Efficient estimation of data combination models by the method of auxiliary-to-study tilting (AST), supplemental material: additional proofs and derivations

This supplemental appendix contains a proof of Theorem 2.1, descriptions of some additional examples of data combination problems, further Monte Carlo results, and a computational algorithm that we have found to work well in practice. It also details some of the more tedious calculations underlying the proofs of Theorems 3.1, 3.2 and 3.3. All notation is as defined in the main text unless stated otherwise. Equation and Table numbering continues in sequence with that established in the main text. References not cited in the main text are listed below.

B Additional Examples

Poverty mapping: Let X be an indicator denoting whether a household's total outlay falls below a poverty line and W a vector of household characteristics. We seek to estimate the poverty rate in a specific study municipality as in Elbers, Lanjouw and Lanjouw (2003) and Tarozzi and Deaton (2009). Available is a random sample of N_s observations of W from this municipality; however, no poverty measurements are available in this sample. Also available is a random sample of size N_a of both X and W from the entire country. Our estimand is $\gamma_0 = \mathbb{E}_s[X]$ which corresponds to setting $\psi_s(Y, W, \gamma) = 0$ and $\psi_a(X, W, \gamma) = X - \gamma$. In this example part two of Assumption 2.1 implies that the conditional probability of begin poor given $W = w$ is the same in the entire country as it is in the specific municipality of interest.

Counterfactual distributions and direct standardization: Let Y be wages of employed Black males and X those of White males. Let W be a vector of worker characteristics. A random study sample of Black, and another auxiliary sample of White, workers are available. We seek to decompose differences in specific quantiles of the Black and White wage distributions into portions due to (i) differences in the distribution of characteristics, and (ii) differences in the mapping from those characteristics into wages, across the two populations. The latter difference is sometimes interpreted as a measure of labor market discrimination, although this interpretation is not assumption free (cf., Darity and Mason, 1998).

This decomposition requires knowledge of the distribution of White wages that would prevail under the Black distribution of worker characteristics. That is, what would the wage distribution look like in a hypothetical White population whose distribution of W coincided with the one in the actual Black population? The αth quantile of this counterfactual distribution, $\gamma_{W|B}^{\alpha}$, is identified by

$$\mathbb{E}_s\left[\mathbf{1}(X \leq \gamma_{W|B}^{\alpha}) - \alpha\right] = 0,$$

which corresponds to setting $\psi_0(Y_0, X, \gamma) = \alpha - \mathbf{1}(X \leq \gamma_{W|B}^{\alpha})$ and $\psi_1(Y_1, X, \gamma)$ to a vector of zeros. The αth quantiles of the actual Black and White earnings distributions are denoted by $\gamma_{B|B}^{\alpha}$ and $\gamma_{W|W}^{\alpha}$. A decomposition into wage structure and compositional effects is then given by

$$\gamma_{B|B}^{\alpha} - \gamma_{W|W}^{\alpha} = \left(\gamma_{B|B}^{\alpha} - \gamma_{W|B}^{\alpha}\right) - \left(\gamma_{W|W}^{\alpha} - \gamma_{W|B}^{\alpha}\right).$$

Barsky, Bound, Charles and Lupton (2002) and Fortin, Lemieux and Firpo (2010) survey alternative decomposition methods. For discretely-valued W these methods are similar to techniques used by demographers to standardize mortality rates across localities (e.g., Kitagawa, 1964).

C Proof of Theorem 2.1

Theorem 2.1 is a slight generalization of Theorem 3 of Chen, Hong and Tarozzi (2008). We provide a proof for completeness. In calculating the efficiency bound for the semiparametric data combination model defined by Assumption

2.1, we use the approach outlined by Bickel, Klaassen, Ritov and Wellner (1993) and, especially, Newey (1990, Section 3). First, we characterize the nuisance tangent space. Second, we demonstrate pathwise differentiability of the parameter of interest, γ_0. The efficient influence function equals the projection of the pathwise derivative onto the tangent space. The third and final step of the proof is to calculate the projection and demonstrate that lies in the model tangent space. The result then follows from an application of Theorem 3.1 in Newey (1990).

Step 1: Characterization of the nuisance tangent space Let $R = (1 - D) X + DY$; the joint density of (R, W, D), making use of parts (ii) and (v) of Assumption 2.1, is given by

$$f(r, w, d) = f(y|w)^d f(x|w)^{1-d} p(w, \delta)^d [1 - p(w, \delta)]^{1-d} f(w).$$

Consider a regular parametric submodel with $f(r, w, d; \eta) = f(r, w, d)$ at $\eta = \eta_0$. The submodel joint density is given by

$$f(r, w, d; \eta) = f(y|w; \eta)^d f(x|w; \eta)^{1-d} p(w, \delta(\eta))^d [1 - p(w, \delta(\eta))]^{1-d} f(w; \eta).$$

The submodel score vector equals

$$
\begin{aligned}
s_\eta(r, w, d; \eta) \quad = \quad & ds_\eta(y|w; \eta) + (1 - d) s_\eta(x|w; \eta) \\
& + \left(\frac{\partial \delta(\eta)}{\partial \eta'} \right)' \frac{d - p(w, \delta(\eta))}{p(w, \delta(\eta)) [1 - p(w, \delta(\eta))]} p_\delta(w, \delta(\eta)) + t_\eta(w; \eta),
\end{aligned}
\tag{35}
$$

where $p_\delta(w, \delta) = \nabla_\delta p(w, \delta)$ and

$$s_\eta(y|w; \eta) = \nabla_\eta \log f(y|w; \eta), \quad s_\eta(x|w; \eta) = \nabla_\eta \log f(x|w; \eta), \quad t_\eta(w; \eta) = \nabla_\eta \log f(w; \eta).$$

By the usual conditional mean zero property of scores we have

$$\mathbb{E}[s_\eta(Y|W)|W] = \mathbb{E}[s_\eta(X|W)|W] = \mathbb{E}[t_\eta(W)] = 0,
\tag{36}$$

where the suppression of η in a function means that it is evaluated at its population value (e.g., $t_\eta(w) = t_\eta(w; \eta_0)$).

From (35) and (36) the tangent set is evidently

$$\mathcal{T} = \{ ds(y|w) + (1 - d) s(x|w) + \mathbf{c}\mathbb{S}_\delta(d, w) + t(w) \},$$

where $\mathbb{S}_\delta(d, w)$ is the score vector associated with the parametric propensity score model evaluated at $\delta_0 = \delta(\eta_0)$:

$$\mathbb{S}_\delta(d, w) = \frac{d - p(w, \delta_0)}{p(w, \delta_0) [1 - p(w, \delta_0)]} p_\delta(w, \delta_0),$$

with $\mathbf{c}$ a matrix of constants, and $s(y|w)$, $s(x|w)$, $\mathbb{S}_\delta(d, w)$ and $t(w)$ satisfying

$$\mathbb{E}[s(Y|W)|W] = \mathbb{E}[s(X|W)|W] = \mathbb{E}[\mathbb{S}_\delta(D, W)|W] = \mathbb{E}[t(W)] = 0,$$

but otherwise unrestricted.

