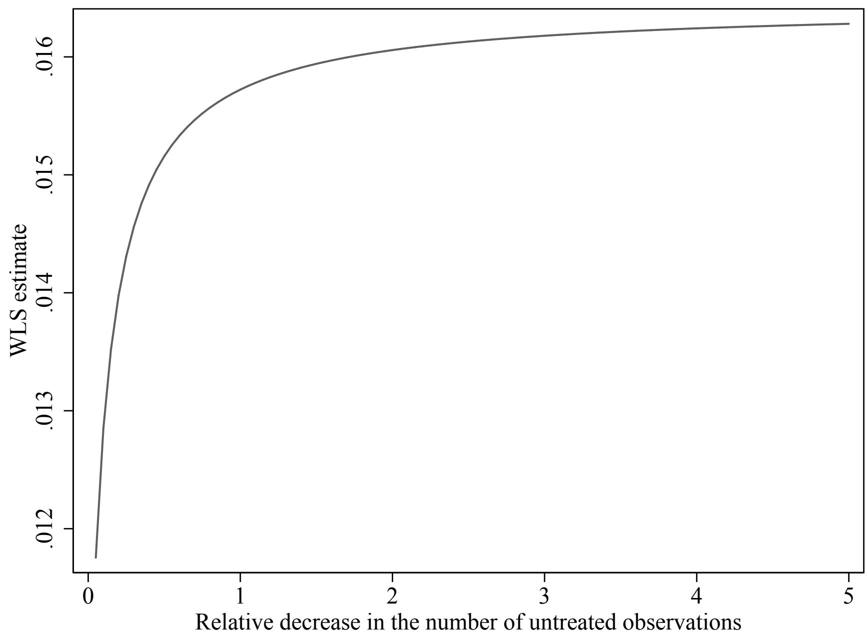
Table E1.1: Alternative Estimates of the Effects of a Training Program on Earnings

	(1)	(2)	(3)	(4)
	Matching on the LPM propensity score			
$\widehat{\text{ATE}}$	–9,227***	–7,504**	–6,245*	–6,581*
	(2,388)	(3,518)	(3,382)	(3,370)
$\widehat{\text{ATT}}$	–3,282***	257	975	–892
	(863)	(694)	(813)	(906)
$\widehat{\text{ATU}}$	–9,295***	–7,594**	–6,328*	–6,646*
	(2,415)	(3,556)	(3,420)	(3,409)
	Matching on the logit propensity score			
$\widehat{\text{ATE}}$	–6,682**	–7,683***	–4,187	–2,961
	(2,773)	(2,421)	(3,012)	(11,900)
$\widehat{\text{ATT}}$	–3,855***	265	2,117**	2,032**
	(854)	(695)	(856)	(860)
$\widehat{\text{ATU}}$	–6,714**	–7,775***	–4,260	–3,018
	(2,804)	(2,448)	(3,046)	(12,037)
	Regression adjustment			
$\widehat{\text{ATE}}$	–6,132***	–6,218**	–4,952*	–4,930
	(1,644)	(2,534)	(2,996)	(3,073)
$\widehat{\text{ATT}}$	–3,417***	–69	623	796
	(628)	(598)	(628)	(639)
$\widehat{\text{ATU}}$	–6,163***	–6,289**	–5,017*	–4,996
	(1,662)	(2,561)	(3,030)	(3,108)
Demographic controls	✓		✓	✓
Earnings in 1974				✓
Earnings in 1975		✓	✓	✓
$\hat{\rho} = \hat{\text{P}}(d = 1)$	0.011	0.011	0.011	0.011
Observations	16,177	16,177	16,177	16,177

Notes: The dependent variable is earnings in 1978. Demographic controls include age, age squared, years of schooling, and indicators for married, high school dropout, black, and Hispanic. For treated individuals, earnings in 1974 correspond to real earnings in months 13–24 prior to randomization, which overlaps with calendar year 1974 for a number of individuals. For "matching on the LPM propensity score" and "matching on the logit propensity score," estimation is based on nearest-neighbor matching on the estimated propensity score (with a single match). The propensity score is estimated using a linear probability model (LPM) or a logit model. For "regression adjustment," estimation is based on the estimator discussed in Kline (2011). Huber–White standard errors (regression adjustment) and Abadie–Imbens standard errors (matching) are in parentheses. Abadie–Imbens standard errors ignore that the propensity score is estimated.
*Statistically significant at the 10% level; **at the 5% level; ***at the 1% level.

