Web Appendix for Busso, DiNardo, and McCrary (2011)
Part I

Part IA of the Web Appendix provides some specialized calculations for the density function of the propensity score conditional on treatment status. Part IB documents large sample properties of reweighting estimators using a parametric model for the propensity score and collects known results from Abadie and Imbens (2006) on the asymptotic variance of nearest neighbor matching on covariates, for the special case of matching on a single continuous covariate. Part IB also compares the asymptotic variance of normalized and unnormalized reweighting and compares the asymptotic variance of normalized reweighting to that of nearest neighbor matching for the special case of the Frölich (2004) designs. Part IC demonstrates that all of the DGPs studied in the main text are well-behaved in the sense that the semiparametric efficiency bound is finite. Part ID gives details on the generation of pseudo-random numbers. Part IE discusses the issue of existence of the population expectation and variance of the estimators we study.

Part IA: Overlap Plots

Here, we record the formulas for plots of the conditional density and conditional probability function for the propensity score conditional on treatment status. We start with the continuous case, corresponding to the Frölich (2004) design. For simplicity, we drop subscripts. Suppose X is scalar and $p(X) = \alpha + \beta\Lambda(c_0 + c_1 X)$, with $c_1 > 0$, where $\Lambda(z) = \exp(z)/(1 + \exp(z))$. Since $p(\cdot)$ is monotonic increasing, we have $p^{-1}(v) = (\ln(v - \alpha) - \ln(\alpha + \beta - v) - c_0)/c_1$ and by construction $\alpha + \beta\Lambda(c_0 + c_1 p^{-1}(v)) = v$. We assume that T is generated according to $T = \mathbf{1}\left(U \leq \alpha + \beta\Lambda(c_0 + c_1 X)\right)$ where U is distributed standard uniform and independent of X. Fix $v \in (\alpha, \alpha + \beta)$. We begin by finding the distribution function for the propensity score. We have

$$P(p(X) \leq v) \quad = \quad P\left(X \leq p^{-1}(v)\right) = \int_{-\infty}^{p^{-1}(v)} f_X(x)dx \tag{1}$$

Next, differentiate with respect to v to obtain the density function, $f_{p(X)}(v)$. By Leibniz's Rule, we have

$$f_{p(X)}(v) \quad = \quad f_X\left(p^{-1}(v)\right)\frac{d}{dv}p^{-1}(v) = f_X\left(p^{-1}(v)\right)\frac{1}{c_1}\frac{\beta}{(v-\alpha)(\alpha+\beta-v)} \tag{2}$$

We turn now to the distribution function for the propensity score conditional on $T = t$. To do so, we first establish a lemma.

Lemma. *If X and Y are independent and continuous, then $P(Y \leq a + bX, X \leq c) = \int_{-\infty}^{c} f_X(x)F_Y(a + bx)dx$. Under the same condition, $P(Y > a + bX, X \leq c) = F_X(c) - \int_{-\infty}^{c} f_X(x)F_Y(a + bx)dx = \int_{-\infty}^{c} f_X(x)(1 - F_Y(a + bx))dx$.*

Proof. First, note that

$$\begin{aligned} P(Y \leq a + bX, X < c) &= \int_{-\infty}^{c}\int_{-\infty}^{a+bx} f_{X,Y}(x, y)dydx = \int_{-\infty}^{c}\int_{-\infty}^{a+bx} f_X(x)f_Y(y)dydx = \int_{-\infty}^{c} f_X(x)\int_{-\infty}^{a+bx} f_Y(y)dydx \\ &= \int_{-\infty}^{c} f_X(x)F_Y(a + bx)dx \end{aligned}$$

Then, note that $P(Y > a + bX, X \leq c) = P(X \leq c) - P(Y \leq a + bX, X \leq c)$ so that the result above establishes the result for the second case as well as the first. ☐

By the lemma we have

$$\begin{aligned} P\left(p(X) \leq v|T = 1\right) &= P\left(p(X) \leq v, T = 1\right)/P(T = 1) = P\left(p(X) \leq v, U \leq \alpha + \beta\Lambda(c_0 + c_1 X)\right)/q & (3) \\ &= \frac{1}{q}P\left(U \leq \alpha + \beta\Lambda(c_0 + c_1 X), X \leq p^{-1}(v)\right) = \frac{1}{q}\int_{-\infty}^{p^{-1}(v)} f_X(x)F_U(\alpha + \beta\Lambda(c_0 + c_1 x))dx & (4) \\ P\left(p(X) \leq v|T = 0\right) &= \frac{1}{1-q}P\left(p(X) \leq v, U > \alpha + \beta\Lambda(c_0 + c_1 X)\right) = \frac{1}{1-q}P\left(U > \alpha + \beta\Lambda(c_0 + c_1 X), X \leq p^{-1}(v)\right) & (5) \\ &= \frac{1}{1-q}\int_{-\infty}^{p^{-1}(v)} f_X(x)(1 - F_U(\alpha + \beta\Lambda(c_0 + c_1 x)))dx & (6) \end{aligned}$$

By differentiation using Leibniz's Rule, we have

$$\begin{aligned} f_{p(X)|T=1}(v) &= \frac{1}{q}f_X(p^{-1}(v))F_U(\alpha + \beta\Lambda(c_0 + c_1 p^{-1}(v)))\frac{d}{dv}p^{-1}(v) & (7) \\ &= \frac{1}{q}f_{p(X)}(v)(\alpha + \beta\Lambda(c_0 + c_1 p^{-1}(v))) = \frac{1}{q}f_{p(X)}(v)v & (8) \\ f_{p(X)|T=0}(v) &= \frac{1}{1-q}f_X(p^{-1}(v))(1 - F_U(\alpha + \beta\Lambda(c_0 + c_1 p^{-1}(v))))\frac{d}{dv}p^{-1}(v) & (9) \\ &= \frac{1}{1-q}f_{p(X)}(v))(1 - \alpha - \beta\Lambda(c_0 + c_1 p^{-1}(v))) = \frac{1}{1-q}f_{p(X)}(v))(1 - v) & (10) \end{aligned}$$

A simpler derivation for the case of discrete covariates shows that the conditional probability function for $p(X)$ has the form

$$f_{p(X)|T=1}(v_j) = \frac{1}{q}\pi_j v_j \tag{11}$$

$$f_{p(X)|T=0}(v_j) = \frac{1}{1-q}\pi_j(1-v_j) \tag{12}$$

where v_j is a point of support of $p(X)$ and π_j is the probability that X takes on a value x_j such that $p(x_j) = v_j$. In the case of no ties, $p(X)$ is distributed uniform over the set $\{p(x_1), p(x_2), \ldots, p(x_J)\}$, and each outcome in that set occurs with probability $\pi_j = 1/J$.

Part IB: Asymptotic Variance Calculations

In the interest of being self-contained, we briefly review notation and context. For every unit i, we observe $Y_i = T_iY_i(1) + (1 - T_i)Y_i(0)$, T_i and X_i. We want to estimate the population parameter $\theta = \mathbb{E}[Y_i(1) - Y_i(0)|T_i = 1]$, referred to in the main text as TOT. We assume that conditional on X_i, T_i is independent of $Y_i(1)$ and $Y_i(0)$ (conditional independence) and that $1 - p(x) \geq \xi > 0$ for almost every x in the support of X_i (strict overlap). For $t \in \{0, 1\}$, define the conditional expectations $\mu_t(x) = \mathbb{E}[Y_i(t)|X_i = x]$, the conditional variances $\sigma_t^2(x) = V[Y_i(t)|X_i = x]$, and the parameters $\alpha_t = \mathbb{E}[\mu_t(X_i)|T_i = 1]$. Finally, define the covariate-specific treatment effects, $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] = \mu_1(x) - \mu_0(x)$ and note that $\theta = \mathbb{E}[\tau(X_i)|T_i = 1] = \alpha_1 - \alpha_0$. The researcher observes Y_i, T_i, and X_i for all units. The first step propensity score is based on a logit model using covariate vector Z_i which contains a constant as well as functions of X_i. An overview of the results recorded here is given in Web Appendix Table 1.

1 The Unnormalized True Weights Estimator, $\widehat{\theta}_{U,tw}$

Consider

$$\widehat{\theta}_{U,tw} = \frac{\sum_i Y_iT_i}{\sum_i T_i} - \frac{\sum_j Y_j(1-T_j)W_j}{\sum_j T_j} \tag{13}$$

where $W_j = p(X_j)/(1-p(X_j))$. Let $\gamma = (\theta, q)'$ denote the TOT parameter and probability of treatment. Then $\widehat{\gamma} = (\widehat{\theta}_{U,tw}, \sum_i T_i/n)$ solves $0 = \frac{1}{n}\sum_i m_i(\widehat{\gamma})$, where $m_i(\gamma)$ and its derivative matrix are given by

$$m_i(\gamma) = \begin{pmatrix} (T_iY_i - (1-T_i)W_iY_i)/q - \theta \\ T_i - q \end{pmatrix} \quad \text{and} \quad M_i(\gamma) = -\begin{pmatrix} 1 & \frac{1}{q^2}(T_iY_i - (1-T_i)W_iY_i) \\ 0 & 1 \end{pmatrix} \tag{14}$$

where $q = P(T_i = 1)$. Evaluated at γ^* and assuming no misspecification of the propensity score, the expectation of the derivative and the variance of the moments are given by

$$M = -\begin{pmatrix} 1 & \theta/q \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} a & b \\ b & q(1-q) \end{pmatrix} \tag{15}$$

where $a = V[Y_iT_i - Y_i(1-T_i)W_i]/q^2$ and $b = \mathbb{E}[\mu_1(X_i)p(X_i)]/q - q\theta$. So the asymptotic variance of $\widehat{\gamma}$ is

$$M^{-1}\Sigma M'^{-1} = \begin{pmatrix} 1 & -\theta/q \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ b & q(1-q) \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\theta/q & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\theta/q \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a - b\theta/q & b \\ b - \theta(1-q) & q(1-q) \end{pmatrix} \tag{16}$$

$$= \begin{pmatrix} a - 2b\theta/q + \theta^2(1-q)/q & b - \theta(1-q) \\ b - \theta(1-q) & q(1-q) \end{pmatrix} \tag{17}$$

The asymptotic variance of $\widehat{\theta}_{U,tw}$ is thus

$$AV(\widehat{\theta}_{U,tw}) = \frac{1}{q^2}\left\{q^2a - 2\theta bq + \theta^2 q(1-q)\right\} = \frac{1}{q^2}\left\{V[Y_iT_i - Y_i(1-T_i)W_i] - 2\theta\left(\mathbb{E}[\mu_1(X_i)p(X_i)] - q^2\theta\right) + \theta^2 q(1-q)\right\} \tag{18}$$

$$= A + \frac{1}{q^2}\left\{\mathbb{E}\left[\mu_1(X_i)^2p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - 2\theta\mathbb{E}[\mu_1(X_i)p(X_i)] + \theta^2 q\right\} \tag{19}$$

$$= \Omega_\theta + \frac{1}{q^2}\left\{\mathbb{E}\left[\mu_1(X_i)^2p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - \mathbb{E}\left[(\tau(X_i)-\theta)^2p(X_i)\right] - 2\theta\mathbb{E}[\mu_1(X_i)p(X_i)] + \theta^2 q\right\}$$

$$= \Omega_\theta + \frac{1}{q^2}\left\{\mathbb{E}\left[\mu_1(X_i)^2p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - \mathbb{E}\left[\tau(X_i)^2p(X_i)\right] + \theta^2 q - 2\theta\mathbb{E}[\mu_1(X_i)p(X_i)] + \theta^2 q\right\} \tag{20}$$

$$= \Omega_\theta + \frac{1}{q^2}\{\mathbb{E}\left[\mu_1(X_i)^2p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - \left(\mathbb{E}\left[\mu_1(X_i)^2p(X_i) - 2\mu_1(X_i)\mu_0(X_i)p(X_i) + \mu_0(X_i)^2p(X_i)\right]\right)$$
$$-2\theta\mathbb{E}[\mu_1(X_i)p(X_i)] + 2\theta^2 q\} \tag{21}$$

$$= \Omega_\theta + \frac{1}{q^2}\{\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\mathbb{E}\left[\mu_1(X_i)\mu_0(X_i)p(X_i)\right] - 2\mathbb{E}\left[\mu_0(X_i)^2p(X_i)\right] - 2\theta\mathbb{E}[\mu_1(X_i)p(X_i)] + 2\theta^2 q\}$$

$$= \Omega_\theta + \frac{1}{q^2}\{\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\mathbb{E}\left[\mu_1(X_i)\mu_0(X_i)p(X_i)\right] - 2\mathbb{E}\left[\mu_0(X_i)^2p(X_i)\right] - 2\theta\mathbb{E}[\mu_0(X_i)p(X_i)]\} \tag{22}$$

$$= \Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}C[\tau(X_i), \mu_0(X_i)|T_i = 1] \tag{23}$$

where we use the definitions

$$A = \frac{1}{q^2}\left\{\mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)}\right]\right\} \tag{24}$$

$$\Omega_\theta = \frac{1}{q^2}\left\{\mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)}\right] + \mathbb{E}\left[(\tau(X_i)-\theta)^2 p(X_i)\right]\right\} \tag{25}$$

and the results

$$\mathbb{E}[Y_iT_i] = \mathbb{E}[\mathbb{E}[Y_iT_i|X_i]] = \mathbb{E}[\mathbb{E}[Y_i(1)T_i|X_i]] = \mathbb{E}[\mu_1(X_i)p(X_i)] \tag{26}$$

$$\mathbb{E}[(1-T_i)W_iY_i] = \mathbb{E}[\mathbb{E}[(1-T_i)W_iY_i|X_i]] = \mathbb{E}[W_i\mathbb{E}[(1-T_i)Y_i(0)|X_i]] = \mathbb{E}[W_i(1-p(X_i))\mu_0(X_i)] \tag{27}$$

$$= \mathbb{E}[\mu_0(X_i)p(X_i)] \tag{28}$$

$$V[Y_iT_i] = \mathbb{E}[Y_i(1)^2T_i] - \mathbb{E}[Y_iT_i]^2 = \mathbb{E}\left[p(X_i)\left(\sigma_1^2(X_i)+\mu(X_i)^2\right)\right] - \mathbb{E}\left[\mu_1(X_i)p(X_i)\right]^2 \tag{29}$$

$$= \mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \mu_1(X_i)^2p(X_i)\right] - \mathbb{E}\left[\mu_1(X_i)p(X_i)\right]^2 \tag{30}$$

$$V[(1-T_i)W_iY_i] = \mathbb{E}\left[(1-T_i)W_i^2Y_i^2\right] - \mathbb{E}\left[(1-T_i)W_iY_i\right]^2 = \mathbb{E}\left[\mathbb{E}\left[(1-T_i)W_i^2Y_i(0)^2|X_i\right]\right] - \mathbb{E}\left[\mu_0(X_i)p(X_i)\right]^2 \tag{31}$$

$$= \mathbb{E}\left[\sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)}\right] + \mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - \mathbb{E}\left[\mu_0(X_i)p(X_i)\right]^2 \tag{32}$$

$$C[Y_iT_i\ ,\ (1-T_i)W_iY_i] = -\mathbb{E}[\mu_1(X_i)p(X_i)]\mathbb{E}[\mu_0(X_i)p(X_i)] \tag{33}$$

$$V[Y_iT_i - Y_i(1-T_i)W_i] = V[Y_iT_i] + V[Y_i(1-T_i)W_i] - 2C[Y_iT_i\ ,\ Y_i(1-T_i)W_i] \tag{34}$$

$$= \mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)}\right] + \mathbb{E}\left[\mu_1(X_i)^2p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - (q\theta)^2 \tag{35}$$

$$\equiv A + \mathbb{E}\left[\mu_1(X_i)^2p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] - (q\theta)^2 \tag{36}$$

$$\mathbb{C}\left[\tau(X_i), \mu_0(X_i)|T_i=1\right] = \frac{1}{q}\left\{\mathbb{E}\left[\mu_1(X_i)\mu_0(X_i)p(X_i)\right] - \mathbb{E}\left[\mu_0(X_i)^2p(X_i)\right] - \theta\mathbb{E}\left[\mu_0(X_i)p(X_i)\right]\right\} \tag{37}$$

