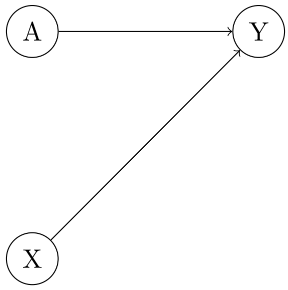
The data is modeled as:

$$
\begin{aligned}
(X_1, Y_1, A_1), \ldots, (X_n, Y_n, A_n) &\overset{iid}{\sim} \mathcal{O} \\
A &\perp X \\
P(A = 1 \mid X) = P(A = 1) = 1 - P(A = 0) &= p
\end{aligned} \tag{1}
$$

for some law $\mathcal{O}$.




The estimand is

$$\psi_0 = E(Y \mid A = 1) - E(Y \mid A = 0).$$

An estimator is obtained as the solution in $\psi$ of

$$\sum_{i=1}^{n} U(Y_i, A_i; \psi) = 0,$$

where

$$U(Y, A; \psi) = (A - p)(Y - \psi A).$$

Consistency and asymptotic normality of this estimator follow from:

**Lemma 0.1.** $E(U(Y, A; \psi_0)) = 0$.

*Proof.*

$$
\begin{aligned}
E(U(Y, A; \psi_0)) &= E[(A - p)(Y - \psi_0 A)] \\
&= E[(A - p)(E(Y \mid A) - \psi_0 A)] \\
&= (E(Y \mid A = 1) - \psi_0 A)(1 - p)p + E(Y \mid A = 0)(-p)(1 - p) \\
&= p(1 - p)[E(Y \mid A = 1) - E(Y \mid A = 0) - \psi_0] = 0.
\end{aligned}
$$

$\square$

We consider estimators obtained as solutions in $\psi$ to equations of the form

$$\sum_i U(Y_i, A_i; \psi) + (A_i - p)h(X_i; \psi) = \sum_i (A_i - p)(Y_i - \psi A_i + h(X_i; \psi)) = 0 \tag{2}$$

for [arbitrary] functions $h$. It follows from Lemma 0.1 and (1) that

$$E(U(Y, A; \psi_0) + (A - p)h(X; \psi)) = 0,$$

so such estimators are also asymptotically normal. An additional benefit is that the asymptotic variance of the resulting estimator may be minimized by varying $h$, perhaps improving on the efficiency of the estimator obtained from $\sum_i U(Y, A; \psi) = 0$. In fact, the minimizing choice of $h$ is determined by the estimating equation given by

$$W(X, Y, A; \psi) = U(Y, A; \psi) - E(U(Y, A; \psi) \mid A, X) + E(U(Y, A; \psi) \mid X). \tag{3}$$

A proof is given in Lemma 0.2, after rewriting the rhs of (3), as follows. The middle term on the rhs of (3) is,

$$\begin{aligned}
E(U(Y,A;\psi)\mid A,X) &= E((A-p)(Y-\psi A)\mid A,X)\\
&= (E(Y\mid A,X)-\psi A)(A-p)\\
&= [E(Y\mid A=1,X)A+E(Y\mid A=0,X)(1-A)](A-p)-\psi A(A-p)\\
&= E(Y\mid A=1,X)A(1-p)-E(Y\mid A=0,X)(1-A)p-\psi A(1-p).
\end{aligned}$$

The last term on the rhs of (3) is then,

$$\begin{aligned}
E(U(Y,A;\psi)\mid X) &= pE(U(Y,A;\psi)\mid A=1,X)+(1-p)E(U(Y,A;\psi)\mid A=0,X)\\
&= p[E(Y\mid A=1,X)(1-p)-\psi(1-p)]-(1-p)[E(Y\mid A=0,X)p]\\
&= p(1-p)(E(Y\mid A=1,X)-E(Y\mid A=0,X)-\psi).
\end{aligned}$$