Step 2: Demonstration of pathwise differentiability Under the parametric submodel $\gamma_0 = \gamma(\eta_0)$ is identified by the moment restriction

$$\mathbb{E}_\eta[\psi_s(Y, W, \gamma(\eta_0)) - \psi_a(X, W, \gamma(\eta_0))| D = 1] = 0.$$

Differentiating under the integral and evaluating at $\eta = \eta_0$ gives

$$
\begin{aligned}
\frac{\partial \gamma(\eta_0)}{\partial \eta'} \quad = \quad & -\mathbb{E}\left[\frac{p_0(W)}{Q_0} \Gamma_0(W) \right]^{-1} \left\{ \mathbb{E}\left[\psi_s(Y, W, \gamma_0) \frac{\partial \log f(Y, W| D = 1; \eta_0)}{\partial \eta'} \right] \right. \\
& \left. -\mathbb{E}\left[\psi_a(X, W, \gamma_0) \frac{\partial \log f(X, W| D = 1; \eta_0)}{\partial \eta'} \right] \right\}.
\end{aligned}
\tag{37}
$$

2

To demonstrate pathwise differentiability of γ we require $F(R, W, D)$ such that

$$\frac{\partial \gamma(\eta_0)}{\partial \eta'} = \mathbb{E}\left[F(R, W, D) s_\eta(R, W, D)'\right]. \tag{38}$$

Recalling that $p_0(w) = p(w, \delta_0)$, $q_s(W) = \mathbb{E}\left[\psi_s(Y, W, \gamma_0)| W\right]$, and $q_a(W) = \mathbb{E}\left[\psi_a(X, W, \gamma_0)| W\right]$. Let

$$
\begin{aligned}
F(R, W, D) \quad = \quad & -\mathbb{E}\left[\frac{p_0(W)}{Q_0}\Gamma_0(W)\right]^{-1} \\
& \times \left[\frac{D}{Q_0}\left\{\psi_s(Y, W, \gamma_0) - q_s(W)\right\} - \frac{1-D}{Q_0}\frac{p_0(W)}{1 - p_0(W)}\left\{\psi_a(X, W, \gamma_0) - q_a(W)\right\} \right. \\
& + \frac{p_0(W)}{Q_0}\left\{q_s(W) - q_a(W)\right\} \\
& \left. + \frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{D}{p_0(W)} - 1\right) p_0(W)\left\{q_s(W) - q_a(W)\right\}\Big| \mathbb{S}_\delta\right]\right].
\end{aligned}
\tag{39}
$$

Letting $Q(\eta) = \int p(w, \delta(\eta)) f(w; \eta)\, \mathrm{d}w$ we have, using part (ii) of Assumption 2.1, and Bayes' Law,

$$f(y, w| d = 1; \eta) = f(y| w; \eta) f(w| d = 1; \eta) = f(y| w; \eta)\frac{p(w, \delta(\eta)) f(w; \eta)}{Q(\eta)},$$

so that

$$\frac{\partial \log f(Y, W| D = 1; \eta_0)}{\partial \eta'} = s_\eta(Y| W)' + \left(\frac{p_\delta(W, \delta(\eta_0))}{p(W, \delta(\eta_0))}\right)'\frac{\partial \delta(\eta_0)}{\partial \eta'} + t_\eta(W) - \frac{Q'_\eta}{Q_0},$$

which allows the first term inside the $\{\cdot\}$ in (37) to be re-written as

$$
\begin{aligned}
\mathbb{E}\left[\psi_s(Y, W, \gamma_0)\frac{\partial \log f(Y, W| D = 1; \eta_0)}{\partial \eta'}\Big| D = 1\right] \quad = \quad & \mathbb{E}\left[\psi_s(Y, W, \gamma_0) s_\eta(Y| W)'\big| D = 1\right] \\
& + \mathbb{E}\left[\psi_s(Y, W, \gamma_0) \mathbb{S}_\delta(D, W)'\big| D = 1\right]\frac{\partial \delta(\eta_0)}{\partial \eta'} \\
& + \mathbb{E}\left[\psi_s(Y, W, \gamma_0) t_\eta(W)'\big| D = 1\right] \\
& - \mathbb{E}\left[\psi_s(Y, W, \gamma_0)| D = 1\right]\frac{Q'_\eta}{Q_0}.
\end{aligned}
$$

Using the conditional mean zero property of $s_\eta(Y| W)$, part (ii) of Assumption 2.1, Bayes' Law, and iterated expectations we can then show that

$$\mathbb{E}\left[\psi_s(Y, W, \gamma_0) s_\eta(Y| W)'\big| D = 1\right] \quad = \quad \mathbb{E}\left[\frac{D}{Q_0}\left\{\psi_s(Y, W, \gamma_0) - q_s(W)\right\} s_\eta(Y| W)'\right] \tag{40}$$

$$\mathbb{E}\left[\psi_s(Y, W, \gamma_0) t_\eta(W)'\big| D = 1\right] \quad = \quad \mathbb{E}\left[\frac{p_0(W)}{Q_0}q_s(W) t_\eta(W)'\right]. \tag{41}$$

We also have

$$
\begin{aligned}
\mathbb{E}\left[\psi_s(Y, W, \gamma_0) \mathbb{S}_\delta(D, W)'\big| D = 1\right] \quad = \quad & \mathbb{E}\left[\psi_s(Y, W, \gamma_0)\frac{p_\delta(W, \delta_0)'}{p(W, \delta_0)}\Big| D = 1\right] \\
= \quad & \frac{1}{Q_0}\mathbb{E}\left[\left(\frac{D}{p_0(W)} - 1\right) p_0(W) q_s(W) \mathbb{S}_\delta(D, W)'\right].
\end{aligned}
\tag{42}
$$

Now observe that, for $\mathbb{S}_\delta = \mathbb{S}_\delta(D, W)$,

$$\mathbb{E}^*\left[\left(\frac{D}{p_0(W)} - 1\right) p_0(W) q_s(W)\Big| \mathbb{S}_\delta\right] = \mathbb{E}\left[\left(\frac{D}{p_0(W)} - 1\right) p_0(W) q_1(W) \mathbb{S}_\delta'\right]\mathbb{E}\left[\mathbb{S}_\delta\mathbb{S}_\delta'\right]^{-1}\mathbb{S}_\delta,$$

so that

$$\mathbb{E}\left[\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\bigg|\,\mathbb{S}_\delta\right]\mathbb{S}_\delta'\right]=\mathbb{E}\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\mathbb{S}_\delta'\right].\tag{43}$$

Using the conditional mean zero property of $s_\eta\left(Y\,|\,W\right)$, part (ii) of Assumption 2.1, Bayes' Law, and iterated expectations we can also show that

$$\mathbb{E}\left[\psi_a\left(X,W,\gamma_0\right)s_\eta\left(X\,|\,W\right)'\big|\,D=1\right]\quad=\quad\mathbb{E}\left[\frac{\left(1-D\right)p_0\left(W\right)\left\{\psi_a\left(X,W,\gamma_0\right)-q_a\left(X\right)\right\}}{\left(1-p_0\left(W\right)\right)Q_0}s_\eta\left(X\,|\,W\right)'\right]\tag{44}$$

$$\mathbb{E}\left[\psi_a\left(X,W,\gamma_0\right)t_\eta\left(W\right)'\big|\,D=1\right]\quad=\quad\mathbb{E}\left[\frac{p_0\left(W\right)}{Q_0}q_a\left(W\right)t_\eta\left(W\right)'\right]\tag{45}$$

$$\mathbb{E}\left[\psi_a\left(X,W,\gamma_0\right)\mathbb{S}_\delta\left(D,W\right)'\big|\,D=1\right]\quad=\quad\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)\mathbb{S}_\delta\left(D,W\right)'\right]\tag{46}$$

Using (40) to (46) it is straightforward to verify that condition (38) holds for $F\left(R,W,D\right)$ as defined in (39) above.