2 The Unnormalized Estimated Weights Estimator with Parametric Logit, $\widehat{\theta}_{U,pw}$

Let $\gamma = (\theta, \pi', q)'$ denote the TOT parameter, the logit coefficients, and and the probability of treatment. The moment function and its derivative matrix are given by

$$m_i(\gamma) = \begin{pmatrix} \frac{1}{q}\left(T_iY_i - (1-T_i)W_iY_i\right) - \theta \\ \left(T_i - \Lambda_i\right)Z_i \\ T_i - q \end{pmatrix} \quad \text{and} \quad M_i(\gamma) = -\begin{pmatrix} 1 & \frac{1}{q}(1-T_i)W_iY_iZ_i' & \frac{1}{q^2}\left(T_iY_i - (1-T_i)W_iY_i\right) \\ 0 & \Lambda_i(1-\Lambda_i)Z_iZ_i' & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{38}$$

where as before we write $\Lambda_i = \Lambda(Z_i'\pi)$. Evaluated at γ^* and assuming no misspecification of the propensity score, the expectation of the derivative and the variance of the moments are given by

$$M = -\begin{pmatrix} 1 & c_0' & \theta/q \\ 0 & \mathcal{I} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} a & d'+c_0' & b \\ d+c_0 & \mathcal{I} & c \\ b & c' & q(1-q) \end{pmatrix} \tag{39}$$

respectively, where as before $a = V[Y_iT_i - Y_i(1-T_i)W_i]/q^2$ and $b = \mathbb{E}[\mu_1(X_i)p(X_i)]/q - q\theta$, and where now we additionally define

$$c_t = \mathbb{E}\left[\mu_t(X_i)p(X_i)Z_i\right]/q \tag{40}$$

$$c = \mathbb{E}\left[p(X_i)(1-p(X_i))Z_i\right] \tag{41}$$

$$d = \mathbb{E}\left[p(X_i)(1-p(X_i))\tau(X_i)Z_i\right]/q \tag{42}$$

So the asymptotic variance of $\widehat{\gamma}$ is

$$M^{-1}\Sigma M'^{-1} = \begin{pmatrix} 1 & -c_0'\mathcal{I}^{-1} & -\theta/q \\ 0 & \mathcal{I}^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a & d'+c_0' & b \\ d+c_0 & \mathcal{I} & c \\ b & c' & q(1-q) \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ -\mathcal{I}^{-1}c_0 & \mathcal{I}^{-1} & 0 \\ -\theta/q & 0 & 1 \end{pmatrix} \tag{43}$$

$$= \begin{pmatrix} 1 & -c_0'\mathcal{I}^{-1} & -\theta/q \\ 0 & \mathcal{I}^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a - (d'+c_0')\mathcal{I}^{-1} - b\theta/q & (d'+c_0')\mathcal{I}^{-1} & b \\ d - c\theta/q & I & c \\ b - c'\mathcal{I}^{-1}c_0 - \theta(1-q) & c'\mathcal{I}^{-1} & q(1-q) \end{pmatrix} \tag{44}$$

$$= \begin{pmatrix} a - 2b\frac{\theta}{q} + \theta^2\frac{1-q}{q} - c_0'\mathcal{I}^{-1}c_0 - 2d'\mathcal{I}^{-1}c_0 + 2\frac{\theta}{q}c'\mathcal{I}^{-1}c_0 & d'\mathcal{I}^{-1} - \frac{\theta}{q}c'\mathcal{I}^{-1} & b - c_0'\mathcal{I}^{-1}c - \theta(1-q) \\ \mathcal{I}^{-1}d - \frac{\theta}{q}\mathcal{I}^{-1}c & \mathcal{I}^{-1} & \mathcal{I}^{-1}c \\ b - c'\mathcal{I}^{-1}c_0 - \theta(1-q) & c'\mathcal{I}^{-1} & q(1-q) \end{pmatrix} \tag{45}$$

It will be useful to note that

$$\mathbb{C}\left[\tau(X_i), (1-p(X_i))Z_i|T_i=1\right] = \frac{1}{q}\mathbb{E}\left[(\tau(X_i)-\theta)p(X_i)(1-p(X_i))Z_i\right] = d - \frac{\theta}{q}c \equiv e \tag{46}$$

The asymptotic variance of $\widehat{\theta}_{U,pw}$ is thus

$$
\begin{aligned}
AV(\widehat{\theta}_{U,pw}) &= a - 2b\frac{\theta}{q} + \theta^2\frac{1-q}{q} - c_0'\mathcal{I}^{-1}c_0 - 2d'\mathcal{I}^{-1}c_0 + 2\frac{\theta}{q}c'\mathcal{I}^{-1}c_0 = AV(\widehat{\theta}_{U,tw}) - c_0'\mathcal{I}^{-1}c_0 - 2d'\mathcal{I}^{-1}c_0 + 2\frac{\theta}{q}c'\mathcal{I}^{-1}c_0 \quad (47)\\
&= AV(\widehat{\theta}_{U,tw}) - c_0'\mathcal{I}^{-1}c_0 - 2e'\mathcal{I}^{-1}c_0 \quad (48)\\
&= \Omega_\theta \quad (49)\\
&+ \frac{1}{q^2}\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] - c_0\mathcal{I}^{-1}c_0 \quad (50)\\
&+ 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)|T_i=1\right] - 2e'\mathcal{I}^{-1}c_0 \quad (51)
\end{aligned}
$$

3 The Unnormalized Estimated Weights Estimator with Fully Saturated Logit, $\widehat{\theta}_{U,ew}$

Now suppose that the covariates X_i are discrete, taking on values $x_1, x_2, \ldots, x_J$ with $P(X_i = x_j) = \eta_j$. Then we have

$$
\begin{aligned}
c_0'\mathcal{I}^{-1}c_0 &= \frac{1}{q^2}\sum_j \frac{\eta_j^2\mu_0(x_j)^2p(x_j)^2}{\eta_j p(x_j)(1-p(x_j))} = \frac{1}{q^2}\sum_j \eta_j\mu_0(x_j)^2\frac{p(x_j)}{1-p(x_j)} \quad (52)\\
&= \frac{1}{q^2}\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] \quad (53)\\
\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)|T_i=1\right] &= \frac{1}{q^2}\mathbb{E}\left[(\tau(X_i)-\theta)\,\mu_0(X_i)p(X_i)\right] \quad (54)\\
&= \frac{1}{q^2}\sum_j \eta_j(\tau(x_j)-\theta)\mu_0(x_j)p(x_j) \quad (55)\\
&= \frac{1}{q^2}\sum_j \frac{\eta_j^2(\tau(x_j)-\theta)\mu_0(x_j)p(x_j)^2(1-p(x_j))}{\eta_j p(x_j)(1-p(x_j))} \quad (56)\\
&= \frac{1}{q^2}\mathbb{E}\left[(\tau(X_i)-\theta)\,p(X_i)(1-p(X_i))Z_i\right]\mathcal{I}^{-1}\mathbb{E}\left[\mu_0(X_i)p(X_i)\right] \quad (57)\\
&= e'\mathcal{I}^{-1}\mathbb{E}\left[\mu_0(X_i)Z_i|T_i=1\right] \quad (58)
\end{aligned}
$$

Because of these results, the terms in equations (50) and (51) both cancel out, and the asymptotic variance of $\widehat{\theta}_{U,ew}$ is simply

$$
AV(\widehat{\theta}_{U,ew}) = \Omega_\theta \quad (59)
$$

4 The Normalized True Weights Estimator, $\widehat{\theta}_{N,tw}$

The normalized true weights estimator is given by

$$
\widehat{\theta}_{N,tw} = \frac{\sum_i T_iY_i}{\sum_i T_i} - \frac{\sum_j(1-T_j)W_jY_j}{\sum_j(1-T_j)W_j} \quad (60)
$$

Define $\alpha_t = \mathbb{E}\left[\mu_t(X_i)p(X_i)\right]/q = \mathbb{E}[Y_i(t)|T_i=1]$ and $\gamma = (\theta, \alpha_0)'$ and consider the moments and their derivative matrix

$$
m_i(\gamma) = V_i(Y_i - \theta T_i - \alpha_0)\begin{pmatrix} T_i \\ 1 \end{pmatrix} \quad \text{and} \quad M_i(\gamma) = -V_i\begin{pmatrix} T_i & T_i \\ T_i & 1 \end{pmatrix} \quad (61)
$$

where $V_i = T_i + (1-T_i)W_i$. These are simply the first- and second-order conditions for a least squares regression of Y_i on a constant and T_i, weighted by V_i. Solving this GMM problem yields the estimate $\widehat{\theta}_{N,tw}$ and the counterfactual mean $\widehat{\alpha}_0$. Evaluated at γ^* and assuming no misspecification of the propensity score, the expectation of the derivative and the variance of the moments are given by

$$
M = -q\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \quad (62)
$$

and so the asymptotic variance of $\widehat{\gamma}$ is given by

$$
\begin{aligned}
M^{-1}\Sigma M'^{-1} &= \frac{1}{q^2}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \quad (63)\\
&= \frac{1}{q^2}\begin{pmatrix} 4\Sigma_{11} - 4\Sigma_{12} + \Sigma_{22} & 3\Sigma_{12} - 2\Sigma_{11} - \Sigma_{22} \\ 3\Sigma_{12} - 2\Sigma_{11} - \Sigma_{22} & \Sigma_{22} - 2\Sigma_{12} + \Sigma_{11} \end{pmatrix} \quad (64)\\
&= \frac{1}{q^2}\begin{pmatrix} \Sigma_{22} & 3\Sigma_{12} - 2\Sigma_{11} - \Sigma_{22} \\ 3\Sigma_{12} - 2\Sigma_{11} - \Sigma_{22} & \Sigma_{22} - 2\Sigma_{12} + \Sigma_{11} \end{pmatrix} \quad (65)
\end{aligned}
$$

where we make use of the fact that since $T_i^2 = T_i$, $\Sigma_{11} = \mathbb{E}\left[m_{i1}^2\right] = \mathbb{E}\left[m_{i1}m_{i2}\right] = \Sigma_{12}$, where the expectations are evaluated at the true γ^*. To compute Σ_{22}, we make use of the following facts:

$$
\begin{aligned}
V_i^2\left(Y_i - \theta T_i - \alpha_0\right)^2 &= V_i^2\left(Y_i^2 + \theta^2 T_i + \alpha_0^2 - 2\theta T_i Y_i - 2\alpha_0 Y_i + 2\alpha_0\theta T_i\right) && (66)\\
\mathbb{E}\left[V_i^2 Y_i^2\right] &= \mathbb{E}\left[T_i Y_i(1)^2 + (1-T_i)Y_i(0)^2 W_i^2\right] && (67)\\
&= \mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)}\right] + \mathbb{E}\left[\mu_1(X_i)^2 p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] && (68)\\
&\equiv q^2 A + \mathbb{E}\left[\mu_1(X_i)^2 p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] && (69)\\
\theta^2\mathbb{E}\left[V_i^2 T_i\right] &= \theta^2 q && (70)\\
\alpha_0^2\mathbb{E}\left[V_i^2\right] &= \alpha_0^2\mathbb{E}\left[T_i + (1-T_i)W_i^2\right] = \alpha_0^2 q + \alpha_0^2\mathbb{E}\left[\frac{p(X_i)^2}{1-p(X_i)}\right] = \alpha_0^2\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right] && (71)\\
-2\theta\mathbb{E}\left[V_i^2 T_i Y_i\right] &= -2\theta\mathbb{E}\left[\mu_1(X_i)p(X_i)\right] = -2\theta\alpha_1 q && (72)\\
-2\alpha_0\mathbb{E}\left[V_i^2 Y_i\right] &= -2\alpha_0\mathbb{E}\left[\mu_1(X_i)p(X_i)\right] - 2\alpha_0\mathbb{E}\left[\mu_0(X_i)\frac{p(X_i)^2}{1-p(X_i)}\right] && (73)\\
&= -2\alpha_0\alpha_1 q - 2\alpha_0\mathbb{E}\left[\mu_0(X_i)\frac{p(X_i)}{1-p(X_i)}\right] + 2\alpha_0\mathbb{E}\left[\mu_0(X_i)p(X_i)\right] && (74)\\
&= -2\alpha_0\mathbb{E}\left[\mu_0(X_i)\frac{p(X_i)}{1-p(X_i)}\right] - 2\alpha_0\theta q && (75)\\
2\alpha_0\theta\mathbb{E}\left[V_i^2 T_i\right] &= 2\alpha_0\theta q && (76)
\end{aligned}
$$