Therefore,

$$\begin{aligned}
W(X,Y,A;\psi) &= U(Y,A;\psi)-E(U(Y,A;\psi)\mid A,X)+E(U(Y,A;\psi)\mid X)\\
&= (A-p)(Y-\psi A)-(A-p)(1-p)E(Y\mid A=1,X)-(A-p)pE(Y\mid A=0,X)+\\
&\quad\ (A-p)(1-p)\psi\\
&= (A-p)[Y-(1-p)E(Y\mid A=1,X)-pE(Y\mid A=0,X)]-p(1-p)\psi\\
&= (A-p)[Y-E(\tilde{Y}\mid X)]-p(1-p)\psi\\
&= U(Y_i,A_i;\psi)+(A-p)[\psi A-E(\tilde{Y}\mid X)]-p(1-p)\psi,
\end{aligned}$$

where $\tilde{Y}$ is determined by the transformation

$$Y=\tilde{Y}\frac{p^A(1-p)^{1-A}}{(1-p)^A p^{1-A}}=\tilde{Y}\left(\frac{p}{1-p}\right)^{2A-1}.$$

In case $p=P(A=1)=P(A=0)=1/2$,

$$W(X,Y,A;\psi)=(A-1/2)(Y-E(Y\mid X))-\psi/4. \tag{4}$$

**Lemma 0.2.** *The asymptotic variance of the estimator obtained as the solution in $\psi$ to*

$$\sum_i U(Y_i,A_i;\psi)+(A_i-p)h(X_i;\psi)=\sum_i(A_i-p)(Y_i-\psi A_i+h(X_i;\psi))=0 \tag{5}$$

*is minimized over arbitrary functions $h$ of $X$ at*

$$h_0(X;\psi)=(A-p)[\psi A-E(\tilde{Y}\mid X)]-p(1-p)\psi.$$

*Proof.* We give the $p=P(A=1)=1/2$ case, in which case

$$h_0(X;\psi)=(A-1/2)[\psi A-E(Y\mid X)]-\psi/4.$$

Under suitable regularity conditions, the asymptotic variance of the solution to the estimating equation (5) is given by the variance of its influence function. Since

$$E\frac{\partial}{\partial\psi}[U(Y_i,A_i;\psi)+(A_i-p)h(X_i;\psi)]=E\frac{\partial}{\partial\psi}U(Y_i,A_i;\psi),$$

the influence function of (5) is

$$-\left(E\,\frac{\partial}{\partial\psi}U(Y,A;\psi)\bigg|_{\psi_0}\right)^{-1}(U(Y,A;\psi_0)+(A-1/2)h(X;\psi)).$$

Thus we wish to show

$$\mathrm{Var}\left[\left(E\frac{\partial}{\partial\psi}U(Y,A;\psi_0)\right)^{-1}(U(Y,A;\psi_0)+(A-1/2)h(X;\psi))\right] \geq$$
$$\mathrm{Var}\left[\left(E\frac{\partial}{\partial\psi}U(Y,A;\psi_0)\right)^{-1}(U(Y,A;\psi_0)+(A-1/2)h_0(X;\psi))\right]$$

or

$$\begin{aligned} &E[(A-1/2)^2h^2(X)]+2E[U(Y,A;\psi_0)(A-1/2)h(X;\psi)] \geq \\ &E[(A-1/2)^2h_0^2(X)]+2E[U(Y,A;\psi_0)(A-1/2)h_0(X;\psi)]. \end{aligned} \tag{6}$$

Since $A, X$ are uncorrelated, and noting that $(A-1/2)(-1)^{1-A}=1/2$, the lhs is

$$\begin{aligned} &E[(A-1/2)^2h^2(X)]+2E[U(Y,A;\psi_0)(A-1/2)h(X;\psi)] \\ &= \mathrm{Var}(A)Eh^2(X)+2E[(A-1/2)E(U(Y,A;\psi_0)h(X;\psi)\mid A)] \\ &= Eh^2(X)/4+2E[(A-1/2)E((Y-\psi_0A)(-1)^{1-A}h(X;\psi)\mid A)] \\ &= Eh^2(X)/4+E[E((Y-\psi_0A)h(X;\psi)\mid A)] \\ &= Eh^2(X)/4+E((Y-\psi_0/2)h(X;\psi)). \end{aligned}$$