Step 3: Calculation of efficient influence function   The variance bound for γ_0 equals the expected square of the projection of $F\left(R,W,D\right)$ – as defined by (39) above – onto the tangent space. Since $F\left(R,W,D\right)$ belongs to the tangent space, this projection equals $F\left(R,W,D\right)$ itself. To verify that $F\left(R,W,D\right)$ lies in the model tangent space note that

$$\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)\left\{q_s\left(W\right)-q_a\left(W\right)\right\}\bigg|\,\mathbb{S}_\delta\right]$$

$$=\underbrace{\mathbb{E}\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)\left\{q_s\left(W\right)-q_a\left(W\right)\right\}\mathbb{S}_\delta'\right]\mathbb{E}\left[\mathbb{S}_\delta\mathbb{S}_\delta'\right]^{-1}}_{\mathbf{c}}\mathbb{S}_\delta$$

plays the role of $\mathbf{c}\mathbb{S}_\delta\left(d,w\right)$, the first two terms in (39) play the role of $ds\left(y\,|\,w\right)$ and $\left(1-d\right)s\left(x\,|\,w\right)$, while the third term plays the role of $t\left(w\right)$.

D   Computation

The first and third steps of AST may be completed using standard software. In this appendix we detail a method for computing the second stage tilting parameters $\widehat{\lambda}_s$ and $\widehat{\lambda}_a$. The algorithm described below is a modified version of the one developed in Graham, Pinto and Egel (2012). Let

$$\varphi\left(v\right)=\frac{v}{G\left(v\right)}+\int_{1/G\left(v\right)}^a G^{-1}\left(\frac{1}{t}\right)\mathrm{d}t,\tag{47}$$

with $G\left(\cdot\right)$ as defined in the main text. The first and second derivatives of $\varphi\left(v\right)$ are

$$\varphi_1\left(v\right)=\frac{1}{G\left(v\right)},\qquad\varphi_2\left(v\right)=-\frac{G_1\left(v\right)}{G\left(v\right)^2},\tag{48}$$

so that (47) is strictly concave.

We compute $\widehat{\lambda}_s$ by solving the following optimization problem

$$\max_{\lambda_s}l_N\left(\lambda_s\right),\quad l_N\left(\lambda_s\right)=\frac{1}{N}\sum_{i=1}^N G\left(r\left(W_i\right)'\widehat{\delta}\right)\left[D_i\varphi\left(r\left(W_i\right)'\widehat{\delta}+t\left(W_i\right)'\lambda_s\right)-t\left(W_i\right)'\lambda_s\right],\tag{49}$$

where $\widehat{\delta}$ is the MLE of the propensity score parameter. Differentiating $l_N\left(\lambda_s\right)$ with respect to λ_s gives an $1+M\times 1$

gradient vector of (for $\dim(\lambda_s) = 1 + M$)

$$
\begin{aligned}
\underset{1+M\times 1}{\nabla_{\lambda_s} l_N(\lambda_s)} &= \frac{1}{N}\sum_{i=1}^{N} G\left(r(W_i)'\widehat{\delta}\right)\left[D_i\varphi_1\left(r(W_i)'\widehat{\delta}+t(W_i)'\lambda_s\right)-1\right]t(W_i) \qquad (50)\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_i}{G\left(r(W_i)'\widehat{\delta}+t(W_i)'\lambda_s\right)}-1\right)G\left(r(W_i)'\widehat{\delta}\right)t(W_i),
\end{aligned}
$$

which coincides with (11) in the main text as required. The $1+M\times 1+M$ Hessian matrix is

$$
\underset{1+M\times 1+M}{\nabla_{\lambda_s\lambda_s} l_N(\lambda_s)} = \frac{1}{N}\sum_{i=1}^{N} D_i\varphi_2\left(r(W_i)'\widehat{\delta}+t(W_i)'\lambda_s\right)G\left(r(W_i)'\widehat{\delta}\right)t(W_i)t(W_i)' \qquad (51)
$$

This is a negative definite function of λ_s; the problem (49) is consequently concave with a unique solution (if one exists).

When $t(W_i)'\lambda_s$ is a large negative number the Hessian (51) will be ill-conditioned. We address this problem by noting that at a valid solution $\sum_{i=1}^{N} D_i G\left(r(W_i)'\widehat{\delta}\right)/G\left(r(W_i)'\widehat{\delta}+t(W_i)'\widehat{\lambda}_s\right)/N = Q_N(\delta)$ for $Q_N(\delta) = \sum_{i=1}^{N} G\left(r(W_i)'\delta\right)/N$ (recall that $t(W_i)$ includes a constant). Since $G(v)$ is bounded below by zero, this means that $D_i G\left(r(W_i)'\widehat{\delta}\right)/G\left(r(W_i)'\widehat{\delta}+t(W_i)'\widehat{\lambda}_s\right) < NQ_N\left(\widehat{\delta}\right)$ for all $i=1,\ldots,N$. Letting $v_i = r(W_i)'\widehat{\delta}+t(W_i)'\lambda_s$ this inequality corresponds to requiring that

$$
G^{-1}\left(D_i G\left(r(W_i)'\widehat{\delta}\right)/NQ_N\left(\widehat{\delta}\right)\right) < v_i, \qquad i=1,\ldots,N \qquad (52)
$$

at $\lambda_s = \widehat{\lambda}_s$. Let $v_N^* = G^{-1}\left(1/NQ_N\left(\widehat{\delta}\right)\right)$; note that $v_N^* \to -\infty$ as $N\to\infty$ suggesting that (52) will be satisfied for most values of λ_s in large enough samples. In small samples (52) may be violated for some i at some iterations of the maximization procedure (although not at a valid solution). For estimation we replace $\varphi(v)$ with a quadratic function when $v \le v_N^*$; this ensures that the denominator in (50) is bounded. This improves the condition of the Hessian with respect to λ_s without changing the solution (cf., Owen 2001, Chapter 12).

Specifically we replace $\varphi(v)$ in (49), (50) and (51) with

$$
\varphi_N^\circ(v) = \begin{cases} \varphi(v) & v > v_N^* \\ a_N + b_N v_N^* + \frac{c_N}{2}(v_N^*)^2 & v \le v_N^* \end{cases}, \qquad (53)
$$

where a_N, b_N and c_N are the solutions to

$$
\begin{aligned}
c_N &= \varphi_2(v_N^*)\\
b_N + c_N v_N^* &= \varphi_1(v_N^*)\\
a_N + b_N v_N^* + \frac{c_N}{2}(v_N^*)^2 &= \varphi_0(v_N^*).
\end{aligned}
$$

This choice of coefficients ensures that $\varphi_N^\circ(v)$ equals $\varphi(v)$, as well as equality of first and second derivatives, *at* $v = v_N^*$.