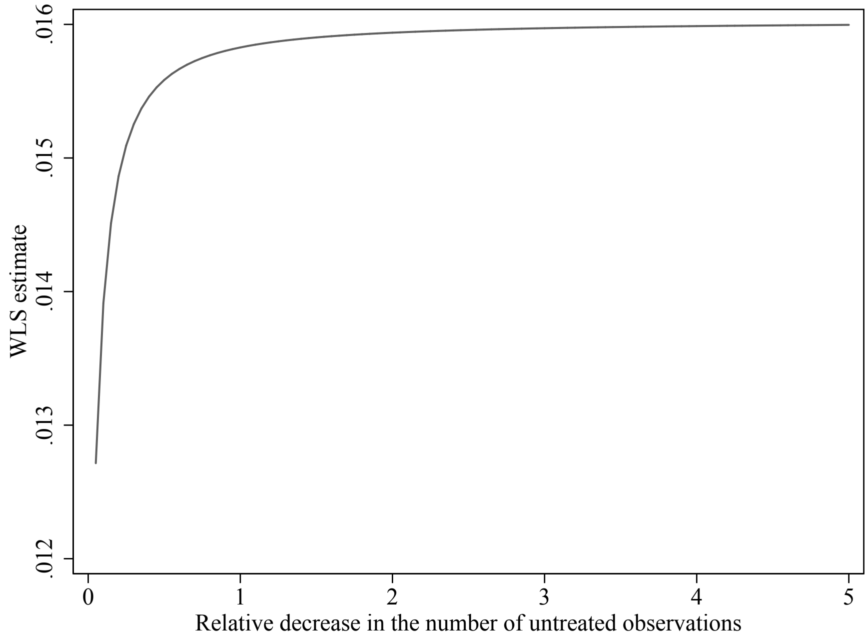
E2 The Effects of Cash Transfers on Longevity

Figure E2.1: WLS Estimates of the Effects of Cash Transfers on Longevity



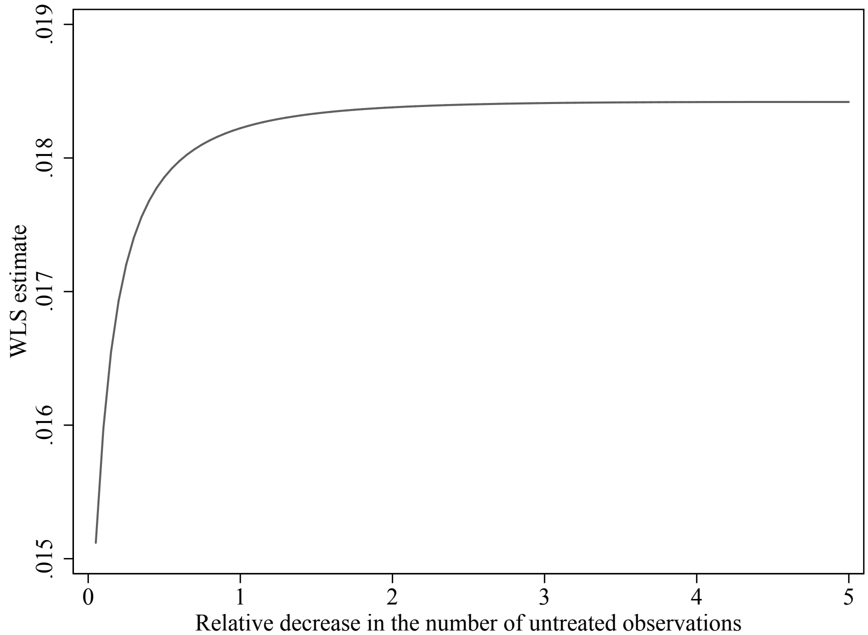
Notes: The vertical axis represents WLS estimates of the effect of cash transfers on log age at death using the model in equation (1) in the main text and the specification in column 1 of Table 2, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