We obtain an expression for the rhs by substituting $h(X;\psi):=h_0(X;\psi)=-(2E(Y\mid X)-\psi_0)$,

$$\begin{aligned} &E[(A-1/2)^2h_0^2(X)]+2E[U(Y,A;\psi_0)(A-1/2)h_0(X;\psi)] \\ &= Eh_0^2(X)/4+E((Y-\psi_0/2)h_0(X;\psi)) \\ &= E[h_0(X;\psi)(h_0(X;\psi)/4+Y-\psi_0/2)] \\ &= E[h_0(X;\psi)(-(2E(Y\mid X)-\psi_0)/4+E(Y\mid X)-\psi_0/2)] \\ &= E[h_0(X;\psi)(E(Y\mid X)/2-\psi_0/4)] \\ &= -Eh_0^2(X)/4 \qquad (7) \\ &= -E[E(Y\mid X)^2]+\psi_0EY-\psi_0^2/4. \end{aligned}$$

Thus (6), which we wish to show, becomes

$$Eh^2(X)/4+E((Y-\psi_0/2)h(X;\psi))+E[E(Y\mid X)^2]-\psi_0EY+\psi_0^2/4\geq 0.$$

This inequality follows by an application of the Cauchy-Schwarz inequality,

$$\begin{aligned} &Eh^2(X)/4+E((Y-\psi_0/2)h(X;\psi))+E[E(Y\mid X)^2]-\psi_0EY+\psi_0^2/4 \\ &= (1/4)E[(h(X;\psi)-\psi_0)^2]+E(Yh(X;\psi))+E[E(Y\mid X)^2]-\psi_0EY \\ &= (1/4)E[(h(X;\psi)-\psi_0)^2]+E[E(Y\mid X)^2]+E[E(Y\mid X)(h(X;\psi)-\psi_0)] \\ &\geq (1/4)E[(h(X;\psi)-\psi_0)^2]+E[E(Y\mid X)^2]-E[(h(X;\psi)-\psi_0)^2]^{1/2}E[E(Y\mid X)^2]^{1/2} \\ &= \{(1/2)E[(h(X;\psi)-\psi_0)^2]^{1/2}-E[E(Y\mid X)^2]^{1/2}\}^2\geq 0. \end{aligned}$$

□

**Remark.** *From (7), $Eh_0^2(X)/4$ is the reduction in the asymptotic variance gained by using (5) over (2).*

The expression for $W(X,Y,A;\psi)$ in (3) contains a term of the form $E(Y\mid A,X)$, generally requiring estimation, whereas the equivalent expression in (4) only requires $E(Y\mid X)$ to be estimated. We investigate whether the second form is more resistant to bias by an unscrupulous analyst.

Consider the following strategies for estimating $\psi$:

□

FIGURE 1. Comparison of the 3 methods of estimating $\psi$.

□

FIGURE 2. Coverage rates of 3 $\psi$ estimates when the sample size is $n = 1000$.

(1) The analyst first forms $2^{p+1}$ estimates of $E(Y \mid A, X)$ by fitting submodels of the linear model
$$E(Y \mid A, X) = A + X_1 + \ldots + X_p + \text{second-order interactions}.$$
For each estimate $E(Y \mid A, X)$, the analyst then obtains an estimate for $\psi$ by solving the estimating equation (3), using the model estimate of $E(Y \mid A_i, X_i), i = 1, \ldots, n$. From the resulting estimates for $\psi$, the analyst reports the largest.
(2) One analyst is given the control data
$$\{(Y_i, X_i) : A_i = 0\}$$
from which he forms estimates of $E(Y \mid A = 0, X)$ using submodels of the linear models
$$E(Y \mid X) = X_1 + \ldots + X_p + \text{second-order interactions}.$$
Another analyst estimates $E(Y \mid A = 1, X)$ analogously. The models are combined to estimate $E(Y \mid A, X)$ and obtain estimates of $\psi$ as before, from which the largest is chosen.
(3) The analyst proceeds as in (1), but omitting $A_i$ from his models. I.e., the analyst first forms estimates of $E(Y \mid X)$ by fitting submodels of the linear model
$$E(Y \mid X) = X_1 + \ldots + X_p + \text{second-order interactions}.$$
For each estimate $E(Y \mid X)$, the analyst then obtains an estimate for $\psi$ by solving the estimating equation (4), using the model estimate of $E(Y \mid X_i), i = 1, \ldots, n$. From the resulting estimates for $\psi$, the analyst reports the largest.
(4) *Not yet done.* As in (2), two analysts each obtain $2^p$ estimates of $E(Y \mid A = 1, X)$ and $E(Y \mid A = 1, X)$ separately, but now the maximum estimate of $\psi$ is obtained ranging over all $2^{p+1}$ pairs of the 2 analysts' sets of estimates.