When $G(v)$ is logit our algorithm is particularly simple to implement. We have $\varphi(v) \propto v - \exp(-v)$, $\varphi_1(v) = 1 + \exp(-v)$, $\varphi_2(v) = -\exp(-v)$, and $v_N^* = \ln\left(\frac{1}{NQ_N(\widehat{\delta})-1}\right)$. Solving for a_N, b_N and c_N yields

$$
\begin{aligned}
a_N &= -\left(NQ_N\left(\widehat{\delta}\right)-1\right)\left[1+\ln\left(\frac{1}{NQ_N\left(\widehat{\delta}\right)-1}\right)+\frac{1}{2}\left[\ln\left(\frac{1}{NQ_N\left(\widehat{\delta}\right)-1}\right)\right]^2\right]\\
b_N &= NQ_N\left(\widehat{\delta}\right)+\left(NQ_N\left(\widehat{\delta}\right)-1\right)\ln\left(\frac{1}{NQ_N\left(\widehat{\delta}\right)-1}\right)\\
c_N &= -\left(NQ_N\left(\widehat{\delta}\right)-1\right).
\end{aligned}
$$

E Additional Monte Carlo experiments

Recently Qin and Zhang (2008) have proposed an empirical likelihood type estimator for the difference-in-differences program evaluation parameter (e.g., Abadie, 2005). This parameter may be viewed as a special case of the average treatment effect on the treated (ATT) parameter. Their procedure, like ours, calibrates estimates of the study population distributions of (Y, W) and (X, W) to features of $\widehat{F}_s^{\text{eff}}(w)$. They use empirical likelihood methods for this purpose, as opposed to our 'tilting' equations (9) and (11). In order to compare our method with the Qin and Zhang (2008) EL procedure we replicated a subset of their Monte Carlo experiments. Adapting their setup to our notation we let

$$W_1 \sim \mathcal{N}(0, 1), \qquad W_2 | W_1 = w_1 \sim \mathcal{N}(1 + 0.6w_1, 1),$$

and

$$Y | W, D \sim \mathcal{N}\left(\mu_Y(W), W_2^2\right), \qquad X | W, D \sim \mathcal{N}\left(\mu_X(W), W_2^2\right).$$

They assume the propensity score takes a logit form with an index linear in $W = (W_1, W_2)'$ (this in turn induces the conditional distributions of W given $D = 0, 1$). The intercept in the logit index is set equal to one across all designs, while the two slope coefficients equal 0.1, 0.2 or 0.5 (corresponding to increasing selection bias). The two conditional mean parameters are set equal to $\mu_Y(W) = 2 + 2W_1 + 2W_2$ and $\mu_X(W) = 2W_1 + 2W_2$ in Design (a) and $\mu_Y(W) = 2 + 2W_1^2 - W_2 + 3W_2^2$ and $\mu_X(W) = 2W_1^2 - W_2 + 3W_2^2$ in Design (b). Analogously to Qin and Zhang (2008) we choose two different specifications for $t(W)$. First, a 'linear' one of $t(W) = (1, W_1, W_2)'$. This corresponds to the locally efficient choice in Design (a). Second, a 'quadratic' one of $t(W) = \left(1, W_1^2, W_2^2\right)'$. This choice in not efficient in either design, but is expected to be more appropriate for Design (b). Across all designs the propensity score is correctly specified with $r(W) = (1, W_1, W_2)'$. We set $N = 1,000$ and perform 1,000 Monte Carlo replications. The Monte Carlo statistics for the EL estimator are as reported in Table 2 of Qin and Zhang (2008, p. 341).

By Theorems 3.1 and 3.2 above, and Theorem 3 of Qin and Zhang (2008, p. 339), both the AST estimator and the EL estimator should be consistent and asymptotically normal across both designs and choices of $t(W)$. Our AST estimator should be efficient in Design (a) when $t(W)$ takes the linear form. (see Table 5 in the supplemental appendix).

In Design (a) the AST and EL estimator perform similarly in terms of bias (see Table 4). However, when $t(W)$ is (correctly) specified to be linear in W, AST has substantially less sampling variation that the EL estimator (consistent with Theorem 3.2). This effect is largest when selection bias is severe. In that case the sampling variation in the AST estimate is just over one half that of the EL one. When $t(W)$ is (incorrectly) specified to be quadratic, this efficiency ranking reverses. In Design (b) the EL estimate exhibits lower sample variation than the corresponding AST estimate when $t(W)$ is (incorrectly) specified to be linear. When $t(W)$ is quadratic, which more closely approximates the efficient choice, this ranking is reversed. As before, the efficiency gains are increasing in the degree of selection bias. In terms of inference the AST Wald confidence intervals generally have actual coverage close to nominal coverage, while the corresponding EL ones tend to be conservative (Qin and Zhang (2008) suggest the use of bootstrap confidence intervals in order to improve coverage).

While Qin and Zhang (2008) do not consider the semiparametric efficiency properties of their procedure, the results in Table 4 suggest that, in contrast to AST, their estimator is not Locally Efficient at Assumption 3.1 (although this is only a conjecture based on the Monte Carlo results). Evidently the comparison of the two estimators when Assumption 3.1 does not hold is more complicated.

Table 5 calculates the variance bound for γ_0 for each of the Qin and Zhang (2008) Monte Carlo designs. The table also reports 'pencil and paper' calculations of the asymptotic variance of the AST estimator across each design and $t(W)$ combination.

Table 4: Monte Carlo results: Qin and Zhang (2008) designs with N =1,000

(β_1, β_2)		$t(W)$	(1) Mean Bias	(2) Sample Var.	(3) Mean Est. Var.	(4) RMSE	(5) Cov. of 95% CI
			Design (a): Linear CEFs				
$(0.1, 0.1)$	AST	Lin	-0.0004	0.0154	0.0151	0.1241	0.936
	AST	Qrd	-0.0083	0.0285	0.0513	0.1690	0.988
	EL	Lin	0.0038	0.0204	0.0311	0.1429	0.981
	EL	Qrd	0.0040	0.0241	0.0357	0.1553	0.978
$(0.2, 0.2)$	AST	Lin	-0.0065	0.0216	0.0195	0.1471	0.930
	AST	Qrd	-0.0039	0.0371	0.0555	0.1926	0.983
	EL	Lin	0.0031	0.0275	0.0402	0.1659	0.975
	EL	Qrd	-0.0009	0.0306	0.0430	0.1749	0.972
$(0.5, 0.5)$	AST	Lin	0.0024	0.0537	0.0428	0.2316	0.907
	AST	Qrd	0.0244	0.1015	0.0867	0.3193	0.920
	EL	Lin	0.0051	0.0900	0.7241	0.3000	0.912
	EL	Qrd	-0.0089	0.1103	0.5842	0.3322	0.891
			Design (b): Quadratic CEFs				
$(0.1, 0.1)$	AST	Lin	0.0009	0.3050	0.2856	0.5520	0.942
	AST	Qrd	-0.0011	0.0168	0.0174	0.1297	0.947
	EL	Lin	0.0347	0.1561	0.2003	0.3966	0.966
	EL	Qrd	0.0029	0.0226	0.1181	0.1504	0.995
$(0.2, 0.2)$	AST	Lin	0.0787	0.3620	0.3201	0.6065	0.916
	AST	Qrd	0.0078	0.0218	0.0217	0.1479	0.951
	EL	Lin	0.0477	0.1227	0.3790	0.3535	0.980
	EL	Qrd	0.0028	0.0309	0.4564	0.1758	0.998
$(0.5, 0.5)$	AST	Lin	0.1943	0.7010	0.4425	0.8591	0.817
	AST	Qrd	0.0095	0.0549	0.0429	0.2343	0.906
	EL	Lin	0.1969	0.2647	3.2656	0.5509	0.959
	EL	Qrd	0.0075	0.1026	2.1138	0.3204	0.993

Table 5: Theoretical properties of AST in Qin and Zhang (2008) Monte Carlo designs with N =1,000

	$t(W)$	(1.a) $I(\gamma_0)^{-1}/N$	(2.a) Asymptotic Variance	(1.b) $I(\gamma_0)^{-1}/N$	(2.b) Asymptotic Variance
		Design (a): Linear CEFs		**Design (b): Quadratic CEFs**	
$\beta_1= \beta_2= 0.1$	Linear	0.0114	0.0114	0.0114	0.2127
	Quadratic		0.0205		0.0125
$\beta_1= \beta_2= 0.2$	Linear	0.0144	0.0144	0.0144	0.2253
	Quadratic		0.0226		0.0154
$\beta_1= \beta_2= 0.5$	Linear	0.0373	0.0373	0.0373	0.4739
	Quadratic		0.0495		0.0386