Figure E2.2: WLS Estimates of the Effects of Cash Transfers on Longevity



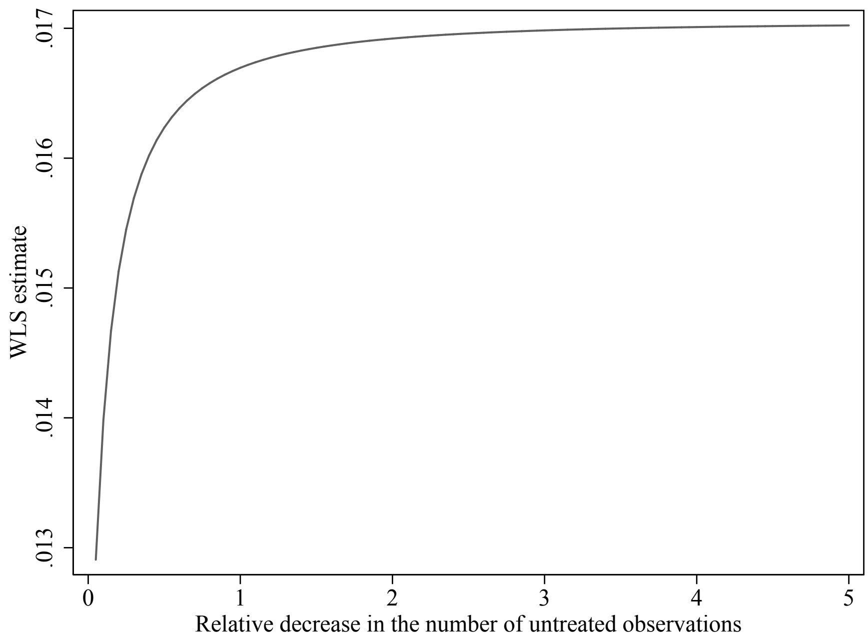
Notes: The vertical axis represents WLS estimates of the effect of cash transfers on log age at death using the model in equation (1) in the main text and the specification in column 2 of Table 2, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

Figure E2.3: WLS Estimates of the Effects of Cash Transfers on Longevity



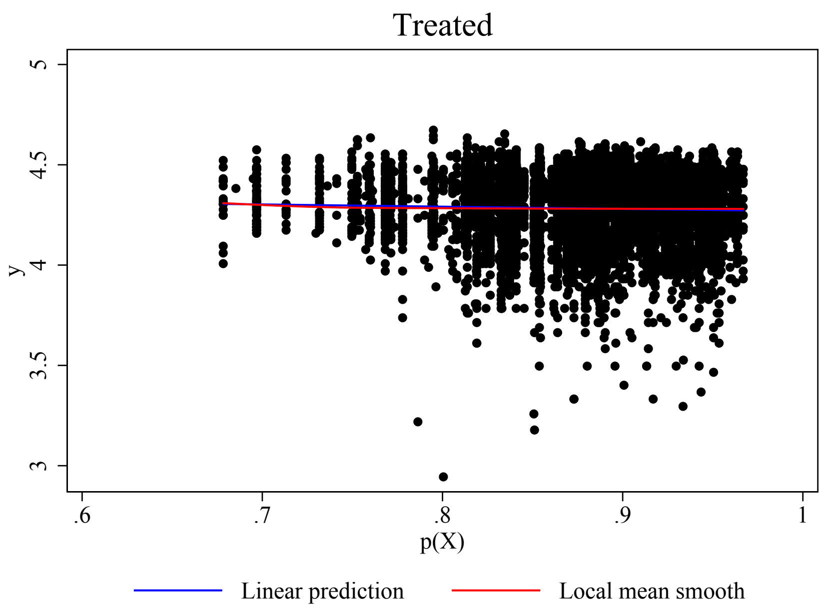
Notes: The vertical axis represents WLS estimates of the effect of cash transfers on log age at death using the model in equation (1) in the main text and the specification in column 3 of Table 2, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