The results of a simulation are in Figure 1. From the simulation, the reported $\psi$ estimates are less biased when computed under method (3) and (2) than method (1), though all methods are seriously biased. Also plotted is the average, rather than maximum $\psi$ estimate, over the sets of models considered under each method. These are close to the true value $\psi = 1$, particularly for larger $n$, as expected.

Since all methods are so biased, confidence intervals at this range of sample sizes ($\leq 300$) are useless. Increasing the sample size to 1000 allows some comparison of error rates. Figure 2 gives the coverage rates of the 3 methods for as the number of covariates given to the analysts ranges.

0.1. **Equivalence to regression estimator.** Define
$$\tilde{Y} = Y - E(Y \mid X)$$
and consider the regression
$$E(\tilde{Y} \mid A) = \beta_0 + \beta_1 A.$$

Then $\beta_0 + \beta_1 = E(\tilde{Y} \mid A = 1) = E(Y \mid A = 1) - E(Y)$ and $\beta_0 = E(\tilde{Y} \mid A = 0) = E(Y \mid A = 0) - E(Y) = E(Y \mid A = 0)/2 - E(Y \mid A = 1)/2$, so

$$\begin{aligned}\beta_0 &= -\psi/2\\ \beta_1 &= \psi.\end{aligned}$$

The influence function of $(\beta_0, \beta_1)$ is obtained as:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \begin{pmatrix} 1 \\ A_i \end{pmatrix} (\tilde{Y}_i - \hat{\beta}_0 - A_i\hat{\beta}_1)\\
&= \sum_{i=1}^{n} \left\{ \begin{pmatrix} 1 \\ A_i \end{pmatrix} (\tilde{Y}_i - \beta_0 - A_i\beta_1) + \begin{pmatrix} -1 & -A_i \\ -A_i & -A_i \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \right\}\\
n^{1/2} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} &= \left( \frac{1}{n} \sum_i \begin{pmatrix} 1 & A_i \\ A_i & A_i \end{pmatrix} \right)^{-1} n^{-1/2} \sum_i \begin{pmatrix} 1 \\ A_i \end{pmatrix} (\tilde{Y}_i - \beta_0 - A_i\beta_1)\\
&= \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}^{-1} n^{-1/2} \sum_i \begin{pmatrix} 1 \\ A_i \end{pmatrix} (\tilde{Y}_i - \beta_0 - A_i\beta_1) + o_P(1)\\
n^{1/2}(\hat{\beta}_1 - \beta_1) &= n^{-1/2} \sum_i \begin{pmatrix} -2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ A_i \end{pmatrix} (\tilde{Y}_i - \beta_0 - A_i\beta_1) + o_P(1)\\
&= 4n^{-1/2} \sum_i (A_i - 1/2)(\tilde{Y}_i - \beta_0 - A_i\beta_1) + o_P(1)\\
&= 4n^{-1/2} \sum_i (A_i - 1/2)(\tilde{Y}_i - (A_i - 1/2)\psi) + o_P(1)\\
&= 4n^{-1/2} \sum_i \{(A_i - 1/2)(Y - E(Y \mid X)) - \psi/4\} + o_P(1).
\end{aligned}$$

By comparison with (4), we find that $\hat{\psi}$, the augmented estimator, is asymptotically equivalent to $\hat{\beta}_1$.

On the other hand, the OLS solution to the above regression gives

$$\hat{\beta}_1 = \hat{\mathrm{Cov}}(\tilde{Y}, A)/\hat{\mathrm{Var}}(A) = \frac{\overline{A\tilde{Y}}/\bar{\tilde{Y}} - \bar{\tilde{Y}}}{1 - \bar{A}}$$

whereas from (4),

$$\hat{\psi} = \bar{\tilde{Y}}/2 - \overline{A\tilde{Y}},$$

so the two estimators are not equal.