F Detailed calculations

F.1 Additional calculations for the proof of Theorem 2.1

Equation (42) follows from the calculations:

$$\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{D}{p_0(X)}-1\right)p_0(X)q_s(W)\mathbb{S}_\delta(D,W)'\right]$$
$$=\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{D}{p_0(W)}-1\right)p_0(W)q_s(W)\frac{D-p(W,\delta_0)}{p(W,\delta_0)[1-p(W,\delta_0)]}p_\delta(W,\delta_0)'\right]$$
$$=\frac{1}{Q_0}\mathbb{E}\left[q_s(W)\frac{D-Dp(W,\delta_0)}{p(W,\delta_0)[1-p(W,\delta_0)]}p_\delta(W,\delta_0)'\right]-0$$
$$=\frac{1}{Q_0}\mathbb{E}\left[q_s(X)p_\delta(W,\delta_0)'\right]$$
$$=\frac{1}{Q_0}\int q_s(w)p_\delta(w,\delta_0)'f(w)\,\mathrm{d}w$$
$$=\int q_s(w)\frac{p_\delta(w,\delta_0)'}{p(w,\delta_0)}f(w|d=1)\,\mathrm{d}w$$
$$=\mathbb{E}\left[q_s(W)\frac{p_\delta(W,\delta_0)'}{p(W,\gamma_0)}\middle|D=1\right]$$
$$=\mathbb{E}\left[\psi_s(Y,W,\gamma_0)\frac{p_\delta(W,\delta_0)'}{p(W,\gamma_0)}\middle|D=1\right].$$

F.2 Additional calculations for the proof of Theorems 3.1, 3.2 and 3.3.

Derivation of (20)   To derive (20) in the main appendix we begin with the partition

$$M=\begin{bmatrix}M_{11}&0&0&0\\M_{21}&M_{22}&0&0\\M_{31}&0&M_{33}&0\\M_{41}&M_{42}&M_{43}&M_{44}\end{bmatrix}.$$

Note that

$$\begin{bmatrix}M_{11}&0&0\\M_{21}&M_{22}&0\\M_{31}&0&M_{33}\end{bmatrix}^{-1}=\begin{bmatrix}M_{11}^{-1}&0&0\\-M_{22}^{-1}M_{21}M_{11}^{-1}&M_{22}^{-1}&0\\-M_{33}^{-1}M_{31}M_{11}^{-1}&0&M_{33}^{-1}\end{bmatrix},$$

8

since

$$-\begin{pmatrix} M_{22}^{-1} & 0 \\ 0 & M_{33}^{-1} \end{pmatrix} \begin{pmatrix} M_{21} \\ M_{31} \end{pmatrix} M_{11}^{-1} = -\begin{pmatrix} M_{22}^{-1} M_{21} M_{11}^{-1} \\ M_{33}^{-1} M_{31} M_{11}^{-1} \end{pmatrix}.$$

This gives an inverse of

$$M^{-1} = \begin{pmatrix} M_{11}^{-1} & 0 & 0 & 0 \\ -M_{22}^{-1} M_{21} M_{11}^{-1} & M_{22}^{-1} & 0 & 0 \\ -M_{33}^{-1} M_{31} M_{11}^{-1} & 0 & M_{33}^{-1} & 0 \\ -M_{44}^{-1} \left(M_{41} M_{11}^{-1} - M_{42} M_{22}^{-1} M_{21} M_{11}^{-1} - M_{43} M_{33}^{-1} M_{31} M_{11}^{-1} \right) & -M_{44}^{-1} M_{42} M_{22}^{-1} & -M_{44}^{-1} M_{43} M_{33}^{-1} & M_{44}^{-1} \end{pmatrix}$$

since

$$-M_{44}^{-1} \begin{pmatrix} M_{41} & M_{42} & M_{43} \end{pmatrix} \begin{pmatrix} M_{11}^{-1} & 0 & 0 \\ -M_{22}^{-1} M_{21} M_{11}^{-1} & M_{22}^{-1} & 0 \\ -M_{33}^{-1} M_{31} M_{11}^{-1} & 0 & M_{33}^{-1} \end{pmatrix}$$

$$= -M_{44}^{-1} \begin{pmatrix} M_{41} M_{11}^{-1} - M_{42} M_{22}^{-1} M_{21} M_{11}^{-1} - M_{43} M_{33}^{-1} M_{31} M_{11}^{-1} & M_{42} M_{22}^{-1} & M_{43} M_{33}^{-1} \end{pmatrix}.$$

Straightforward matrix multiplication then gives the last K elements of $-M^{-1} m(Z, \theta)$ equal to

$$-M_{44}^{-1} m_4 (Z, \delta, \lambda_a, \lambda_s, \gamma) + M_{44}^{-1} \left(M_{41} M_{11}^{-1} - M_{42} M_{22}^{-1} M_{21} M_{11}^{-1} - M_{43} M_{33}^{-1} M_{31} M_{11}^{-1} \right) m_1 (Z, \delta)$$

$$+ M_{44}^{-1} M_{42} M_{22}^{-1} m_2 (Z, \delta, \lambda_a) + M_{44}^{-1} M_{43} M_{33}^{-1} m_3 (Z, \delta, \lambda_s)$$

$$= -M_{44}^{-1} \left\{ \begin{array}{c} m_4 (Z, \delta, \lambda_a, \lambda_s, \gamma) - \left(M_{41} M_{11}^{-1} - M_{42} M_{22}^{-1} M_{21} M_{11}^{-1} - M_{43} M_{33}^{-1} M_{31} M_{11}^{-1} \right) m_1 (Z, \delta) \\ -M_{42} M_{22}^{-1} m_2 (Z, \delta, \lambda_a) - M_{43} M_{33}^{-1} m_3 (Z, \delta, \lambda_s) \end{array} \right\}$$

$$= -M_{44}^{-1} \left\{ \begin{array}{c} m_4 (Z, \delta, \lambda_a, \lambda_s, \gamma) - M_{41} M_{11}^{-1} m_1 (Z, \delta) \\ + M_{42} M_{22}^{-1} \left(M_{21} M_{11}^{-1} m_1 (Z, \delta) - m_2 (Z, \delta, \lambda_a) \right) + M_{43} M_{33}^{-1} \left(M_{31} M_{11}^{-1} m_1 (Z, \delta) - m_3 (Z, \delta, \lambda_s) \right) \end{array} \right\},$$

from which the result directly follows.

Derivation of (22) To derive (22) we start with the direct derivative calculation

$$M_{21} = \frac{\partial}{\partial \delta'} \frac{1}{Q_0} \mathbb{E} \left[\left(\frac{1 - D}{1 - G \left(r(W)' \delta + t(W)' \lambda_a \right)} - 1 \right) G \left(r(W)' \delta \right) t(W) \right]$$

$$= \frac{1}{Q_0} \mathbb{E} \left[\left(\frac{1 - D}{\left[1 - G \left(r(W)' \delta + t(W)' \lambda_a \right) \right]^2} \right) G \left(r(W)' \delta \right) G_1 \left(r(W)' \delta + t(W)' \lambda_a \right) t(W) r(W)' \right] + 0$$

$$= \frac{1}{Q_0} \mathbb{E} \left[\frac{G \left(r(W)' \delta \right)}{1 - G \left(r(W)' \delta + t(W)' \lambda_a \right)} G_1 \left(r(W)' \delta + t(W)' \lambda_a \right) t(W) r(W)' \right],$$

now observe that

$$\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}-1\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)m_1\left(Z,\delta\right)'\right]$$

$$=\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}-1\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\left(\frac{D-G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1-G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)r\left(W\right)\right)'\right]$$

$$=\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\left(\frac{D-G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1-G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)r\left(W\right)\right)'\right]-0$$

$$=0-\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\frac{G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1-G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)r\left(W\right)'\right]$$

$$=-\frac{1}{Q_0}\mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}G_1\left(r\left(W\right)'\delta\right)t\left(W\right)r\left(W\right)'\right],$$

which gives (22) after recalling that the population value of λ_a is zero.