Adding these terms together yields

$$
\begin{aligned}
\Sigma_{22} &= q^2 A + \mathbb{E}\left[\mu_1(X_i)^2 p(X_i) + \mu_0(X_i)^2\frac{p(X_i)^2}{1-p(X_i)}\right] + \theta^2 q + \alpha_0^2\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right] - 2\alpha_1\theta q - 2\alpha_0\mathbb{E}\left[\mu_0(X_i)\frac{p(X_i)}{1-p(X_i)}\right] && (77)\\
&= q^2 A + \mathbb{E}\left[\mu_1(X_i)^2 p(X_i) - \mu_0(X_i)^2 p(X_i)\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] + \theta^2 q - 2\alpha_1\theta q && (78)\\
&= q^2 A + \mathbb{E}\left[\mu_1(X_i)^2 p(X_i) - \mu_0(X_i)^2 p(X_i)\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] && (79)\\
&+ \left(\alpha_1^2 - 2\alpha_1\alpha_0 + \alpha_0^2\right)q - 2\alpha_1^2 q + 2\alpha_1\alpha_0 q && (80)\\
&= q^2 A + \mathbb{E}\left[\mu_1(X_i)^2 p(X_i) - \mu_0(X_i)^2 p(X_i)\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] - \alpha_1^2 q + \alpha_0^2 q && (81)\\
&= q^2 A + \mathbb{E}\left[\left(\mu_1(X_i) - \alpha_1\right)^2 p(X_i)\right] - \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2 p(X_i)\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] && (82)\\
&= q^2 A + qV\left[\mu_1(X_i)|T_i = 1\right] - qV\left[\mu_0(X_i)|T_i = 1\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] && (83)\\
&= q^2\Omega_\theta - qV\left[\tau(X_i)|T_i = 1\right] + qV\left[\mu_1(X_i)|T_i = 1\right] - qV\left[\mu_0(X_i)|T_i = 1\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] && (84)\\
&= q^2\Omega_\theta + 2q\mathbb{C}\left[\tau(X_i), \mu_0(X_i)|T_i = 1\right] + \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] && (85)
\end{aligned}
$$

Although we do not need it here, a similar analysis shows that

$$\Sigma_{11} = \mathbb{E}\left[\sigma_1^2(X_i)p(X_i)\right] + \mathbb{E}\left[p(X_i)\left(\mu_1(X_i) - \alpha_1\right)^2\right] \qquad (86)$$

These results mean that the asymptotic variance of $\widehat{\theta}_{N,tw}$ is given by

$$AV(\widehat{\theta}_{N,tw}) = \Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)|T_i = 1\right] \qquad (87)$$

5 The Normalized Estimated Weights Estimator with Parametric Logit, $\widehat{\theta}_{N,pw}$

Let $\gamma = (\theta, \alpha_0, \pi')'$. The moment function and its derivative matrix are given by

$$m_i(\gamma) = \begin{pmatrix} V_i(Y_i - \theta T_i - \alpha_0)T_i \\ V_i(Y_i - \theta T_i - \alpha_0) \\ (T_i - \Lambda_i)Z_i \end{pmatrix} \quad \text{and} \quad M_i(\gamma) = -\begin{pmatrix} V_i T_i & V_i T_i & 0 \\ V_i T_i & V_i & -\left(Y_i(0) - \alpha_0\right)(1-T_i)\frac{\Lambda_i}{1-\Lambda_i}Z_i \\ 0 & 0 & \Lambda_i(1-\Lambda_i)Z_i Z_i' \end{pmatrix} \qquad (88)$$

where as before $V_i = T_i + (1-T_i)W_i$. Evaluated at γ^* and assuming no misspecification of the propensity score, the expectation of the derivative and the variance of the moments are given by

$$M = -q\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & -\frac{1}{q}f_0' \\ 0 & 0 & \frac{1}{q}\mathcal{I} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11} & g_1' \\ \Sigma_{11} & q^2 AV(\widehat{\theta}_{N,tw}) & h' \\ g_1 & h & \mathcal{I} \end{pmatrix} \qquad (89)$$

where for $t \in \{0,1\}$ we define

$$f_t = \mathbb{E}\left[p(X_i)\left(\mu_t(X_i) - \alpha_t\right) Z_i\right] \tag{90}$$
$$g_t = \mathbb{E}\left[p(X_i)(1-p(X_i))(\mu_t(X_i) - \alpha_t) Z_i\right] \tag{91}$$
$$g_1 + g_0 - f_0 = h \tag{92}$$

With this notation, note that $g_1 - g_0 = \mathbb{E}\left[p(X_i)(1-p(X_i))(\tau(X_i) - \theta) Z_i\right]$. The results on variance are calculated as

$$\begin{aligned}
\Sigma_{13} &= \mathbb{E}\left[V_i\left(Y_i - \theta T_i - \alpha_0\right) T_i\left(T_i - \Lambda_i\right) Z_i\right] \equiv \mathbb{E}\left[V_i \varepsilon_i T_i\left(T_i - \Lambda_i\right) Z_i\right] = \mathbb{E}\left[T_i \varepsilon_i\left(1 - \Lambda_i\right) Z_i\right] && (93)\\
&= \mathbb{E}\left[T_i\left(T_i Y_i(1) + (1-T_i) Y_i(0) - \theta T_i - \alpha_0\right)(1-\Lambda_i) Z_i\right] = \mathbb{E}\left[\left(T_i Y_i(1) - \theta T_i - \alpha_0 T_i\right)(1-\Lambda_i) Z_i\right] && (94)\\
&= \mathbb{E}\left[T_i(1-\Lambda_i)(Y_i(1) - (\theta + \alpha_0)) Z_i\right] && (95)\\
&= \mathbb{E}\left[p(X_i)(1-p(X_i))\left(\mu_1(X_i) - \alpha_1\right) Z_i\right] && (96)\\
&\equiv g_1 && (97)\\
\Sigma_{23} &= \mathbb{E}\left[V_i\left(Y_i - \theta T_i - \alpha_0\right)\left(T_i - \Lambda_i\right) Z_i\right] = \mathbb{E}\left[T_i \varepsilon_i\left(1 - \Lambda_i\right) Z_i\right] + \mathbb{E}\left[(1-T_i) W_i \varepsilon_i\left(T_i - \Lambda_i\right) Z_i\right] && (98)\\
&= \Sigma_{13} + \mathbb{E}\left[(1-T_i) W_i \varepsilon_i\left(T_i - \Lambda_i\right) Z_i\right] = \Sigma_{13} - \mathbb{E}\left[(1-T_i) W_i \varepsilon_i \Lambda_i Z_i\right] && (99)\\
&= \Sigma_{13} - \mathbb{E}\left[(1-T_i) W_i\left(T_i Y_i(1) + (1-T_i) Y_i(0) - \theta T_i - \alpha_0\right)\Lambda_i Z_i\right] = \Sigma_{13} - \mathbb{E}\left[(1-T_i) W_i\left(Y_i(0) - \alpha_0\right)\Lambda_i Z_i\right] && (100)\\
&= \Sigma_{13} - \mathbb{E}\left[p(X_i)^2\left(\mu_0(X_i) - \alpha_0\right) Z_i\right] && (101)\\
&= \mathbb{E}\left[p(X_i)(1-p(X_i))\left(\mu_1(X_i) - \alpha_1\right) Z_i\right] + \mathbb{E}\left[p(X_i)(1-p(X_i))\left(\mu_0(X_i) - \alpha_0\right) Z_i\right] - f_0 && (102)\\
&\equiv h = g_1 + g_0 - f_0 && (103)
\end{aligned}$$

The asymptotic variance of $\widehat{\gamma}$ is thus given by

$$\begin{aligned}
M^{-1}\Sigma M'^{-1} &= \frac{1}{q^2}\begin{pmatrix} 2 & -1 & -f_0'\mathcal{I}^{-1} \\ -1 & 1 & f_0'\mathcal{I}^{-1} \\ 0 & 0 & q\mathcal{I}^{-1} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{11} & g_1' \\ \Sigma_{11} & q^2 AV(\widehat{\theta}_{N,tw}) & h' \\ g_1 & h & \mathcal{I} \end{pmatrix}\begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ -\mathcal{I}^{-1}f_0 & \mathcal{I}^{-1}f_0 & q\mathcal{I}^{-1} \end{pmatrix} && (104)\\
&= \frac{1}{q^2}\begin{pmatrix} 2 & -1 & -f_0'\mathcal{I}^{-1} \\ -1 & 1 & f_0'\mathcal{I}^{-1} \\ 0 & 0 & q\mathcal{I}^{-1} \end{pmatrix}\begin{pmatrix} \Sigma_{11} - g_1'\mathcal{I}^{-1}f_0 & g_1'\mathcal{I}^{-1}f_0 & qg_1'\mathcal{I}^{-1} \\ 2\Sigma_{11} - q^2 AV(\widehat{\theta}_{N,tw}) - h'\mathcal{I}^{-1}f_0 & -\Sigma_{11} + q^2 AV(\widehat{\theta}_{N,tw}) + h'\mathcal{I}^{-1}f_0 & qh'\mathcal{I}^{-1} \\ 2g_1 - h - f_0 & -g_1 + h + f_0 & q\mathrm{I} \end{pmatrix}
\end{aligned}$$

which shows that

$$\begin{aligned}
AV(\widehat{\theta}_{N,pw}) &= \frac{1}{q^2}\left\{2\left(\Sigma_{11} - g_1'\mathcal{I}^{-1}f_0\right) - \left(2\Sigma_{11} - q^2 AV(\widehat{\theta}_{N,tw}) - h'\mathcal{I}^{-1}f_0\right) - f_0'\mathcal{I}^{-1}\left(2g_1 - h - f_0\right)\right\} && (105)\\
&= AV(\widehat{\theta}_{N,tw}) + \frac{1}{q^2}\left(2h - 4g_1 + f_0\right)'\mathcal{I}^{-1}f_0 && (106)\\
&= AV(\widehat{\theta}_{N,tw}) - \frac{1}{q^2}f_0'\mathcal{I}^{-1}f_0 - 2\frac{1}{q^2}(g_1 - g_0)'\mathcal{I}^{-1}f_0 && (107)\\
&= \Omega_\theta && (108)\\
&+ \frac{1}{q^2}\mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] - \frac{1}{q^2}f_0'\mathcal{I}^{-1}f_0 && (109)\\
&+ 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i) | T_i = 1\right] - 2\frac{1}{q^2}\left(g_1 - g_0\right)'\mathcal{I}^{-1}f_0 && (110)
\end{aligned}$$

6 The Normalized Estimated Weights Estimator with Fully Saturated Logit, $\widehat{\theta}_{N,ew}$

Now suppose that the covariates X_i are discrete, as before. Then we have

$$\begin{aligned}
f_0'\mathcal{I}^{-1}f_0 &= \sum_j \frac{\eta_j^2 p(x_j)^2\left(\mu_0(x_j) - \alpha_0\right)^2}{\eta_j p(x_j)(1-p(x_j))} = \sum_j \eta_j\left(\mu_0(x_j) - \alpha_0\right)^2\frac{p(x_j)}{1-p(x_j)} && (111)\\
&= \mathbb{E}\left[\left(\mu_0(X_i) - \alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] && (112)\\
(g_1 - g_0)'\mathcal{I}^{-1}f_0 &= \sum_j \frac{\eta_j^2 p(x_j)^2(1-p(x_j))\left(\tau(x_j) - \theta\right)\left(\mu_0(x_j) - \alpha_0\right)}{\eta_j p(x_j)(1-p(x_j))} && (113)\\
&= \sum_j \eta_j p(x_j)\left(\tau(x_j) - \theta\right)\left(\mu_0(x_j) - \alpha_0\right) = \mathbb{E}\left[p(X_i)\left(\tau(X_i) - \theta\right)\left(\mu_0(X_i) - \alpha_0\right)\right] && (114)\\
&= q\mathbb{C}\left[\tau(X_i), \mu_0(X_i) | T_i = 1\right] && (115)
\end{aligned}$$

Because of these results, the terms in equations (109) and (110) both cancel out, and the asymptotic variance of $\widehat{\theta}_{N,ew}$ is simply

$$AV(\widehat{\theta}_{N,ew}) = \Omega_\theta \tag{116}$$

7 GPE Reweighting with Parametric Logit, $\widehat{\theta}_{GPE,pw}$

Define $\gamma = (\theta, \alpha_0, \pi')'$. The moment function and its derivative matrix are given by

$$m_i(\gamma) = \begin{pmatrix} V_i(Y_i - \theta T_i - \alpha_0)T_i \\ V_i(Y_i - \theta T_i - \alpha_0) \\ \left(T_i - (1-T_i)\frac{\Lambda_i}{1-\Lambda_i}\right) Z_i \end{pmatrix} \quad \text{and} \quad M_i(\gamma) = -\begin{pmatrix} V_iT_i & V_iT_i & 0 \\ V_iT_i & V_i & -\left(Y_i(0)-\alpha_0\right)(1-T_i)\frac{\Lambda_i}{1-\Lambda_i}Z_i' \\ 0 & 0 & (1-T_i)\frac{\Lambda_i}{1-\Lambda_i}Z_iZ_i' \end{pmatrix} \tag{117}$$

where as before $V_i = T_i + (1-T_i)W_i$. Evaluated at γ^* and assuming no misspecification of the propensity score, the expectation of the derivative and the variance of the moments are given by

$$M = -q\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & -\frac{1}{q}f_0' \\ 0 & 0 & \frac{1}{q}B \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11} & f_1' \\ \Sigma_{11} & q^2 AV(\widehat{\theta}_{N,tw}) & (f_1 + f_0 - r)' \\ f_1 & f_1 + f_0 - r_0 & C \end{pmatrix} \tag{118}$$

where $B = \mathbb{E}[p(X_i)Z_iZ_i']$, $C = \mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}Z_iZ_i'\right]$, and $r_0 = \mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}(\mu_0(X_i)-\alpha_0)Z_i\right]$. The results on variance are calculated as

$$\begin{aligned}
\Sigma_{13} &= \mathbb{E}\left[V_i(Y_i - \theta T_i - \alpha_0)T_i\left(T_i - (1-T_i)\frac{\Lambda_i}{1-\Lambda_i}\right)Z_i\right] && (119)\\
&= \mathbb{E}\left[V_iT_i(Y_i - \theta T_i - \alpha_0)Z_i\right] = \mathbb{E}\left[T_i(Y_i(1) - \alpha_1)Z_i\right] = \mathbb{E}\left[p(X_i)(\mu_1(X_i) - \alpha_1)Z_i\right] = f_1 && (120)\\
\Sigma_{23} &= \mathbb{E}\left[V_i(Y_i - \theta T_i - \alpha_0)\left(T_i - (1-T_i)\frac{\Lambda_i}{1-\Lambda_i}\right)Z_i\right] && (121)\\
&= \mathbb{E}\left[V_iT_i(Y_i - \theta T_i - \alpha_0)Z_i\right] - \mathbb{E}\left[V_i(1-T_i)\frac{\Lambda_i}{1-\Lambda_i}(Y_i - \theta T_i - \alpha_0)Z_i\right] && (122)\\
&= f_1 - \mathbb{E}\left[(1-T_i)\frac{\Lambda_i^2}{(1-\Lambda_i)^2}(Y_i(0) - \alpha_0)Z_i\right] = f_1 - \mathbb{E}\left[\frac{p(X_i)^2}{1-p(X_i)}(\mu_0(X_i) - \alpha_0)Z_i\right] && (123)\\
&= f_1 - \mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}(\mu_0(X_i) - \alpha_0)Z_i\right] + \mathbb{E}\left[p(X_i)(\mu_0(X_i) - \alpha_0)Z_i\right] && (124)\\
&= f_1 + f_0 - \mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}(\mu_0(X_i) - \alpha_0)Z_i\right] \equiv f_1 + f_0 - r_0 && (125)
\end{aligned}$$

where we use the fact that $V_iT_i = T_i$, $V_i(1-T_i) = (1-T_i)\Lambda_i/(1-\Lambda_i)$, and $\theta + \alpha_0 = \alpha_1$.