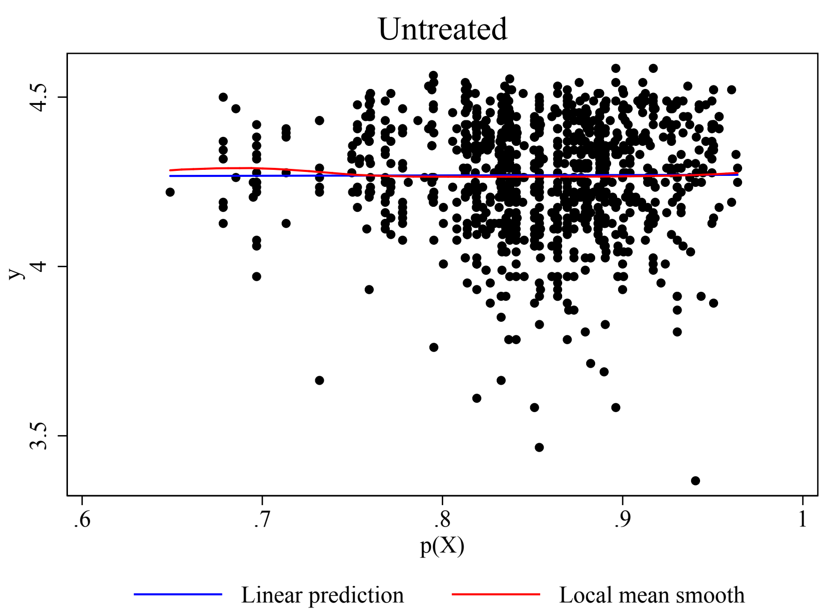
Figure E2.4: WLS Estimates of the Effects of Cash Transfers on Longevity



Notes: The vertical axis represents WLS estimates of the effect of cash transfers on log age at death using the model in equation (1) in the main text and the specification in column 4 of Table 2, with weights of 1 for treated and $\frac{1}{k}$ for untreated units. The horizontal axis represents k.

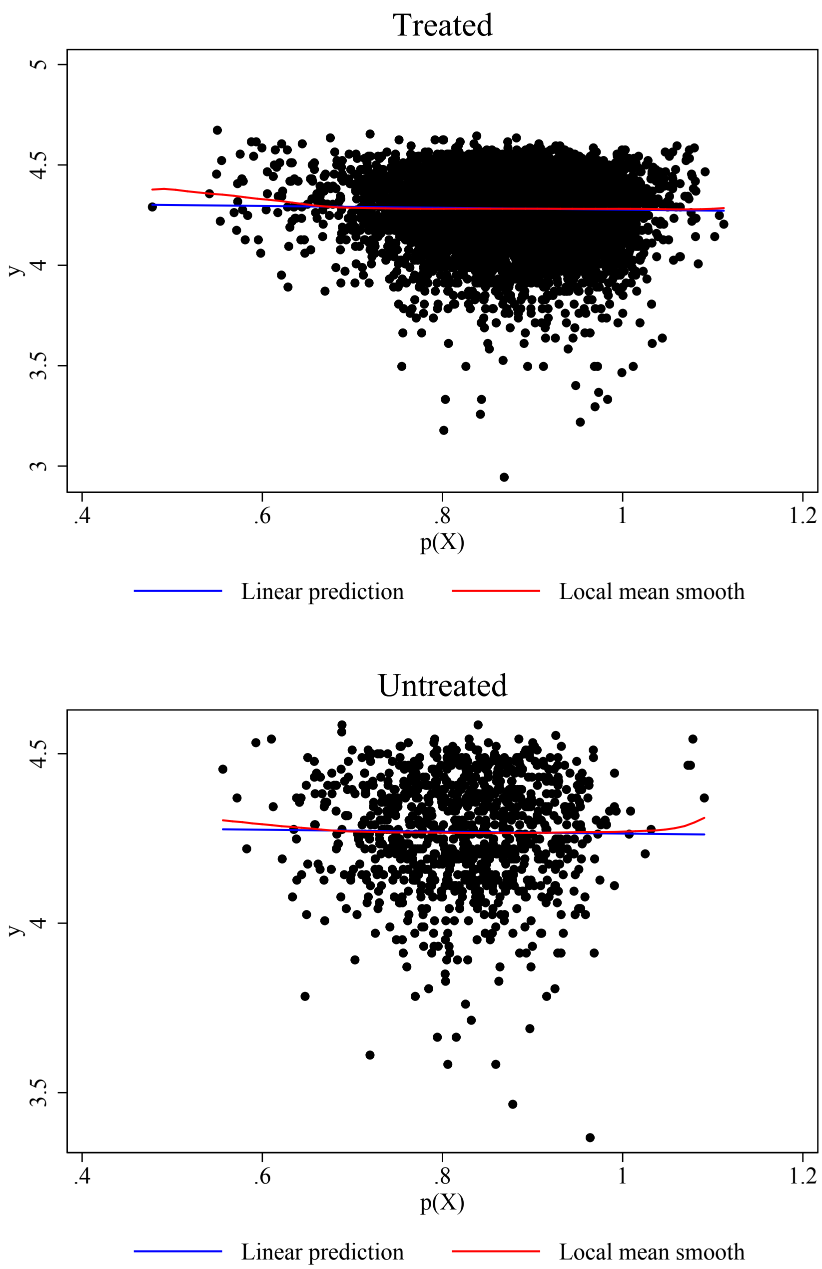
Figure E2.5: Relationship Between Longevity and $p(X)$