Derivation of (24) and (25)   Equations (24) and (25) follow from

$$M_{22}=\frac{\partial}{\partial\lambda_a'}\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}-1\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\right]$$

$$=\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{1-D}{\left[1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)\right]^2}\right)G\left(r\left(W\right)'\delta\right)G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)t\left(W\right)t\left(W\right)'\right]$$

$$=\mathbb{E}\left[\frac{1}{Q_0}\frac{G\left(r\left(W\right)'\delta\right)}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)t\left(W\right)t\left(W\right)'\right].$$

and

$$M_{42}=\frac{\partial}{\partial\lambda_a'}\mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\left[\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}\psi_s\left(Y,W,\gamma_0\right)-\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}\psi_a\left(X,W,\gamma_0\right)\right]\right]$$

$$=-\mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\frac{1-D}{\left[1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)\right]^2}G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)\psi_a\left(X,W,\gamma_0\right)t\left(W\right)'\right]$$

$$=-\frac{1}{Q_0}\mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)\psi_a\left(X,W,\gamma_0\right)t\left(W\right)'\right].$$

Derivation of (23)   To derive (23) we start with the direct derivative calculation

$$M_{31}=\frac{\partial}{\partial\delta'}\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}-1\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\right]$$

$$=-\frac{1}{Q_0}\mathbb{E}\left[\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)^2}G\left(r\left(W\right)'\delta\right)G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)t\left(W\right)r\left(W\right)'\right]+0$$

$$=-\frac{1}{Q_0}\mathbb{E}\left[G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)t\left(W\right)r\left(W\right)'\right].$$

Next observe that

$$\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}-1\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\left(\frac{D-G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1-G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)r\left(W\right)\right)'\right]$$

$$=\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}\right)G\left(r\left(W\right)'\delta\right)\left(\frac{D-G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1-G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)\right)t\left(W\right)r\left(W\right)'\right]-0$$

$$=\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}\right)\left(\frac{D-DG\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1-G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)\right)t\left(W\right)r\left(W\right)'\right]$$

$$=\frac{1}{Q_0}\mathbb{E}\left[\left(\frac{G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}\right)G_1\left(r\left(W\right)'\delta\right)t\left(W\right)r\left(W\right)'\right]$$

$$=\frac{1}{Q_0}\mathbb{E}\left[G_1\left(r\left(W\right)'\delta\right)t\left(W\right)r\left(W\right)'\right]$$

which gives (23) after recalling that the population value of λ_s is zero.

Derivation of (26) and (27) Equations (26) and (27) follow from

$$
\begin{aligned}
M_{33} &= \frac{\partial}{\partial\lambda_s'}\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}-1\right)G\left(r\left(W\right)'\delta\right)t\left(W\right)\right]\\
&= -\mathbb{E}\left[\frac{1}{Q_0}\left(\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)^2}\right)G\left(r\left(W\right)'\delta\right)G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)t\left(W\right)t\left(W\right)'\right]\\
&= -\frac{1}{Q_0}\mathbb{E}\left[G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)t\left(W\right)t\left(W\right)'\right].
\end{aligned}
$$

and

$$
\begin{aligned}
M_{43} &= \frac{\partial}{\partial\lambda_s'}\mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\left[\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)}\psi_s\left(Y,W,\gamma_0\right)-\frac{1-D}{1-G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_a\right)}\psi_a\left(X,W,\gamma_0\right)\right]\right]\\
&= -\mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\frac{D}{G\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)^2}G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)\psi_s\left(X,W,\gamma_0\right)t\left(W\right)'\right]\\
&= -\frac{1}{Q_0}\mathbb{E}\left[G_1\left(r\left(W\right)'\delta+t\left(W\right)'\lambda_s\right)\psi_s\left(X,W,\gamma_0\right)t\left(W\right)'\right].
\end{aligned}
$$

Derivation of (21) To derive (21) we start with the direct derivative calculation

$$
\begin{aligned}
M_{41} &= \frac{\partial}{\partial \delta'} \mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\left[\frac{D}{G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)}\psi_s\left(Y, W, \gamma_0\right) - \frac{1-D}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\psi_a\left(X, W, \gamma_0\right)\right]\right] \\
&= \mathbb{E}\left[\frac{G_1\left(r\left(W\right)'\delta\right)}{Q_0}\left[\frac{D}{G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)}\psi_s\left(Y, W, \gamma_0\right) - \frac{1-D}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\psi_a\left(X, W, \gamma_0\right)\right]r\left(W\right)'\right] \\
&\quad + \mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\left[-\frac{D}{G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)^2}\psi_s\left(Y, W, \gamma_0\right)\right.\right. \\
&\quad\quad \left.\left. - \frac{1-D}{\left[1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)\right]^2}\psi_a\left(X, W, \gamma_0\right)\right]G_1\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)r\left(W\right)'\right] \\
&= \frac{1}{Q_0}\mathbb{E}\left[\left[\frac{D}{G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)}G_1\left(r\left(W\right)'\delta\right) - \frac{D}{G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)}G_1\left(r\left(W\right)'\delta\right)\right]\psi_s\left(Y, W, \gamma_0\right)r\left(W\right)'\right] \\
&\quad - \frac{1}{Q_0}\mathbb{E}\left[\frac{1-D}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}G_1\left(r\left(W\right)'\delta\right)\psi_a\left(X, W, \gamma_0\right)r\left(W\right)'\right] \\
&\quad - \mathbb{E}\left[\frac{G\left(r\left(W\right)'\delta\right)}{Q_0}\left[\frac{1-D}{\left[1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)\right]^2}\psi_a\left(X, W, \gamma_0\right)\right]G_1\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_s\right)r\left(W\right)'\right] \\
&= 0 - \frac{1}{Q_0}\mathbb{E}\left[\frac{\left(1-D\right)\left[1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right) + G\left(r\left(W\right)'\delta\right)\right]}{\left[1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)\right]^2}G_1\left(r\left(W\right)'\delta\right)\psi_a\left(X, W, \gamma_0\right)r\left(W\right)'\right] \\
&= 0 - \frac{1}{Q_0}\mathbb{E}\left[\frac{\left(1-D\right)}{\left[1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)\right]^2}G_1\left(r\left(W\right)'\delta\right)\psi_a\left(X, W, \gamma_0\right)r\left(W\right)'\right] \\
&= -\frac{1}{Q_0}\mathbb{E}\left[\frac{G_1\left(r\left(W\right)'\delta\right)}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\psi_a\left(X, W, \gamma_0\right)r\left(W\right)'\right].
\end{aligned}
$$

Now observe that

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{Q_0}\left[\frac{1-D}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\right]\psi_a\left(X, W, \gamma_0\right)m_1\left(Z, \delta\right)'\right] \\
&= \frac{1}{Q_0}\mathbb{E}\left[\frac{1-D}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\psi_a\left(X, W, \gamma_0\right)\left(\frac{D - G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1 - G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)r\left(W\right)\right)'\right] \\
&= -\frac{1}{Q_0}\mathbb{E}\left[\frac{1-D}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\psi_a\left(X, W, \gamma_0\right)\left(\frac{G\left(r\left(W\right)'\delta\right)}{G\left(r\left(W\right)'\delta\right)\left[1 - G\left(r\left(W\right)'\delta\right)\right]}G_1\left(r\left(W\right)'\delta\right)r\left(W\right)\right)'\right] \\
&= -\frac{1}{Q_0}\mathbb{E}\left[\left[\frac{G_1\left(r\left(W\right)'\delta\right)}{1 - G\left(r\left(W\right)'\delta + t\left(W\right)'\lambda_a\right)}\right]\psi_a\left(X, W, \gamma_0\right)r\left(W\right)'\right].
\end{aligned}
$$

which gives (21).