The asymptotic variance of $\widehat{\gamma}$ is thus given by $M^{-1}\Sigma M'^{-1}$, or

$$\frac{1}{q^2}\begin{pmatrix} 2 & -1 & -f_0'B^{-1} \\ -1 & 1 & f_0'B^{-1} \\ 0 & 0 & qB^{-1} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{11} & f_1' \\ \Sigma_{11} & q^2 AV(\widehat{\theta}_{N,tw}) & (f_1 + f_0 - r_0)' \\ f_1 & f_1 + f_0 - r_0 & C \end{pmatrix}\begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ -B^{-1}f_0 & B^{-1}f_0 & qB^{-1} \end{pmatrix} \tag{126}$$

$$= \frac{1}{q^2}\begin{pmatrix} 2 & -1 & -f_0'B^{-1} \\ -1 & 1 & f_0'B^{-1} \\ 0 & 0 & qB^{-1} \end{pmatrix} \tag{127}$$

$$\times \begin{pmatrix} \Sigma_{11} - f_1'B^{-1}f_0 & f_1'B^{-1}f_0 & qf_1'B^{-1} \\ 2\Sigma_{11} - q^2AV(\widehat{\theta}_{N,tw}) - (f_1 + f_0 - r_0)'B^{-1}f_0 & -\Sigma_{11} + q^2AV(\widehat{\theta}_{N,tw}) + (f_1 + f_0 - r_0)'B^{-1}f_0 & q(f_1 + f_0 - r_0)'B^{-1} \\ f_1 - f_0 + r_0 - CB^{-1}f_0 & f_0 - r_0 + CB^{-1}f_0 & qCB^{-1} \end{pmatrix}$$

and the asymptotic variance of $\widehat{\theta}_{GPE,pw}$ is then

$$\begin{aligned}
AV(\widehat{\theta}_{GPE,pw}) &= \frac{1}{q^2}\left\{2\left(\Sigma_{11} - f_1'B^{-1}f_0\right) - \left(2\Sigma_{11} - q^2AV(\widehat{\theta}_{N,tw}) - (f_1 + f_0 - r_0)'B^{-1}f_0\right) - f_0'B^{-1}\left(f_1 - f_0 + r_0 - CB^{-1}f_0\right)\right\} \\
&= AV(\widehat{\theta}_{N,tw}) + \frac{1}{q^2}\left\{-2(f_1 - f_0)'B^{-1}f_0 - 2r_0'B^{-1}f_0 + f_0'B^{-1}CB^{-1}f_0\right\} && (128)\\
&= \Omega_\theta && (129)\\
&+ \frac{1}{q^2}\mathbb{E}\left[(\mu_0(X_i) - \alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] + \frac{1}{q^2}f_0'B^{-1}CB^{-1}f_0 - 2\frac{1}{q^2}r_0'B^{-1}f_0 && (130)\\
&+ \frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i) | T_i = 1\right] - 2\frac{1}{q^2}(f_1 - f_0)'B^{-1}f_0 && (131)
\end{aligned}$$

8 GPE Reweighting with Fully Saturated Model, $\widehat{\theta}_{GPE,ew}$

Now suppose that the covariates X_i are discrete, as before. Then we have

$$B^{-1}CB^{-1} = \text{diag}\left\{\frac{1}{\eta_j p(x_j)}\right\}\text{diag}\left\{\eta_j\frac{p(x_j)}{1-p(x_j)}\right\}\text{diag}\left\{\frac{1}{\eta_j p(x_j)}\right\} = \text{diag}\left\{\frac{1}{\eta_j p(x_j)(1-p(x_j))}\right\} \quad (132)$$

$$f_t'B^{-1}f_0 = \sum_j \frac{\eta_j^2 p(x_j)^2(\mu_t(x_j)-\alpha_t)(\mu_0(x_j)-\alpha_0)}{\eta_j p(x_j)} = \sum_j \eta_j p(x_j)(\mu_t(x_j)-\alpha_t)(\mu_0(x_j)-\alpha_0) \quad (133)$$

$$= \mathbb{E}\left[p(X_i)(\mu_t(X_i)-\alpha_t)(\mu_0(X_i)-\alpha_0)\right] = q\mathbb{C}\left[\mu_t(X_i),\mu_0(X_i)|T_i=1\right] \quad (134)$$

$$f_0'B^{-1}CB^{-1}f_0 = \sum_j \frac{\eta_j^2 p(x_j)^2\left(\mu_0(x_j)-\alpha_0\right)^2}{\eta_j p(x_j)(1-p(x_j))} = \sum_j \eta_j\frac{p(x_j)}{1-p(x_j)}\left(\mu_0(x_j)-\alpha_0\right)^2 \quad (135)$$

$$= \mathbb{E}\left[\left(\mu_0(X_i)-\alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] \quad (136)$$

$$r_0'B^{-1}f_0 = \sum_j \frac{\eta_j^2(\mu_0(x_j)-\alpha_0)^2 p(x_j)^2/(1-p(x_j))}{\eta_j p(x_j)} = \sum_j \eta_j\frac{p(x_j)}{1-p(x_j)}(\mu_0(x_j)-\alpha_0)^2 \quad (137)$$

$$= \mathbb{E}\left[\left(\mu_0(X_i)-\alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] \quad (138)$$

so that

$$-\frac{1}{q^2}2(f_1-f_0)'B^{-1}f_0 = -\frac{1}{q}\mathbb{C}\left[\tau(X_i),\mu_0(X_i)|T_i=1\right] \quad (139)$$

$$f_0'B^{-1}CB^{-1}f_0 - 2r_0B^{-1}f_0 = -\mathbb{E}\left[\left(\mu_0(X_i)-\alpha_0\right)^2\frac{p(X_i)}{1-p(X_i)}\right] \quad (140)$$

Because of these results, the terms in equations (130) and (131) cancel out, and the asymptotic variance of $\widehat{\theta}_{GPE,ew}$ is simply

$$AV(\widehat{\theta}_{GPE,ew}) = \Omega_\theta \quad (141)$$

9 Matching on a Scalar Covariate

As shown in Abadie and Imbens (2006), nearest neighbor matching on a scalar covariate has variance conditional on $\mathbf{X}$ and $\mathbf{T}$ of

$$\mathbb{V}\left[\widetilde{\theta}|\mathbf{X},\mathbf{T}\right] = \frac{1}{n_1^2}\sum_i\left(T_i-(1-T_i)\frac{K_k(i)}{k}\right)^2\sigma_{T_i}^2(X_i) \quad (142)$$

$$= \frac{1}{n_1^2}\sum_i\left(T_i\sigma_1^2(X_i)+(1-T_i)\sigma_0^2(X_i)\frac{K_k(i)^2}{k^2}\right) \quad (143)$$

where $\mathbf{X}$ is the matrix with ith row X_i', $\mathbf{T}$ is the vector with ith row T_i, $K_k(i)$ is the number of times unit i is matched given that k matches per unit are used, and n_1 is the number of treated units. Since n_1/n converges almost surely to q, we have

$$q^2 n\mathbb{V}\left[\widetilde{\theta}|\mathbf{X},\mathbf{T}\right] = \frac{1}{n}\sum_i\left(T_i\sigma_1^2(X_i)+(1-T_i)\sigma_0^2(X_i)\frac{K_k(i)^2}{k^2}\right)+o_p(1) \quad (144)$$

The (marginal) asymptotic variance of $\widetilde{\theta}$ is the expectation of $n\mathbb{V}\left[\widetilde{\theta}|\mathbf{X},\mathbf{T}\right]$ over $\mathbf{X}$ and $\mathbf{T}$, plus the variance of the conditional expectation. To compute the expectation of $n\mathbb{V}\left[\widetilde{\theta}|\mathbf{X},\mathbf{T}\right]$, we use the fact that

$$\mathbb{E}\left[K_k(i)^2|T_i=0,X_i=x\right] = k\frac{p(x)}{1-p(x)}+\frac{k(2k+1)}{2}\left(\frac{p(x)}{1-p(x)}\right)^2+o(1) \quad (145)$$

as shown by Abadie and Imbens (2006) in their supplemental proofs. Then, following the logic described there, we have

$$q^2AV\left(\widetilde{\theta}\right) = q^2\mathbb{E}\left[n\mathbb{V}\left[\widetilde{\theta}|\mathbf{X},\mathbf{T}\right]\right]+\mathbb{E}\left[(\tau(X_i)-\theta)^2p(X_i)\right] \quad (146)$$

$$= \mathbb{E}\left[T_i\sigma_1^2(X_i)+(1-T_i)\sigma_0^2(X_i)\frac{K_k(i)^2}{k^2}\right]+\mathbb{E}\left[(\tau(X_i)-\theta)^2p(X_i)\right] \quad (147)$$

$$= \mathbb{E}\left[p(X_i)\sigma_1^2(X_i)+(1-p(X_i))\sigma_0^2(X_i)\frac{1}{k^2}\left(k\frac{p(X_i)}{1-p(X_i)}+\frac{k(2k+1)}{2}\left(\frac{p(X_i)}{1-p(X_i)}\right)^2\right)\right] \quad (148)$$

$$+\mathbb{E}\left[(\tau(X_i)-\theta)^2p(X_i)\right] \quad (149)$$

$$= \mathbb{E}\left[p(X_i)\sigma_1^2(X_i)+\frac{1}{k}\sigma_0^2(X_i)\left(p(X_i)+\frac{2k+1}{2}\frac{p(X_i)^2}{1-p(X_i)}\right)\right]+\mathbb{E}\left[(\tau(X_i)-\theta)^2p(X_i)\right] \quad (150)$$

$$= \mathbb{E}\left[p(X_i)\sigma_1^2(X_i)+\frac{p(X_i)^2}{1-p(X_i)}\sigma_0^2(X_i)\right]+\frac{1}{2k}\mathbb{E}\left[\sigma_0^2(X_i)\left(\frac{p(X_i)^2}{1-p(X_i)}+2p(X_i)\right)\right]+\mathbb{E}\left[(\tau(X_i)-\theta)^2p(X_i)\right] \quad (151)$$

$$= q^2\Omega_\theta+\frac{1}{2k}\mathbb{E}\left[\sigma_0^2(X_i)\left(\frac{p(X_i)}{1-p(X_i)}+p(X_i)\right)\right] \quad (152)$$

10 Puzzle # 1 of Section III: Normalized Reweighting and Nearest Neighbor Matching

Recall that the Frölich (2004) DGPs involve matching on a single continuous covariate in DGPs with homogenous treatment effects and a homoscedastic outcome equation error term with variance σ^2. For this special case, we have

$$\Omega_\theta = \frac{1}{q^2}\sigma^2\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right] \tag{153}$$

and the asymptotic variance of normalized reweighting under homogenous treatment effects and homoskedasticity is

$$AV\left(\widehat{\theta}_{N,pw}\right) = \widetilde{a} + \Omega_\theta = \widetilde{a} + \widetilde{b}\sigma^2 \tag{154}$$

where
$$\widetilde{a} = \frac{1}{q^2}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] - \frac{1}{q^2}\mathbb{E}\left[p(X_i)\left(\mu_0(X_i)-\alpha_0\right)Z_i'\right]\mathcal{I}^{-1}\mathbb{E}\left[p(X_i)\left(\mu_0(X_i)-\alpha_0\right)Z_i\right] \tag{155}$$

and $\widetilde{b} = \frac{1}{q^2}\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right]$. The asymptotic variance of nearest neighbor matching under the same conditions is given by

$$AV\left(\widetilde{\theta}\right) = \Omega_\theta\left(1+\frac{1}{2k}\right) + \frac{1}{2k}\frac{\sigma^2}{q} = \left\{\widetilde{b}\left(1+\frac{1}{2k}\right)+\frac{1}{2k}\frac{1}{q}\right\}\sigma^2 \tag{156}$$

Thus, under homoskedasticity and constant treatment effects, we have

$$AV\left(\widetilde{\theta}\right) - AV\left(\widehat{\theta}_N\right) = \frac{1}{2k}\left(\widetilde{b}+\frac{1}{q}\right)\sigma^2 - \widetilde{a} \tag{157}$$

This implies that under homoskedasticity and constant treatment effects nearest neighbor matching on a scalar covariate has a larger asymptotic variance than normalized reweighting if and only if

$$\sigma^2 > 2k\,\widetilde{a}/(\widetilde{b}+1/q) \tag{158}$$

$$= 2k\frac{\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] - \mathbb{E}[p(X_i)\left(\mu_0(X_i)-\alpha_0\right)Z_i']\mathcal{I}^{-1}\mathbb{E}[p(X_i)\left(\mu_0(X_i)-\alpha_0\right)Z_i]}{\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}+p(X_i)\right]} \tag{159}$$

For pair matching ($k=1$), this condition is met for all 30 Frölich DGPs for $\sigma^2 = 0.1$, but is only met for some of these DGPs when $\sigma^2 = 0.01$.