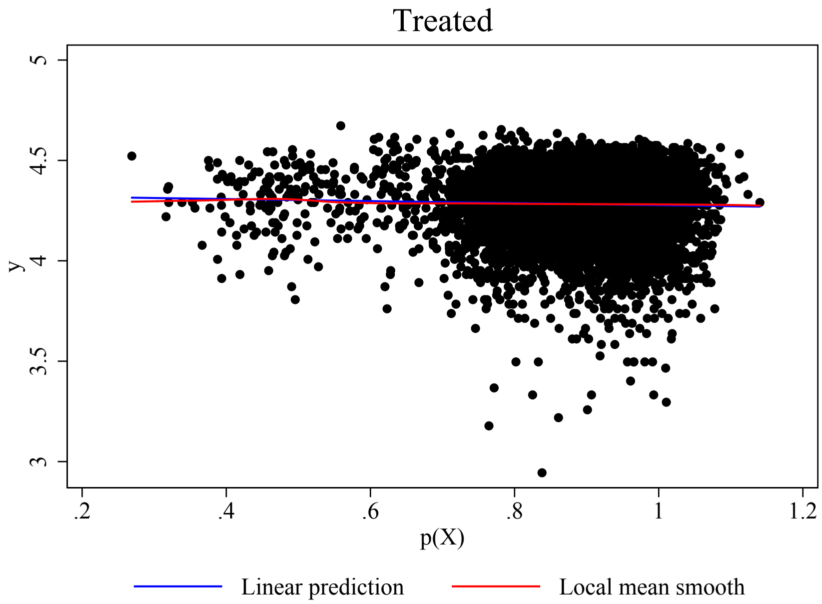
Notes: The vertical axis represents log age at death, as reported in the MP records. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 1 of Table 2. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

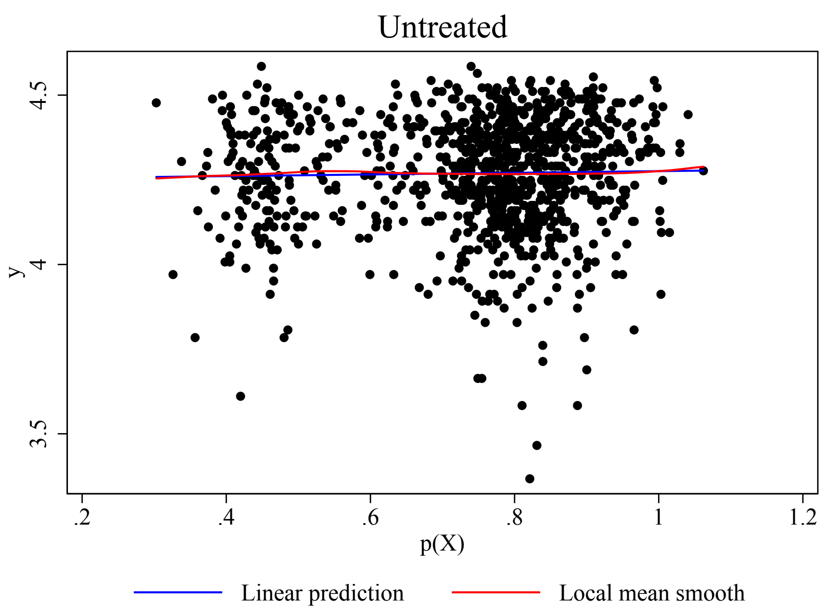
Figure E2.6: Relationship Between Longevity and $p(X)$