Details of final steps of the derivation of the influence function appearing in Theorem 3.1 The influence function (20) consists of four parts. The first part, $m_4\left(Z_i, \delta_0, \lambda_{a0}, \lambda_{s0}, \gamma_0\right) - M_{41}M_{11}^{-1}m_1\left(Z_i, \delta_0\right)$, using expressions derived previously, equals, after some rearrangement,

$$
\begin{aligned}
&\frac{D}{Q_0}\left\{\psi_s\left(Y, W, \gamma_0\right) - q_s\left(W\right)\right\} - \frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1 - p_0\left(W\right)}\left\{\psi_a\left(X, W, \gamma_0\right) - q_a\left(W\right)\right\} \\
&+ \frac{D}{Q_0}q_s\left(W\right) - \frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1 - p_0\left(W\right)}q_a\left(W\right) \\
&+ \frac{1}{Q_0}\mathbb{E}^*\left[\left.\frac{1-D}{1 - p_0\left(W\right)}q_a\left(W\right)\right|\mathbb{S}_\delta\right].
\end{aligned} \tag{54}
$$

The second term, $M_{42}M_{22}^{-1}\left(M_{21}M_{11}^{-1}m_1\left(Z_i,\delta_0\right)-m_2\left(Z_i,\delta_0,\lambda_{a0}\right)\right)$, equals

$$-\frac{\Pi_a^*}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)t\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)t\left(W\right)\right\} \tag{55}$$

$$=\quad-\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a^*\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a^*\left(W\right)\right\}$$

$$=\quad-\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)\right\}$$

$$-\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)\{q_a^*\left(W\right)-q_a\left(W\right)\}\bigg|\mathbb{S}_\delta\right]\right\}$$

$$+\frac{1}{Q_0}\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)\{q_a^*\left(W\right)-q_a\left(W\right)\}.$$

The third term, $M_{43}M_{33}^{-1}\left(M_{31}M_{11}^{-1}m_1\left(Z_i,\delta_0\right)-m_3\left(Z_i,\delta_0,\lambda_{s0}\right)\right)$, equals

$$\frac{\Pi_s^*}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)t\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)t\left(W\right)\right\} \tag{56}$$

$$=\quad\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s^*\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s^*\left(W\right)\right\}$$

$$=\quad\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\right\}$$

$$+\frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)\{q_s^*\left(W\right)-q_s\left(W\right)\}\bigg|\mathbb{S}_\delta\right]$$

$$-\frac{1}{Q_0}\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)\{q_s^*\left(W\right)-q_s\left(W\right)\}.$$

Adding (54), (55) and (56) yields

$$\frac{D}{Q_0}\{\psi_s\left(Y,W,\gamma_0\right)-q_s\left(W\right)\}-\frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1-p_0\left(W\right)}\{\psi_a\left(X,W,\gamma_0\right)-q_a\left(W\right)\}$$

$$+\frac{D}{Q_0}q_s\left(W\right)-\frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1-p_0\left(W\right)}q_a\left(W\right)$$

$$+\frac{1}{Q_0}\mathbb{E}^*\left[\frac{1-D}{1-p_0\left(W\right)}q_a\left(W\right)\bigg|\mathbb{S}_\delta\right]$$

$$-\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)\right\}$$

$$+\frac{1}{Q_0}\left\{\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\bigg|\mathbb{S}_\delta\right]-\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\right\}$$

$$+R_s\left(D,W\right)-R_a\left(D,W\right)$$

$$=\quad\frac{D}{Q_0}\{\psi_s\left(Y,W,\gamma_0\right)-q_s\left(W\right)\}-\frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1-p_0\left(W\right)}\{\psi_a\left(X,W,\gamma_0\right)-q_a\left(W\right)\} \tag{57}$$

$$+\frac{D}{Q_0}q_s\left(W\right)-\frac{1}{Q_0}\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)$$

$$-\frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1-p_0\left(W\right)}q_a\left(W\right)+\frac{1}{Q_0}\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)$$

$$+\frac{1}{Q_0}\mathbb{E}^*\left[\frac{1-D}{1-p_0\left(W\right)}q_a\left(W\right)\bigg|\mathbb{S}_\delta\right]-\frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)q_a\left(W\right)\bigg|\mathbb{S}_\delta\right]$$

$$+\frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{D}{p_0\left(W\right)}-1\right)p_0\left(W\right)q_s\left(W\right)\bigg|\mathbb{S}_\delta\right]+R_s\left(D,W\right)-R_a\left(D,W\right).$$

By linearity of the LP operator and the conditional mean zero property of the score (i.e., $\mathbb{E}\left[\mathbb{S}_\delta \mid W\right] = 0$ implies that $\mathbb{S}_\delta$ and $q_a(W)$ are uncorrelated).

$$\frac{1}{Q_0}\mathbb{E}^*\left[\frac{1-D}{1-p(W)}q_a(W)\,\middle|\,\mathbb{S}_\delta\right] - \frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{1-D}{1-p_0(W)}-1\right)p_0(W)q_a(W)\,\middle|\,\mathbb{S}_\delta\right]$$

$$= \frac{1}{Q_0}\mathbb{E}^*\left[\frac{1-D}{1-p(W)}q_a(W) - \left(\frac{1-D}{1-p_0(W)}-1\right)p_0(W)q_a(W)\,\middle|\,\mathbb{S}_\delta\right]$$

$$= \frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{(1-D)(1-p_0(W))}{1-p_0(W)}+p_0(W)\right)q_a(W)\,\middle|\,\mathbb{S}_\delta\right]$$

$$= -\frac{1}{Q_0}\mathbb{E}^*\left[\left(-1+(D-p_0(W))\right)q_a(W)\mid\mathbb{S}_\delta\right]$$

$$-\frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{D}{p_0(W)}-1\right)p_0(W)q_a(W)\,\middle|\,\mathbb{S}_\delta\right].$$