11 Puzzle # 2 of Section III: Normalized and Unnormalized Reweighting

From the work above, we know that the asymptotic variance of normalized reweighting is given by

$$AV\left(\widehat{\theta}_N\right) = \Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i),\mu_0(X_i)|T_i=1\right] \tag{160}$$

$$- \frac{1}{q^2}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]'\mathcal{I}^{-1}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right] \tag{161}$$

$$- 2\frac{1}{q^2}\mathbb{E}\left[p(X_i)(1-p(X_i))(\tau(X_i)-\theta)Z_i\right]\mathcal{I}^{-1}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)p(X_i)Z_i\right] \tag{162}$$

where $\Omega_\theta = \frac{1}{q^2}\mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)} + (\tau(X_i)-\theta)^2p(X_i)\right]$, Z_i is the vector of predictors for the logit model, $\tau(X_i) = \mu_1(X_i) - \mu_0(X_i)$, $\alpha_0 = \mathbb{E}\left[\mu_0(X_i)|T_i=1\right]$, and $\mathcal{I} = \mathbb{E}\left[p(X_i)(1-p(X_i))Z_iZ_i'\right]$ is the information matrix for the logit model. The asymptotic variance of unnormalized reweighting has precisely the form given in equations (160), (161), and (162), but with zero replacing α_0.

This variance expression is somewhat complex. We next briefly discuss each of the five terms. The first term, Ω_θ, is common to reweighting and matching estimators and is a particular type of efficiency bound first derived by Hahn (1998) for this problem and hence non-negative. In typical empirical DGPs, this term is the largest in magnitude of the five terms. The second term is also non-negative, but can be zero for normalized reweighting when $\mu_0(x)$ is constant in x, because then $\mu_0(x) = \alpha_0$ for every x in the support of X_i. For unnormalized reweighting, the second term can only be zero if $\mu_0(x)$ is zero for every x in the support of X_i. The third term can either be positive or negative and pertains to heterogeneity in treatment effects. It is zero when the treatment effect is homogenous. Reweighting using the known propensity score, rather than the estimated propensity score, has an asymptotic variance given by these first three terms. The fourth and fifth terms at least partially offset the second and third terms, respectively. As noted above, in some specialized circumstances, they exactly offset the second and third terms, in which case the asymptotic variance of reweighting is merely Ω_θ.

Having briefly discussed the terms in this variance expression, we now return resolving the puzzle that normalized reweighting exhibits much lower variance in the Frölich (2004) simulations than unnormalized reweighting. This result has little to do with the third and fifth terms, since the Frölich DGPs have homogenous treatment effects, leading both terms to be zero. A potential explanation for this pattern is the second term. In particular, the second term for normalized reweighting, $\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2p(X_i)/(1-p(X_i))\right]$, will often be smaller than the second term for unnormalized reweighting, $\mathbb{E}\left[\mu_0(X_i)^2p(X_i)/(1-p(X_i))\right]$. For example, in the context of the 30 Frölich DGPs, the latter is 6 to 42 times as large as the former.

However, since we are studying the performance of reweighting using an *estimated* propensity score rather than the true propensity score, the fourth term factors into the asymptotic variance as well, and the fourth term typically compensates for the second term. That is, the relevant quantity for the difference in the overall asymptotic variances is the difference in the *sum* of the second and fourth terms. In particular, for the special case of homogenous treatment effects, we have

$$q^2\left(AV\left(\widehat{\theta}_U\right) - AV\left(\widehat{\theta}_N\right)\right) \tag{163}$$

$$= \left(\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] - \mathbb{E}\left[p(X_i)\mu_0(X_i)Z_i\right]'\mathcal{I}^{-1}\mathbb{E}\left[p(X_i)\mu_0(X_i)Z_i\right]\right) \tag{164}$$

$$-\left(\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] - \mathbb{E}\left[p(X_i)\left(\mu_0(X_i)-\alpha_0\right)Z_i\right]'\mathcal{I}^{-1}\mathbb{E}\left[p(X_i)\left(\mu_0(X_i)-\alpha_0\right)Z_i\right]\right) \tag{165}$$

We computed this difference for the 30 Frölich designs. In 20 cases, we find that while the asymptotic variance of normalized reweighting is below that of unnormalized reweighting. In 10 cases, the pattern is the opposite. This holds for both $\sigma^2 = 0.01$ and $\sigma^2 = 0.10$.

Part IC: Finite Semiparametric Efficiency Bound

As noted in Chamberlain (1986), $\sqrt{n}$-consistent semiparametric estimators for a given DGP exist if and only if the semiparametric efficiency bound (SEB) is finite. Since the SEB is the supremum of parametric efficiency bounds over a class of parametric models, establishing that the SEB is finite also establishes that the parametric efficiency bound is finite. We next show the finiteness of the SEB established by Hahn (1998) for TOT, assuming an unknown propensity score, for the special case of the DGPs studied in the main text. The SEB for this case is given by

$$\Omega_\theta = \frac{1}{q^2}\mathbb{E}\left[\sigma_1^2(X_i)p(X_i)\right] + \frac{1}{q^2}\mathbb{E}\left[\sigma_0^2(X_i)p(X_i)^2/(1-p(X_i))\right] + \frac{1}{q^2}\mathbb{E}\left[(\tau(X_i)-\theta)^2\,p(X_i)\right] \tag{166}$$

Consider first the case of the Frölich (2004) DGPs. These DGPs have homoskedastic errors and homogenous treatment effects, which means that

$$\Omega_\theta = \frac{\sigma^2}{q^2}\mathbb{E}\left[p(X_i)\right] + \frac{\sigma^2}{q^2}\mathbb{E}\left[p(X_i)^2/(1-p(X_i))\right] = \frac{\sigma^2}{q^2}\mathbb{E}\left[p(X_i)\right] + \frac{\sigma^2}{q^2}\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)} - p(X_i)\right] \tag{167}$$

$$= \frac{\sigma^2}{q^2}\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right] \tag{168}$$

Next, note that when $\alpha+\beta<1$, strict overlap is satisfied and Ω_θ is immediately finite. However, when $\alpha+\beta=1$, we have the intermediary case where strict overlap is violated but overlap is satisfied. For such a case, we must show that $\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right]$ is finite. To study this case in detail, set $\alpha = 1-\beta$, and note that

$$\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right] = \mathbb{E}\left[\frac{1-\beta+\beta\Lambda(\sqrt{2}X_i)}{\beta-\beta\Lambda(\sqrt{2}X_i)}\right] = \mathbb{E}\left[\frac{1-\beta\left(1-\Lambda(\sqrt{2}X_i)\right)}{\beta\left(1-\Lambda(\sqrt{2}X_i)\right)}\right] \tag{169}$$

$$= \frac{1}{\beta}\mathbb{E}\left[\frac{1}{1-\Lambda(\sqrt{2}X_i)}\right] - 1 \tag{170}$$

$$= \frac{1}{\beta}\mathbb{E}\left[1+\exp(\sqrt{2}X_i)\right] - 1 \tag{171}$$

where we recall that X_i is distributed standard normal. This is then a standard calculation for the mean of the exponential of a normal with mean zero and variance 4, which is known to be finite. Consequently, the SEB for all 30 of the Frölich (2004) DGPs is finite.

Consider next the case of the NSW DGPs. There are two such DGPs. In the first such DGP, X_i is distributed discrete and for each such X_i, $0 < p(X_i) < 1$. Both DGPs have homoskedastic errors and homogenous treatment effects, so that we need only demonstrate that $\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right]$ is finite. For the first DGP with discrete covariates, this expectation is immediately finite since $0 < p(x) < 1$ for every x in the support of X_i, by virtue of the logit model used to model the probability of treatment. For the second DGP with continuous covariates, this expectation is a weighted average of the analogous conditional expectations, one for each group. Each group has

$$p(X_i) = \Lambda(c_0 + c_1 Y_i) \tag{172}$$

for different, but finite, values of c_0 and c_1, where Y_i is distributed normal (here, Y_i is a scalar which is a linear combination of normal covariates). For each such group, the conditional expectation is finite, since

$$\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}\right] = \mathbb{E}\left[\exp(c_0 + c_1 Y_i)\right] \tag{173}$$

which again is a standard calculation in normal theory for the mean of the exponential of a normal random variable.

Consider finally the case of the CPS DGPs. There are 7 such DGPs. Each such DGP is characterized by discrete covariates X_i with $0 < p(x) < 1$ for every x in the support of X_i. For these cases, Ω_θ is immediately finite.

Part ID: Pseudo-Random Number Generation

As noted in the text, we do all of our computations in Stata, version 11.0. Additionally, much of the code we use is based on the Stata's matrix programming sublanguage, Mata. Uniform pseudo-random variables were generated using the `uniform()` command, and Gaussian pseudo-random variables were generated using `invnorm(uniform())`, as opposed to the recommended `rnormal()` command. This last method is Stata's implementation of the Marsaglia sawtooth method (Knuth 1998). As of the time of this writing, an error in Stata's implementation of this method leads the sequence of signs of pseudo-random numbers generated by `rnormal()` to be nearly identical for different seeds (personal communications, William Gould and Owen Ozier). This is potentially problematic for our purposes, because we use multiple seeds (approximately 60) in order to take advantage of multiple processors. This embarassingly parallel approach is essential to keeping computing times under a month. The problem with signs in `rnormal()` was first brought to our attention by Owen Ozier. At the time, we had done all of our simulations using the `rnormal()` command. To ensure that our results were not sensitive to this software problem, we re-ran all of our computations. The simulation estimates of bias and variance quantities differed by minor amounts, with the final reported digit changing for only a few entries.

We note that Stata's `uniform()` random number generator passes Marsaglia's DIEHARD test sequence and that the numerical approximation errors associated with `invnorm`, which Marsaglia's sawtooth method would in theory avoid, are trivial except in the far tails.[1] As a partial test of the proposition that errors in the tails of `invnorm` are not worth worrying about in this context, we compared bias and variance estimates for normalized reweighting based on `invnorm(uniform())` with bias and variance estimates based on Marsaglia's polar method. This method generates a pair (U, V) distributed uniformly in the unit circle (by sequentially generating pairs distributed uniformly in the unit square centered at zero and discarding those falling outside of the interior of the unit circle), computes $S = U^2 + V^2$, and then generates a pair (X, Y) according to $X = U\sqrt{-\ln(S)/S}$ and $Y = V\sqrt{-\ln(S)/S}$. Constructed in this manner, X and Y are independent pseudo-random standard Gaussians, and no approximation to the Gaussian inverse distribution function is required. The estimated bias and variance quantities obtained using this method are nearly the same as those obtained using the simpler `invnorm(uniform())` method.

Part IE: Existence of Moments

We assume throughout the manuscript that each replication for a given estimator is drawn randomly with replacement from some distribution characterized by finite mean and variance. One implication of the assumption of finite mean and variance is that the variance of sample means of the estimator based on k replications should be proportional to k^{-1}. A simple way to informally examine this implication is as follows. First, randomly assign estimator replications to groups and subgroups so that for group k, there are m subgroups (where m is a fixed number chosen by the researcher), and there are k replications in each subgroup. That is, group 1 has m subgroups of size 1, group 2 has m subgroups of size 2, group 3 has m subgroups of size 3, and so on. Second, construct sample means of the estimator separately for each subgroup. For a given group k, this creates m sample means, each of which is based on k replications. Third, for each group construct the sample variance of those sample means. The sample variance for group k approximates the population variance for group k, which under a hypothesis of finite mean and variance is proportional to k^{-1}. This suggests the fourth step, which is to simply regress the sample variances on k^{-1}. The constant term can be suppressed, or it can be estimated in order to verify that it has little explanatory power.

Web Appendix Table 2 presents the R^2 from both types of regressions for the Frölich Design (Design 1, Curve 1), and for the NSW Design (empirical distribution). We choose these designs because for both strict overlap is violated, and it is in such a context that existence of moments is more likely to be a problem. These tests use $m = 128$.[2] Columns A and B correspond to regressions without and with a constant term, respectively.

To get a sense of what kinds of R^2 values are expected when neither the variance nor the mean exists, Web Appendix Table 3 presents the 5th and 95th percentiles of R^2 for Student's t distribution with degrees of freedom one through six, for both versions of the R^2 statistic. The percentiles reported hold fixed six particular groups of 10,000 replications (one for each degree of freedom); the percentiles are computed over 10,000 randomized permutations of the group-subgroup structure.

These results indicate that the high R^2 values in Web Appendix Table 2 are unlikely to occur for a distribution such as Student's t, unless the degrees of freedom is 3 (in which case 2 moments exist), 4 (in which case 3 moments exist), or higher. This suggests that the results in Web Appendix Table 2 are consistent with an assumption of two finite moments, but might also be inconsistent with third or fourth finite moments.

Another approach to this problem applies the method of moments to the same group-subgroup structure. This approach requires that we strengthen the null hypothesis to existence of the first four moments. To explain these ideas requires additional notation. For simplicity, we now write X_r for a replication of a given estimator. We have iid draws $X_1, X_2, \ldots, X_R$, with $\mathbb{E}[X_r] = \mu_0$, $\mathbb{V}[X_r] = \sigma_0^2$, $\mathbb{E}[(X_r - \mu_0)^3] = \widetilde{\gamma}_0$ and $\mathbb{E}[(X_r - \mu_0)^4] = \widetilde{\delta}_0$, where the null hypothesis is that $|\mu_0| < \infty$, $\sigma_0^2 < \infty$, $|\widetilde{\gamma}_0| < \infty$, and $\widetilde{\delta}_0 < \infty$. As before, we organize these replications as X_{ijk} with groups k, subgroups j within groups, and replications i within subgroups. Recall that by construction each subgroup j is comprised of k replications (see above for discussion). The sample means $\overline{X}_{jk} = \frac{1}{k}\sum_i X_{ijk}$ have centered moments $\mathbb{E}[(\overline{X}_{jk} - \mu_0)] = 0$, $\mathbb{E}[(\overline{X}_{jk} - \mu_0)^2] = \sigma_0^2/k$, $\mathbb{E}[(\overline{X}_{jk} - \mu_0)^3] \equiv \gamma_0$, and $\mathbb{E}[(\overline{X}_{jk} - \mu_0)^4] \equiv \delta_0$, where $|\gamma_0| < \infty$ and $\delta_0 < \infty$ since $|\widetilde{\gamma}_0| < \infty$ and $\widetilde{\delta}_0 < \infty$. For each group k, define

$$M_k = \frac{1}{m}\sum_{j=1}^{m} \overline{X}_{jk} \quad \text{and} \quad W_k^2 = \frac{1}{m}\sum_{j=1}^{m} (\overline{X}_{jk} - \mu_0)^2 \tag{174}$$

[1] See `http://www.stata.com/support/cert/diehard/index.html`.