Notes: The vertical axis represents log age at death, as reported in the MP records. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 2 of Table 2. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

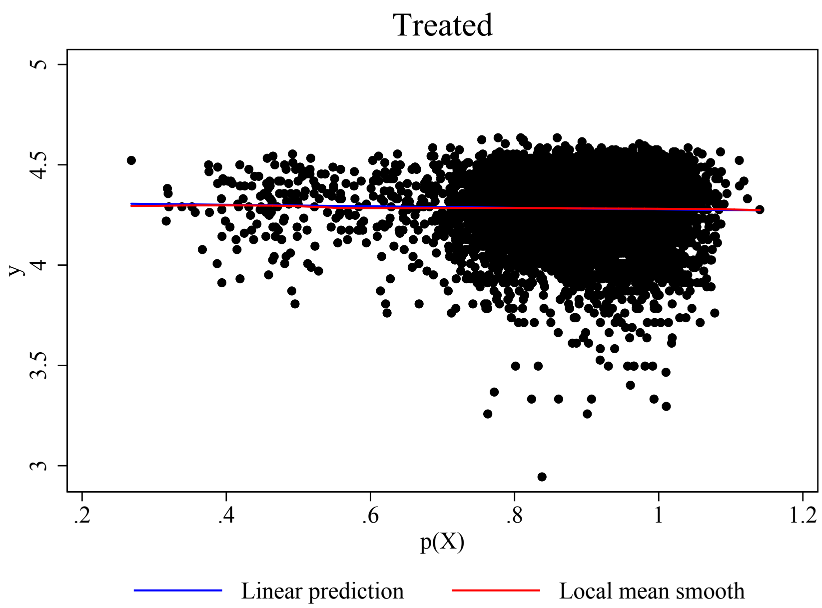
Figure E2.7: Relationship Between Longevity and $p(X)$

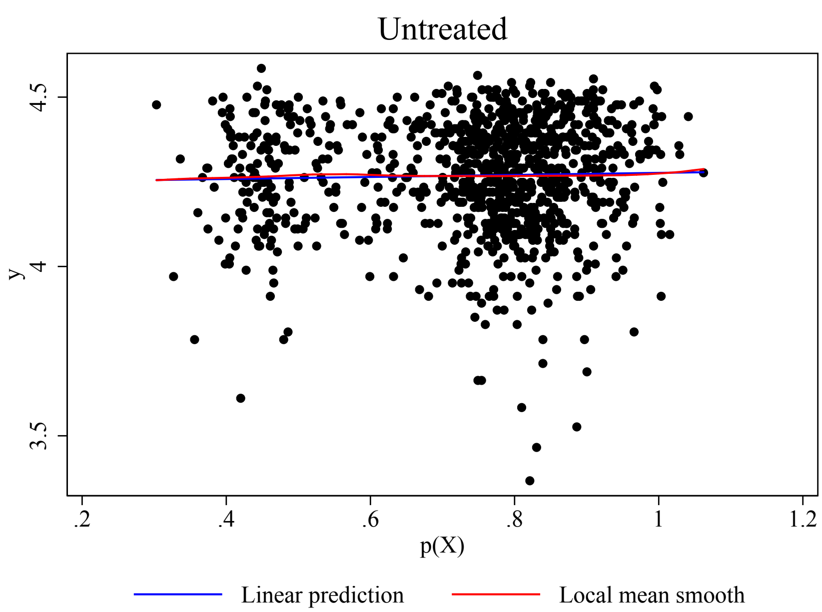




Notes: The vertical axis represents log age at death, as reported in the MP records. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 3 of Table 2. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

Figure E2.8: Relationship Between Longevity and $p(X)$





Notes: The vertical axis represents log age at death, as reported on the death certificate. The horizontal axis represents the LPM propensity score. The propensity score is estimated using the specification in column 4 of Table 2. "Local mean smooth" is estimated using the Epanechnikov kernel and a rule-of-thumb bandwidth.