We also have that

$$\frac{D}{Q_0}q_s(W) - \frac{1}{Q_0}\left(\frac{D}{p_0(W)}-1\right)p_0(W)q_s(W)$$

$$= \frac{1}{Q_0}\left\{D - \left(\frac{D}{p_0(W)}-1\right)p_0(W)\right\}q_s(W)$$

$$= \frac{1}{Q_0}\left\{D - (D-p_0(W))\right\}q_s(W)$$

$$= \frac{p_0(W)}{Q_0}q_s(W)$$

and

$$-\frac{1-D}{Q_0}\frac{p_0(W)}{1-p_0(W)}q_a(W) + \frac{1}{Q_0}\left(\frac{1-D}{1-p_0(W)}-1\right)p_0(W)q_a(W)$$

$$= \frac{1}{Q_0}\left\{-(1-D)\frac{p_0(W)}{1-p_0(W)}+\left(\frac{1-D}{1-p_0(W)}-1\right)p_0(W)\right\}q_a(W)$$

$$= \frac{1}{Q_0}\left\{-(1-D)\frac{p_0(W)}{1-p_0(W)}+\left((1-D)\frac{p_0(W)}{1-p_0(W)}-p_0(W)\right)\right\}q_a(W)$$

$$= -\frac{p_0(W)}{Q_0}q_a(W).$$

Using these results we get (57), less the inverse Jacobian component, equal to

$$\frac{D}{Q_0}\left\{\psi_s(Y,W,\gamma_0)-q_s(W)\right\} - \frac{1-D}{Q_0}\frac{p_0(W)}{1-p_0(W)}\left\{\psi_a(X,W,\gamma_0)-q_a(W)\right\}$$

$$+\frac{p_0(W)}{Q_0}\left\{q_s(W)-q_a(W)\right\}$$

$$+\frac{1}{Q_0}\mathbb{E}^*\left[\left(\frac{D}{p_0(W)}-1\right)p_0(W)\left\{q_s(W)-q_a(W)\right\}\,\middle|\,\mathbb{S}_\delta\right]$$

$$+R_s(D,W)-R_a(D,W),$$

and hence the claimed form of $\phi^{\mathrm{AST}}(Z,\gamma_0)$.

Derivation of (30). To derive (30) we manipulate:

$$\mathbb{E}\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)\left\{q_a^*\left(W\right)-q_a\left(W\right)\right\}\mathbb{S}_\delta'\right]$$

$$=\mathbb{E}\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)\left\{q_a^*\left(W\right)-q_a\left(W\right)\right\}\frac{D-p_0\left(W\right)}{p_0\left(W\right)\left(1-p_0\left(W\right)\right)}G_1\left(r\left(W\right)'\delta_0\right)r\left(W\right)'\right]$$

$$=\mathbb{E}\left[\left(\frac{1-D-\left(1-p_0\left(W\right)\right)}{1-p_0\left(W\right)}\right)p_0\left(W\right)\left\{q_a^*\left(W\right)-q_a\left(W\right)\right\}\frac{D-p_0\left(W\right)}{p_0\left(W\right)\left(1-p_0\left(W\right)\right)}G_1\left(r\left(W\right)'\delta_0\right)r\left(W\right)'\right]$$

$$=-\mathbb{E}\left[\frac{D-p_0\left(W\right)}{1-p_0\left(W\right)}p_0\left(W\right)\left\{q_a^*\left(W\right)-q_a\left(W\right)\right\}\frac{D-p_0\left(W\right)}{p_0\left(W\right)\left(1-p_0\left(W\right)\right)}G_1\left(r\left(W\right)'\delta_0\right)r\left(W\right)'\right]$$

$$=-\mathbb{E}\left[\left\{q_a^*\left(W\right)-q_a\left(W\right)\right\}\frac{\left(D-p_0\left(W\right)\right)^2}{\left(1-p_0\left(W\right)\right)^2}G_1\left(r\left(W\right)'\delta_0\right)r\left(W\right)'\right]$$

$$=-\mathbb{E}\left[\left\{q_a^*\left(W\right)-q_a\left(W\right)\right\}\frac{p_0\left(W\right)}{1-p_0\left(W\right)}G_1\left(r\left(W\right)'\delta_0\right)r\left(W\right)'\right].$$

Derivation of (31). To derive (31) we begin by calculating

$$\mathbb{V}\left(R_s\left(D,W\right)\right)=\frac{1}{Q_0^2}\mathbb{E}\left[\left(\frac{D}{p_0\left(W\right)}-1\right)^2p_0\left(W\right)^2\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)'\right]$$

$$=\frac{1}{Q_0^2}\mathbb{E}\left[p_0\left(W\right)\left(1-p_0\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)'\right]$$

$$\mathbb{V}\left(R_a\left(D,W\right)\right)=\frac{1}{Q_0^2}\mathbb{E}\left[\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)^2p_0\left(W\right)^2\left(q_a\left(W\right)-q_a^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right]$$

$$=\frac{1}{Q_0^2}\mathbb{E}\left[\left(D-p_0\left(W\right)\right)^2\frac{p_0\left(W\right)^2}{\left(1-p_0\left(W\right)\right)^2}\left(q_a\left(W\right)-q_a^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right]$$

$$=\frac{1}{Q_0^2}\mathbb{E}\left[\frac{p_0\left(W\right)^3}{1-p_0\left(W\right)}\left(q_a\left(W\right)-q_a^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right]$$

$$\mathbb{C}\left(R_s\left(D,W\right),R_a\left(D,W\right)\right)=\frac{1}{Q_0^2}\mathbb{E}\left[\left(\frac{D}{p_0\left(W\right)}-1\right)\left(\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)^2\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right]$$

$$=-\frac{1}{Q_0^2}\mathbb{E}\left[\left(\frac{D}{p_0\left(W\right)}+\frac{1-D}{1-p_0\left(W\right)}-1\right)p_0\left(W\right)^2\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right]$$

$$=-\frac{1}{Q_0^2}\mathbb{E}\left[p_0\left(W\right)^2\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right],$$

which then gives (assuming that $\dim\left(\gamma_0\right)=1$ to keep the expressions compact)

$$\mathbb{V}\left(R_s\left(D,W\right)-R_a\left(D,W\right)\right)$$

$$=\mathbb{E}\left[\frac{p_0\left(W\right)^2}{Q_0^2}\left\{\frac{1-p_0\left(W\right)}{p_0\left(W\right)}\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)'\right.\right.$$

$$+\frac{p_0\left(W\right)}{1-p_0\left(W\right)}\left(q_a\left(W\right)-q_a^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'$$

$$\left.\left.+2\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right\}\right]$$

$$=\mathbb{E}\left[\frac{1}{Q_0^2}\frac{p_0\left(W\right)^2}{p_0\left(W\right)\left(1-p_0\left(W\right)\right)}\left\{\left(1-p_0\left(W\right)\right)^2\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)'\right.\right.$$

$$+p_0\left(W\right)^2\left(q_a\left(W\right)-q_a^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'$$

$$\left.\left.+2p_0\left(W\right)\left(1-p_0\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)'\right\}\right]$$

$$=\frac{1}{Q_0^2}\mathbb{E}\left[\frac{p_0\left(W\right)}{1-p_0\left(W\right)}\left\{\left(1-p_0\left(W\right)\right)\left(q_s\left(W\right)-q_s^*\left(W\right)\right)+p_0\left(W\right)\left(q_a\left(W\right)-q_a^*\left(W\right)\right)\right\}^2\right]$$

as asserted.

References

[1] Bickel, Peter J., Chris A.J. Klaassen, Ya'acov Ritov and Jon A. Wellner. (1993). *Efficient and adaptive estimation for semiparametric models.* New York: Springer-Verlag, Inc.

[2] Chen, Xiaohong, Han Hong and Alessandro Tarozzi. (2008). "Semiparametric efficiency in GMM models with auxiliary data," *Annals of Statistics* 36 (2): 808 - 843.

[3] Darity; William A. and Patrick L. Mason. (1998). "Evidence on discrimination in employment: codes of color, codes of gender," *Journal of Economic Perspectives* 12 (2): 63 - 90.

[4] Owen, Art. B. (2001). *Empirical Likelihood.* New York: Chapman & Hall/CRC.