[2] Since the number of replications is fixed, there is a tradeoff between the number of groups K and the number of subgroups within a group, or m. Ideally, both m and K will be large; if m is large then we obtain a good estimate of the variance of the sample means, and if K is large then we obtain a good estimate of the relationship between the sample variances and k^{-1}. However, if there are m subgroups within a group and K groups then the total number of replications required is $mK(K+1)/2$. One can show that if the number of replications is limited to R, then the largest number of groups possible, as a function of the number of subgroups m, is given by the integer portion of $0.5(-1 + \sqrt{1 + 8 \times R/m})$. We set $m = 128$, which leads to $K = 12$. This allows us to use $mK(K+1)/2 = 9,984$ out of the 10,000 replications.

and note that $\mathbb{E}[M_k] = \mu_0$ and $\mathbb{E}[W_k^2] = \sigma_0^2/k$. Further note that M_k and W_k^2 are independent of $M_{k'}$ and $W_{k'}^2$ for $k \neq k'$. Next, note that

$$\begin{aligned}
\mathbb{V}[M_k] &= \frac{1}{m}\mathbb{V}[\overline{X}_{jk}] = \frac{1}{m}\frac{\sigma_0^2}{k} && (175)\\
\mathbb{C}[M_k, W_k^2] &= \frac{1}{m^2}\sum_{j=1}^{m}\sum_{\ell=1}^{m}\mathbb{C}[\overline{X}_{jk}, (\overline{X}_{\ell k} - \mu_0)^2] = \frac{1}{m}\gamma_0 && (176)\\
\mathbb{V}[W_k^2] &= \frac{1}{m}\mathbb{V}[(\overline{X}_{jk} - \mu_0)^2] = \frac{1}{m}\left(\delta_0 - \left(\frac{\sigma_0^2}{k}\right)^2\right) && (177)
\end{aligned}$$

We now cast this problem into a method of moments framework. Define $\theta = (\mu, \sigma^2)$ and the moment vector

$$g_m(\theta) \equiv \left(M_1 - \mu \;,\; W_1^2 - \sigma^2 \;,\; M_2 - \mu \;,\; W_2^2 - \sigma^2/2 \;,\; \ldots \;,\; M_K - \mu \;,\; W_K^2 - \sigma^2/K\right)' \tag{178}$$

Following Newey (1985), note that $g_m(\theta)$ can be thought of as an average itself and write

$$g_m(\theta) = \frac{1}{m}\sum_{j=1}^{m} g(Z_j, \theta) \tag{179}$$

where $Z_j = (\overline{X}_{j1}, \overline{X}_{j2}, \ldots, \overline{X}_{jK})$ is the data for all groups. Note that this is well-behaved problem with $\mathbb{E}[g(Z_j, \theta)] = 0$ if and only if $\theta = \theta_0 \equiv (\mu_0, \sigma_0^2)$. Then define

$$V = \mathbb{E}[g(Z_j, \theta_0) g(Z_j, \theta_0)'] \tag{180}$$

noting that since the M_k sequence is independent and the W_k^2 sequence is independent, this is a $2K \times 2K$ block diagonal matrix with 2×2 blocks with the kth block given by

$$\begin{bmatrix} \sigma_0^2/k & \gamma_0 \\ \gamma_0 & \delta_0 - (\sigma_0^2/k)^2 \end{bmatrix} \tag{181}$$

Next, define the $2K \times 2$ expected derivative matrix evaluated at θ_0,

$$H = \mathbb{E}\left[\frac{\partial}{\partial\theta} g(Z_j, \theta_0)\right] = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1/2 \\ & \cdots \\ 1 & 0 \\ 0 & 1/K \end{bmatrix} \tag{182}$$

Finally, for a fixed weighting matrix A, define the method of moments estimator

$$\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma}^2) = \arg\min_{\theta} g_m(\theta)' \, A \, g_m(\theta) \tag{183}$$

This is precisely the sort of method of moments problem where there is a wisdom to avoiding the efficient choice of A (see Altonji and Segal (1996) for background). We select $A = \text{diag}\{m, m, 2m, 2m, \ldots, Km, Km\}$ because the different groups have different numbers of replications underlying them and because this choice of A does not rely on any estimated quantities.

The specification tests we are interested in are tests of the null hypothesis that

$$L\mathbb{E}[g(Z_j, \theta_0)] = 0 \tag{184}$$

where L is a matrix that focuses the test in particular directions. We focus on $L = (e_2, e_4, \ldots, e_{2K})$, where e_ℓ has a one in position ℓ and zeros everywhere else. This choice focuses attention on the pattern of the variance of the sample means across the groups. The test statistic we are interested in is given by

$$J = m \, g_m(\widehat{\theta})' \, L' Q_m^- L \, g_m(\widehat{\theta}) \tag{185}$$

where Q_m^- is a generalized inverse of the matrix $Q_m = LPV_mP'L'$ where $P \equiv I - H(H'AH)^{-1}H'A$ and where V_m is any consistent estimator for V. Under the null hypothesis that the first four moments are finite, J is distributed χ^2 with degrees of freedom equal to the rank of Q, where Q_m has probability limit Q. This rank can be shown here to be equal to $K - 1$.

A particular issue in this context is the quality of the estimator V_m. Intuitively, V is related to the second, third, and fourth moments of the individual replications X_r. To allow V_m to take advantage of these microdata estimates, we need to relate γ_0 to $\widetilde{\gamma}_0$ and δ_0 to $\widetilde{\delta}_0$. To do so, note that

$$\begin{aligned}
\gamma_0 &= \mathbb{E}[(\overline{X}_{jk} - \mu_0)^3] = \mathbb{C}[\overline{X}_{jk} - \mu_0, (\overline{X}_{jk} - \mu_0)^2] && (186)\\
&= \mathbb{C}[\overline{X}_{jk} - \mu_0, \overline{X}_{jk}^2] - 2\mu_0\mathbb{C}[\overline{X}_{jk} - \mu_0, \overline{X}_{jk}] && (187)
\end{aligned}$$

Then we have

$$\mathbb{C}[\overline{X}_{jk} - \mu_0, \overline{X}_{jk}^2] = \frac{1}{k}\sum_{i=1}^{k}\mathbb{C}[X_{ijk} - \mu_0, \overline{X}_{jk}^2] = \frac{1}{k^3}\sum_{i=1}^{k}\sum_{\ell=1}^{k}\sum_{\ell'=1}^{k}\mathbb{C}[X_{ijk} - \mu_0, X_{\ell jk}X_{\ell' jk}] \tag{188}$$

When $i \neq \ell$ and $i \neq \ell'$, this covariance is zero. When $i = \ell = \ell'$ the covariance is given by $\widetilde{\gamma}_0 + 2\mu_0\sigma_0^2$. This occurs k times. When $i = \ell$ but $\ell \neq \ell'$ the covariance is given by $\sigma_0^2\mu_0$. This occurs $k(k-1)$ times. Symmetrically, when $i = \ell'$ but $\ell \neq \ell'$ the covariance is given by $\sigma_0^2\mu_0$. This occurs $k(k-1)$ times. Overall then we have

$$\mathbb{C}[\overline{X}_{jk} - \mu_0, \overline{X}_{jk}^2] \quad = \quad \frac{1}{k^3}\left(k(\gamma_0 + 2\sigma_0^2\mu_0) + 2k(k-1)\sigma_0^2\mu_0\right) = \frac{1}{k^2}\widetilde{\gamma}_0 + \frac{2}{k}\sigma_0^2\mu_0 \tag{189}$$

Turning now to the second piece,

$$\mathbb{C}[\overline{X}_{jk} - \mu_0, \overline{X}_{jk}] \quad = \quad \frac{1}{k}\sum_{i=1}^{k}\mathbb{C}[X_{ijk} - \mu_0, \overline{X}_{jk}] = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{\ell=1}^{k}\mathbb{C}[X_{ijk} - \mu_0, X_{\ell jk}] \tag{190}$$

This covariance is zero except when $i = \ell$, when it is equal to σ_0^2. This occurs k times. Putting these results together, we have

$$\gamma_0 \quad = \quad \mathbb{E}[(\overline{X}_{jk} - \mu_0)^3] = \frac{1}{k^2}\widetilde{\gamma}_0 + \frac{2}{k}\sigma_0^2\mu_0 - \frac{2}{k}\mu_0\sigma_0^2 = \frac{1}{k^2}\widetilde{\gamma}_0 \tag{191}$$

Consider next the fourth centered moment. Note that

$$\delta_0 \quad = \quad \mathbb{E}[(\overline{X}_{jk} - \mu_0)^4] = \mathbb{V}[(\overline{X}_{jk} - \mu_0)^2] + \mathbb{E}[(\overline{X}_{jk} - \mu_0)^2]^2 \tag{192}$$

A standard result from introductory statistics[3] is that

$$\mathbb{V}[(\overline{X}_{jk} - \mu_0)^2] \quad = \quad \frac{1}{k^3}(\widetilde{\delta}_0 - 3\sigma_0^4) + \frac{2}{k^2}\sigma_0^4 \tag{193}$$

and of course $\mathbb{E}[(\overline{X}_{jk} - \mu_0)^2] = \mathbb{V}[\overline{X}_{jk}] = \sigma_0^2/k$. Putting these results together, we have

$$\delta_0 \quad = \quad \mathbb{E}[(\overline{X}_{jk} - \mu_0)^4] = \frac{1}{k^3}\widetilde{\delta}_0 + 3\sigma_0^4\frac{k-1}{k^3} \tag{194}$$

In summary, an accurate estimator for V is given by

$$V_m \quad = \quad \text{diag}\left\{\begin{bmatrix} \widehat{\sigma}^2/k & \widehat{\gamma}/k^2 \\ \widehat{\gamma}/k^2 & (1/k^3)(\widehat{\delta} - 3\widehat{\sigma}^4) + 2\widehat{\sigma}^4/k^2 \end{bmatrix}\right\} \tag{195}$$

where $\widehat{\sigma}^2$ corresponds to the estimate $\widehat{\theta}$ and $\widehat{\gamma} = \frac{1}{R}\sum_{r=1}^{R}(X_r - \overline{X})^3$ and $\widehat{\delta} = \frac{1}{R}\sum_{r=1}^{R}(X_r - \overline{X})^4$ are based on all available replications, where $\overline{X} = \frac{1}{R}\sum_{r=1}^{R}X_r$. Under the null hypothesis of finite moments, the test statistic J defined in equation (185) is distributed chi-square with $K-1$ degrees of freedom for every randomized set of groups and subgroups. As before, to test the null hypothesis, we focus on design 1, curve 1 of the Frölich DGPs and on the NSW DGP with covariates drawn from the empirical distribution.

For the estimator replications corresponding to these DGPs, we generated 10,000 group-subgroup structures, computing J for each DGP for each estimator for each structure. We then used an undersmoothed kernel density estimator to examine departures of the resulting distribution of J from the reference distribution of χ^2_{K-1}. To confirm visual impressions, we also computed the Kolmogorov statistic for each estimator. The results of this exercise are given in Web Appendix Figures 1 (Frölich) and 2 (NSW).

For most estimators, there is little evidence against the finite moments hypothesis. The Kolmogorov statistics are occasionally statistically significant at conventional levels, but usually when this is the case the extent of the discrepancy in the densities (and hence in the cumulative distribution functions) is very minor. The clear exception to this is the pair of results for the unnormalized reweighting estimator. For this estimator, there is rather clear evidence against the hypothesis of finite moments, with most J values being either quite small or quite large relative to the reference distribution. For some of the other estimators, particularly those involving cross-validation, there is some evidence against the null hypothesis, but the results are not decisive. Finally, it is interesting to note that the simple regression procedure did not detect any difficulties with the unnormalized reweighting estimator. One interpretation of this pattern is that the first two moments exist, but that the fourth, or perhaps even the third, moment does not exist.

To investigate this further, we took six batches of 10,000 draws from Student's t distribution, with degrees of freedom equal to 1, 2, 3, 4, 5, and 6. We then followed the same testing procedure described above. The results of this exercise are presented in Web Appendix Figure 3. Comparing the figures in Web Appendix Figure 3 to those in Web Appendix Figures 1 and 2, it appears as though unnormalized reweighting has finite first moment, and possibly a finite second moment, but that the estimator may not have a finite third moment and almost surely does not have a finite fourth moment.

[3] See `http://www.math.uah.edu/stat/sample/Variance.pdf` for a particularly clear derivation.

References

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

Altonji, Joseph G. and Lewis M. Segal, "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 1996, *14* (3), pp. 353–366.

Chamberlain, Gary, "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Econometrics*, 1986, *32*, 189–218.

Frölich, Markus, "Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, February 2004, *86* (1), 77–90.

Hahn, Jinyong, "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, March 1998, *66* (2), 315–331.

Knuth, Donald E., *The Art of Computer Programming*, 3rd ed., Vol. 2, New York: Addison-Wesley, 1998.

Newey, Whitney, "Generalized Method of Moments Specification Tests," *Journal of Econometrics*, 1985, *29*, 229–256.