Table E2.1: Alternative Estimates of the Effects of Cash Transfers on Longevity

	(1)	(2)	(3)	(4)
	Matching on the LPM propensity score			
$\widehat{\text{ATE}}$	0.0110	0.0147*	0.0022	0.0011
	(0.0070)	(0.0089)	(0.0099)	(0.0098)
$\widehat{\text{ATT}}$	0.0106	0.0143	–0.0002	–0.0002
	(0.0073)	(0.0096)	(0.0109)	(0.0107)
$\widehat{\text{ATU}}$	0.0144**	0.0179**	0.0194**	0.0100
	(0.0059)	(0.0082)	(0.0084)	(0.0085)
	Matching on the logit propensity score			
$\widehat{\text{ATE}}$	0.0111	0.0183**	–0.0019	–0.0054
	(0.0073)	(0.0081)	(0.0166)	(0.0166)
$\widehat{\text{ATT}}$	0.0107	0.0181**	–0.0043	–0.0105
	(0.0077)	(0.0087)	(0.0187)	(0.0186)
$\widehat{\text{ATU}}$	0.0145**	0.0193**	0.0152*	0.0309***
	(0.0059)	(0.0083)	(0.0085)	(0.0083)
	Regression adjustment			
$\widehat{\text{ATE}}$	0.0105*	0.0100	0.0140	0.0130
	(0.0063)	(0.0070)	(0.0110)	(0.0110)
$\widehat{\text{ATT}}$	0.0096	0.0092	0.0133	0.0124
	(0.0064)	(0.0073)	(0.0121)	(0.0121)
$\widehat{\text{ATU}}$	0.0164***	0.0160***	0.0184***	0.0170***
	(0.0058)	(0.0061)	(0.0065)	(0.0065)
State fixed effects	✓			
County fixed effects			✓	✓
Cohort fixed effects	✓	✓	✓	✓
State characteristics		✓	✓	✓
County characteristics		✓		
Individual characteristics		✓	✓	✓
$\hat{p} = \hat{\text{P}}(d = 1)$	0.875	0.875	0.875	0.875
Observations	7,860	7,859	7,859	7,857

Notes: The dependent variable is log age at death, as reported in the MP records (columns 1 to 3) or on the death certificate (column 4). State characteristics include manufacturing wages, age of school entry, minimum age for work permit, an indicator for a continuation school requirement, state laws concerning MP transfers (work requirement, reapplication requirement, and maximum amounts for first and second child), and log expenditures on education, charity, and social programs. County characteristics include average value of farm land, mean and SD of socio-economic index, poverty rate, female lfp rate, and shares of urban population, widowed women, children living with single mothers, and children working. Individual characteristics include child age at application, age of oldest and youngest child in family, number of letters in name, and indicators for the number of siblings, the marital status of the mother, and whether date of birth is incomplete. For "matching on the LPM propensity score" and "matching on the logit propensity score," estimation is based on nearest-neighbor matching on the estimated propensity score (with a single match). The propensity score is estimated using a linear probability model (LPM) or a logit model. For "regression adjustment," estimation is based on the estimator discussed in Kline (2011). Huber–White standard errors (regression adjustment) and Abadie–Imbens standard errors (matching) are in parentheses. Abadie–Imbens standard errors ignore that the propensity score is estimated.
*Statistically significant at the 10% level; **at the 5% level; ***at the 1% level.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88:265–296.

Angrist, J. D. (1998). Estimating the labor market impact of voluntary military service using Social Security data on military applicants. *Econometrica*, 66:249–288.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Aronow, P. M. and Samii, C. (2016). Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60:250–267.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8:436–455.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Johns Hopkins University Press.

Elder, T. E., Goddeeris, J. H., and Haider, S. J. (2010). Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics*, 17:284–290.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4A. North-Holland.

Freedman, D. A. (2008a). On regression adjustments in experiments with several treatments. *Annals of Applied Statistics*, 2:176–196.

Freedman, D. A. (2008b). On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40:180–193.

Humphreys, M. (2009). Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities. Unpublished.

Kline, P. (2011). Oaxaca–Blinder as a reweighting estimator. *American Economic Review: Papers & Proceedings*, 101:532–537.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, 7:295–318.

Negi, A. and Wooldridge, J. M. (2019). Revisiting regression adjustment in experiments with heterogeneous treatment effects. Unpublished.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14:693–709.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, 50:301–316.