Web Appendix Table 1. Asymptotic Variances of Selected Estimators for TOT

Estimator	Asymptotic Variance
$\widehat{\theta}_{U,tw}$	$\Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)\lvert T_i = 1\right]$
$\widehat{\theta}_{U,pw}$	$\Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[\mu_0(X_i)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)\lvert T_i = 1\right]$ $-\frac{1}{q^2}\mathbb{E}\left[\mu_0(X_i)p(X_i)Z_i\right]'\mathcal{I}^{-1}\mathbb{E}\left[\mu_0(X_i)p(X_i)Z_i\right]$ $-2\frac{1}{q^2}\mathbb{E}\left[p(X_i)(1-p(X_i))(\tau(X_i)-\theta)Z_i\right]\mathcal{I}^{-1}\mathbb{E}\left[\mu_0(X_i)p(X_i)Z_i\right]$
$\widehat{\theta}_{U,ew}$	Ω_θ
$\widehat{\theta}_{N,tw}$	$\Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)\lvert T_i = 1\right]$
$\widehat{\theta}_{N,pw}$	$\Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)\lvert T_i = 1\right]$ $-\frac{1}{q^2}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]'\mathcal{I}^{-1}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]$ $-2\frac{1}{q^2}\mathbb{E}\left[p(X_i)(1-p(X_i))(\tau(X_i)-\theta)Z_i\right]\mathcal{I}^{-1}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)p(X_i)Z_i\right]$
$\widehat{\theta}_{N,ew}$	Ω_θ
$\widehat{\theta}_{GPE,pw}$	$\Omega_\theta + \frac{1}{q^2}\mathbb{E}\left[(\mu_0(X_i)-\alpha_0)^2\frac{p(X_i)}{1-p(X_i)}\right] + 2\frac{1}{q}\mathbb{C}\left[\tau(X_i), \mu_0(X_i)\lvert T_i = 1\right]$ $+\frac{1}{q^2}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]'B^{-1}CB^{-1}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]$ $-2\frac{1}{q^2}\mathbb{E}\left[\frac{p(X_i)}{1-p(X_i)}(\mu_0(X_i)-\alpha_0)Z_i\right]'B^{-1}CB^{-1}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]$ $-2\frac{1}{q^2}\mathbb{E}\left[p(X_i)(\tau(X_i)-\theta)Z_i\right]B^{-1}\mathbb{E}\left[p(X_i)(\mu_0(X_i)-\alpha_0)Z_i\right]$
$\widehat{\theta}_{GPE,ew}$	Ω_θ
$\widetilde{\theta}_{NN}$	$\Omega_\theta + \frac{1}{q^2}\frac{1}{2k}\mathbb{E}\left[\sigma_0^2(X_i)\left(\frac{p(X_i)}{1-p(X_i)} + p(X_i)\right)\right]$

Notes: Here, $\mathcal{I} = \mathbb{E}[p(X_i)(1-p(X_i))Z_iZ_i']$, $B = \mathbb{E}[p(X_i)Z_iZ_i']$, $C = \mathbb{E}[\frac{p(X_i)}{1-p(X_i)}Z_iZ_i']$, $q = \mathbb{E}\left[p(X_i)\right]$, $q^2\Omega_\theta = \mathbb{E}\left[\sigma_1^2(X_i)p(X_i) + \sigma_0^2(X_i)\frac{p(X_i)^2}{1-p(X_i)} + (\tau(X_i)-\theta)^2p(X_i)\right]$, and k is number of neighbors.

Web Appendix Table 2. Link Between $\mathbb{V}[\overline{X}]$ and k^{-1}: R^2 Values

Estimator	Frölich Design A	Frölich Design B	NSW Design A	NSW Design B
Propensity Score Matching				
Pair ($k = 1$)	0.991	0.993	0.992	0.992
NN ($k = 4$)	0.991	0.982	0.987	0.975
BCM ($k = 4$)	0.992	0.984	0.987	0.976
NN (CV)	0.988	0.985	0.961	0.966
BCM (CV)	0.999	0.998	0.995	0.990
LL (CV)	0.982	0.988	0.976	0.973
Covariate Matching				
Pair ($k = 1$)	0.966	0.974	0.937	0.912
NN ($k = 4$)	0.995	0.990	0.987	0.977
BCM ($k = 4$)	0.992	0.991	0.992	0.988
NN (CV)	0.995	0.992	0.968	0.956
BCM (CV)	0.996	0.994	0.992	0.985
Reweighting				
Unnorm.	0.981	0.974	0.981	0.970
Norm.	0.993	0.992	0.991	0.987
GPE	0.990	0.985	0.983	0.978

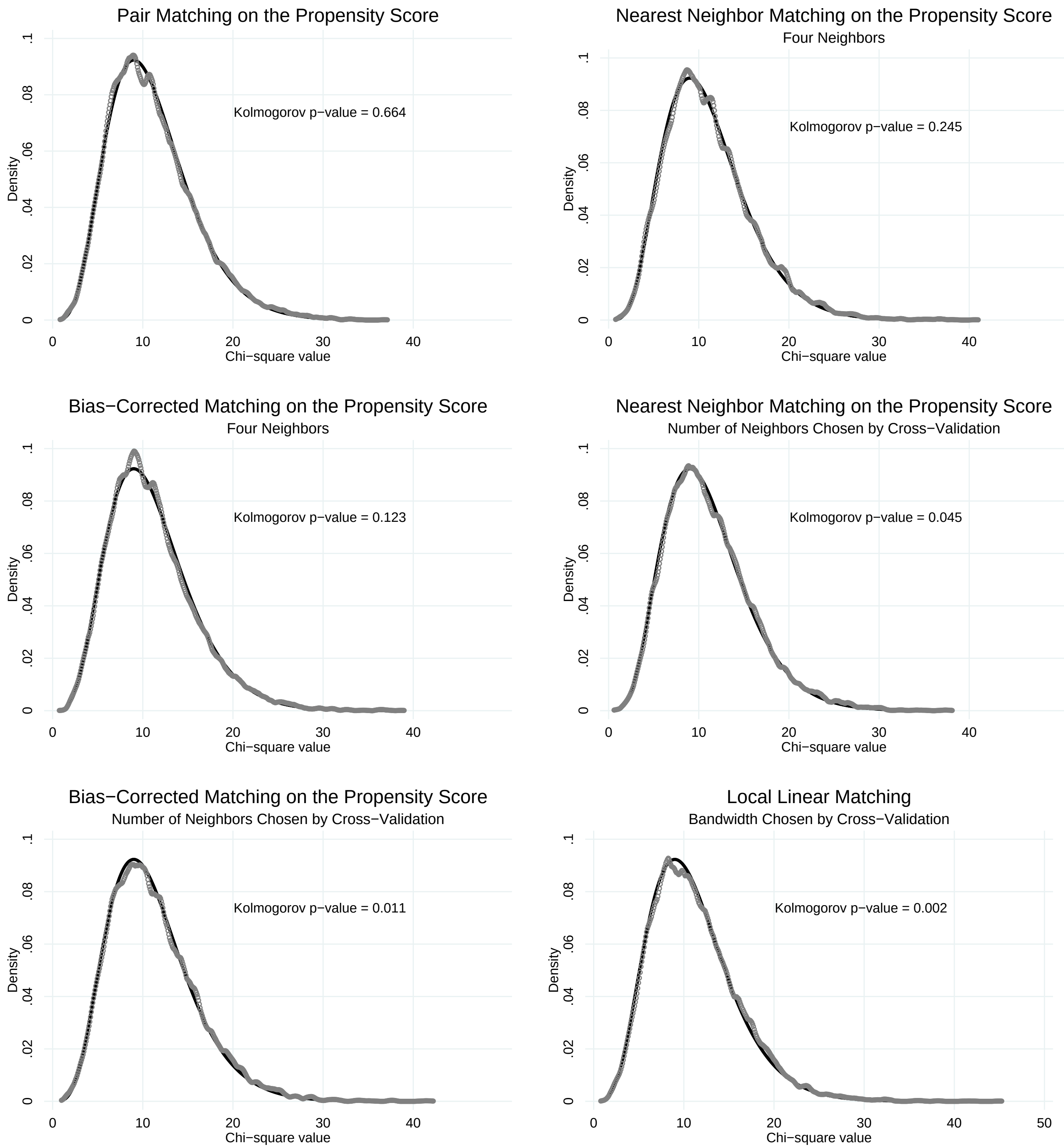
Notes: Table gives R^2 for a regression of sample variances for groups indexed by k on k^{-1}. Columns A correspond to a regression without a constant term and columns B correspond to a regression with a constant term. See Web Appendix text for details.

Web Appendix Table 3. Percentiles of R^2

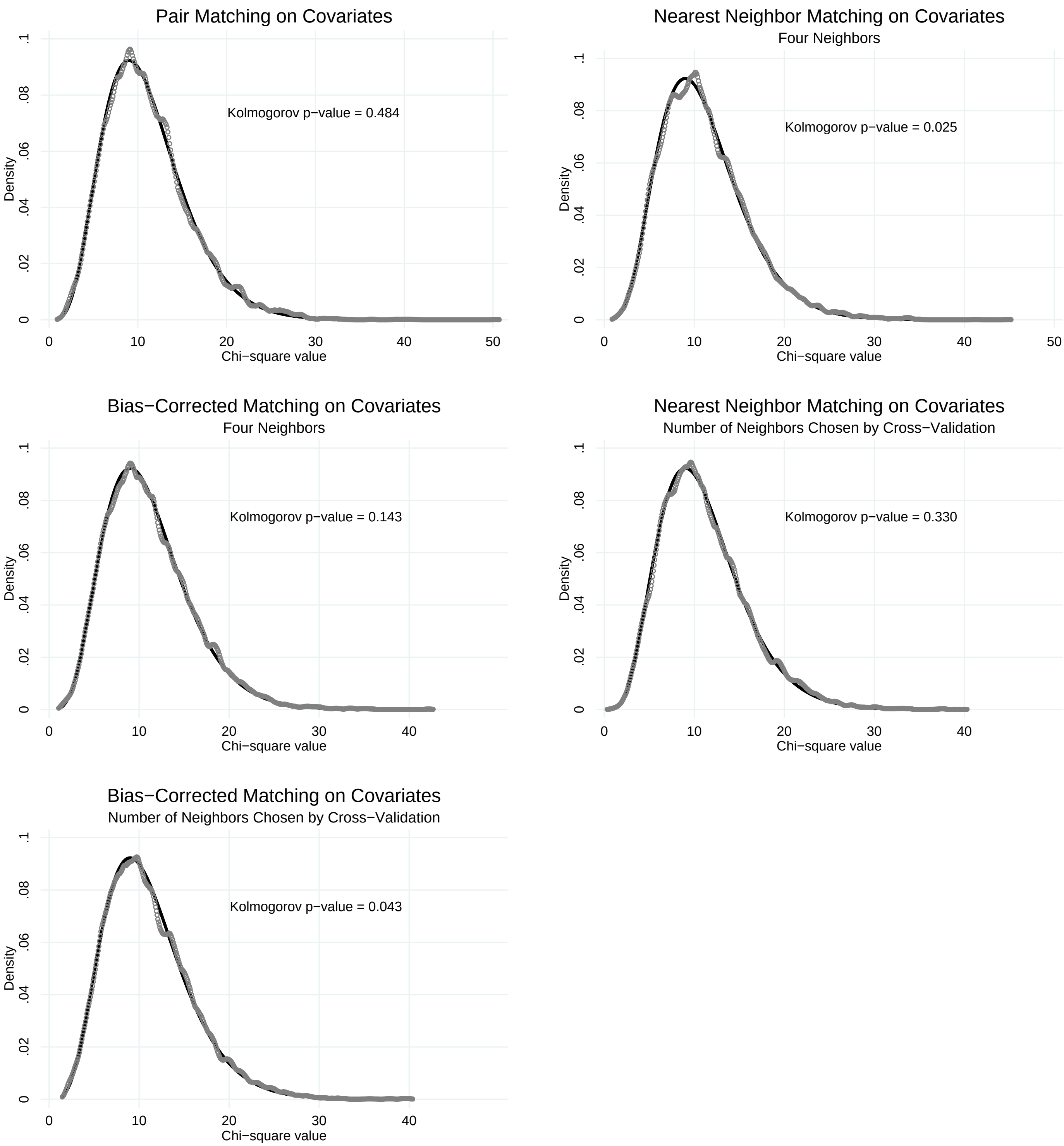
dof	A 5th	A 95th	B 5th	B 95th
1	0.008	0.653	0.000	0.693
2	0.160	0.955	0.011	0.930
3	0.779	0.992	0.606	0.987
4	0.914	0.996	0.853	0.994
5	0.946	0.997	0.913	0.995
6	0.956	0.997	0.929	0.996

Notes: Table gives 5th and 95th percentiles of R^2 across 10,000 possible group-subgroup structures, for a fixed set of 10,000 draws from Student's t distriubtion with varying degrees of freedom. Recall that Student's t with ν degrees of freedom has $\nu - 1$ moments, but that the νth moment either does not exist or is infinite.

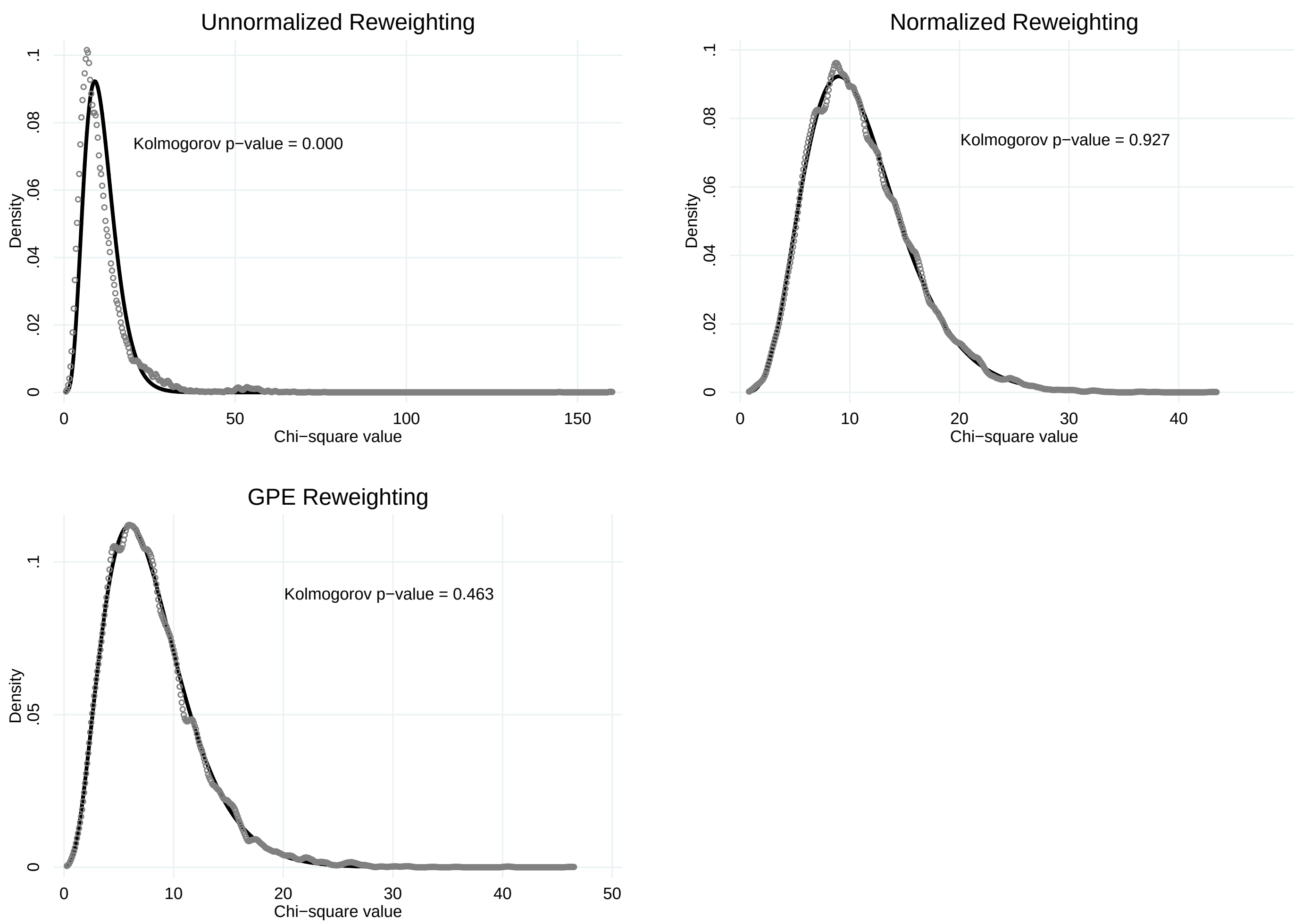
Web Appendix Figure 1. Testing for Finite Moments:
Density Estimates for J in Design 1, Curve 1



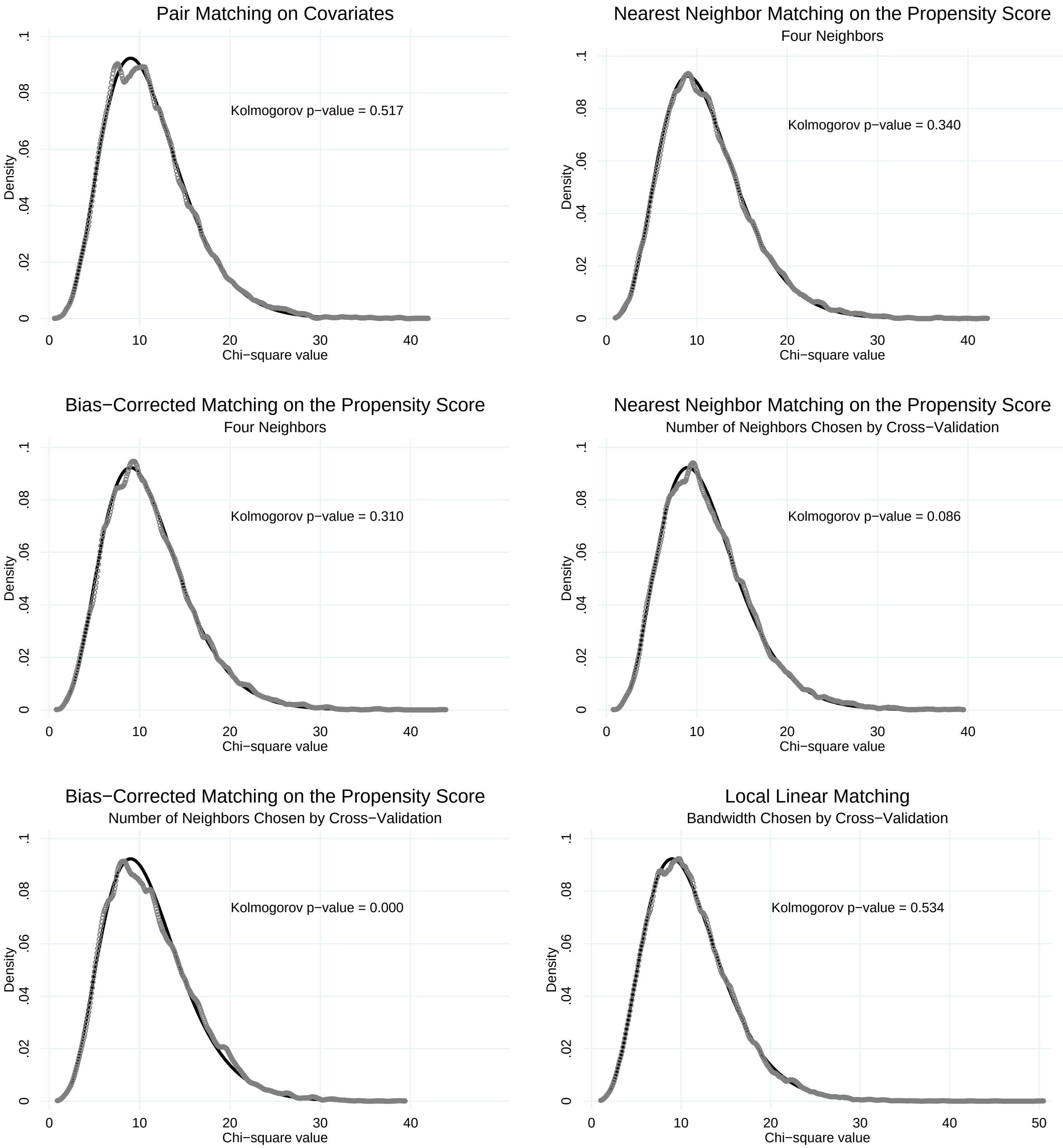
Web Appendix Figure 1. Testing for Finite Moments:
Density Estimates for J in Design 1, Curve 1 (cont.)



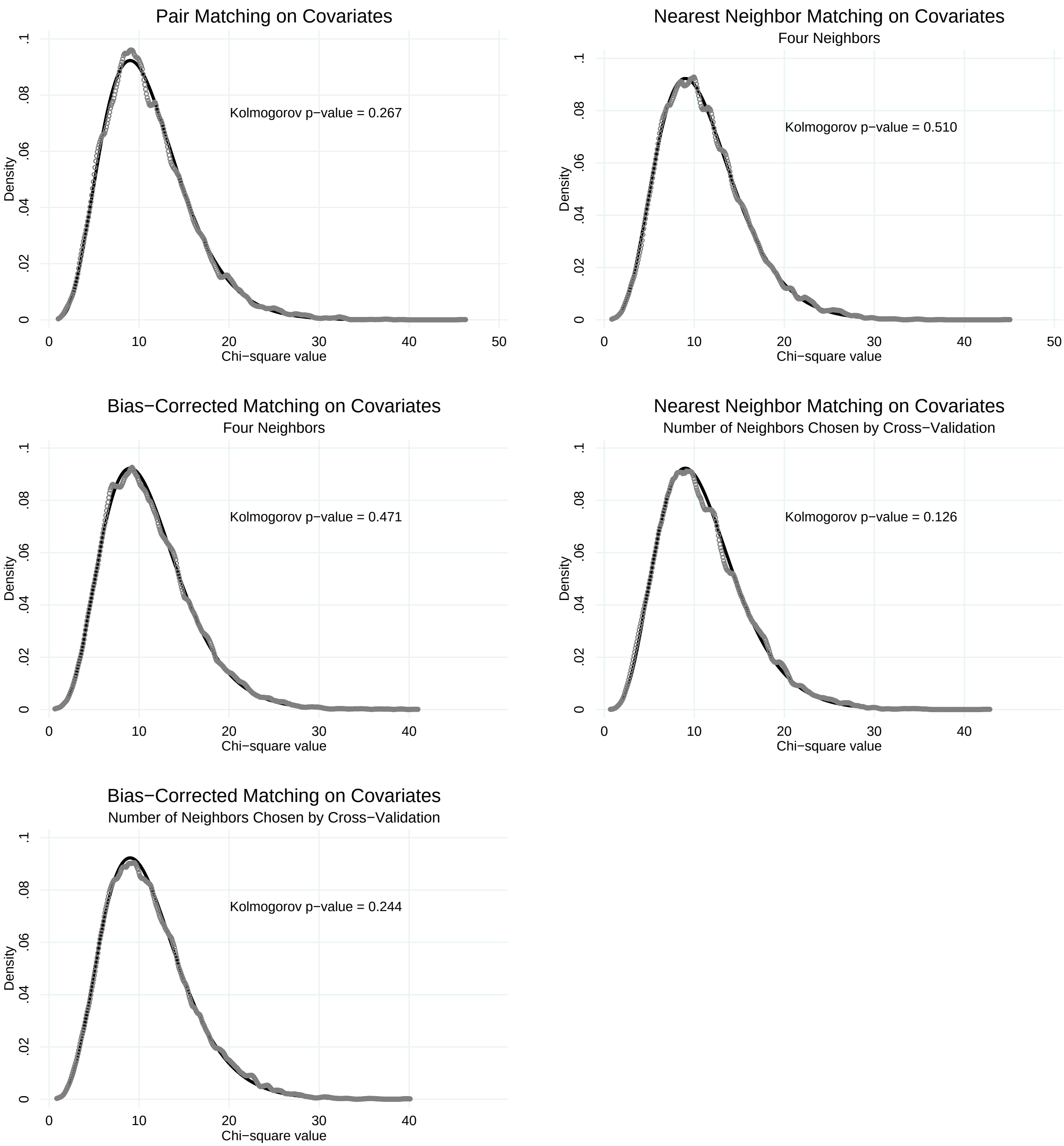
Web Appendix Figure 1. Testing for Finite Moments: Density Estimates for J in Design 1, Curve 1 (cont.)



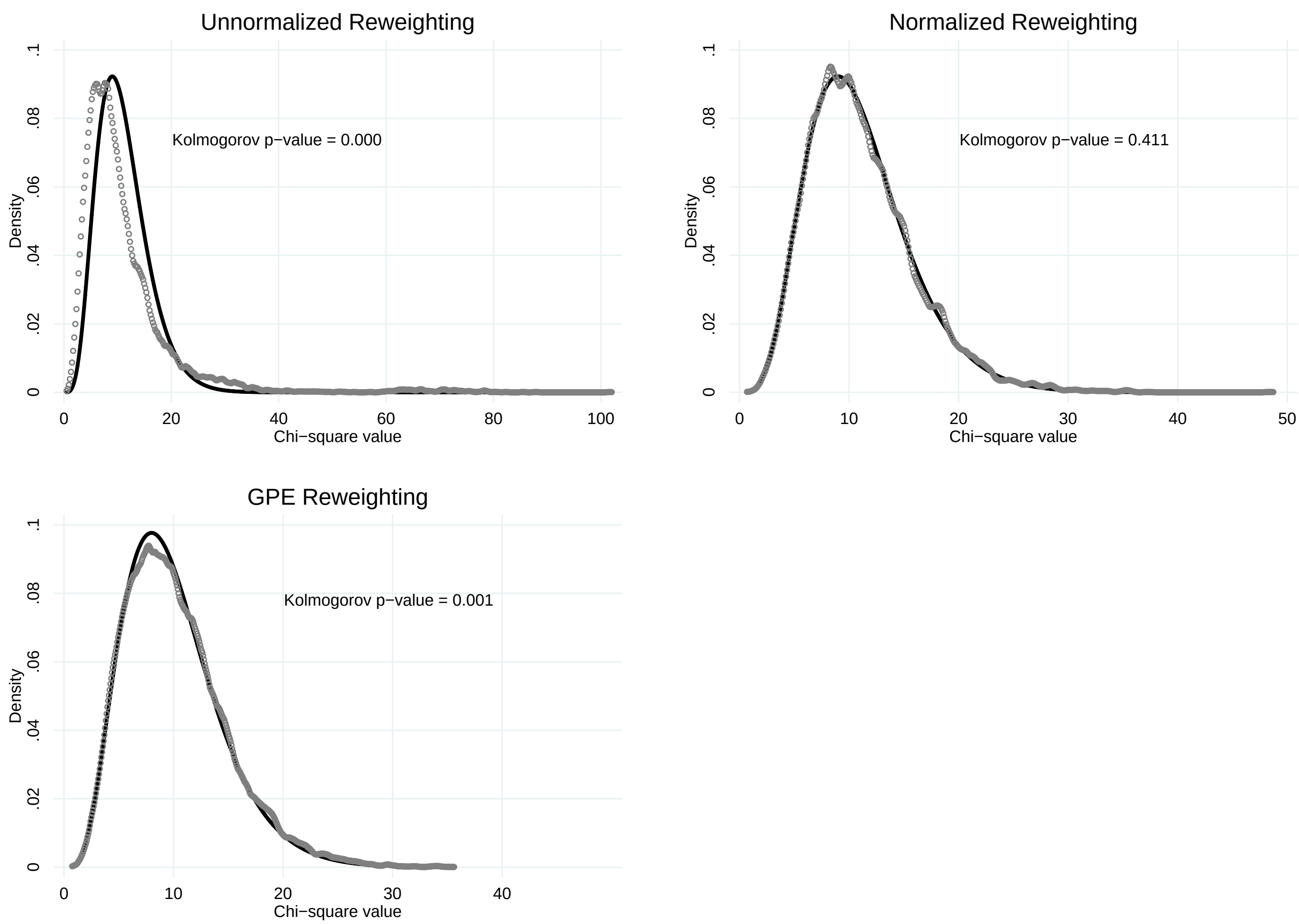
Web Appendix Figure 2. Testing for Finite Moments:
Density Estimates for J in NSW Design



Web Appendix Figure 2. Testing for Finite Moments:
Density Estimates for J in NSW Design (cont.)



Web Appendix Figure 2. Testing for Finite Moments:
Density Estimates for J in NSW Design (cont.)



Web Appendix Figure 3. Testing for Finite Moments:
Density Estimates for J Based on Student's t Distribution