## 1. Other estimands

As above, the full data is $Y_i^* = (Y_i^*(0), Y_i^*(1)), i = 1, \ldots, n$, the observed data is $(Y_i, A_i, X_i), i = 1, \ldots, n$, and we assume

$$\begin{aligned}&Y = AY^*(1) + (1 - A)Y^*(0),\\ &P(A = 1) = p \in (0, 1)\\ &A \perp X, A \perp Y^*.\end{aligned}$$

Besides the mean treatment difference $E(Y \mid A = 1) - E(Y \mid A = 0)$ discussed above, we consider other estimands:

(1) $\psi_0 = \log \frac{E(Y^*(1))}{E(Y^*(0))} = \log \frac{E(Y|A=1)}{E(Y|A=0)}$
(2) the slope in the model

$$\mathrm{logit}(AE(Y^*(1)) + (1 - A)E(Y^*(0))) = \mathrm{logit}(P(Y = 1 \mid A)) = \psi_0 + \psi_1 A,$$

for a binary-valued response $Y$

In each case, we obtain the efficient augmented influence function following the approach of [[Tsiatis ch. 13]]:

(1) obtain a full-data influence function $\phi^F(Y^*)$
(2) obtain an observed data influence function $\phi(Y, A, X)$ corresponding to $\phi^F$ under the mapping $\phi \mapsto E(\phi \mid Y^*)$
(3) compute the efficient augmentation term

$$\begin{aligned} h^*(Y, A, X) &= (A - p)(E(\phi \mid A = 1, X) - E(\phi \mid A = 0, X)) \\ &= E(\phi \mid A, X) - E(\phi \mid X) \end{aligned}$$

We then eliminate regressions on treatment level, i.e., the terms $E(Y \mid A = 1, X)$ and $E(Y \mid A = 0, X)$.

1.1. $\log \frac{E(Y|A=1)}{E(Y|A=0)}$. The problem is to estimate

$$\psi_0 = \log \frac{E(Y^*(1))}{E(Y^*(0))} = \log \frac{E(Y \mid A = 1)}{E(Y \mid A = 0)}.$$

A full-data estimator is given by the solution to

$$\sum_i (Y_i^*(1) - e^{\psi_0} Y_i^*(0)) = 0,$$

with influence function

$$\begin{aligned} \phi^F(Y, A, X; \psi) &= (e^{\psi} E(Y^*(0)))^{-1}(Y^*(1) - e^{\psi} Y^*(0)) \\ &= (E(Y^*(1)))^{-1}(Y^*(1) - e^{\psi} Y^*(0)). \end{aligned}$$

An influence functions $\phi$ of the observed data satisfies

$$\begin{aligned} E(Y_i^*(1))\phi(Y, A, X) &= \left(\frac{A}{p} - e^{\psi_0}\frac{1-A}{1-p}\right) Y + h(Y, A, X) \\ &= (A - p)\left(\frac{A}{(A-p)p} - e^{\psi_0}\frac{1-A}{(A-p)(1-p)}\right) Y + h(Y, A, X) \\ &= \frac{A-p}{p(1-p)}(A + e^{\psi_0}(1-A))Y + h(Y, A, X) \\ &= \frac{A-p}{p(1-p)} e^{(1-A)\psi_0} Y + h(Y, A, X), \end{aligned}$$

where $h$ satisfies $E(h(Y, A, X) \mid Y^*) = 0$. The minimizing $h$ is given by subtracting out

$$\begin{aligned} h^*(Y, A, X) &= (A - p)[E(\frac{A-p}{p(1-p)} e^{(1-A)\psi_0} Y \mid A = 1, X) - E(\frac{A-p}{p(1-p)} e^{(1-A)\psi_0} Y \mid A = 0, X)] \\ &= (A - p)[\frac{1}{p} E(Y \mid A = 1, X) + \frac{e^{\psi_0}}{1-p} E(Y \mid A = 0, X)]. \end{aligned}$$

The efficient influence function is therefore

$$\begin{aligned} \phi^*(Y, A, X) &= (E(Y^*(1)))^{-1}\phi(Y, A, X) - h^*(Y, A, X) \\ &= (E(Y^*(1)))^{-1}(A - p)\left(\frac{e^{(1-A)\psi_0}}{p(1-p)} Y - \frac{1}{p} E(Y \mid A = 1, X) - \frac{e^{\psi_0}}{1-p} E(Y \mid A = 0, X)\right). \end{aligned}$$

Let $\hat{\psi}_n$ be a consistent estimator of $\psi_0$. Under the transformation

$$Y = \frac{p^{2A}(1-p)^{2(1-A)}}{e^{(1-A)\hat{\psi}_n}} \tilde{Y}$$

the efficient influence function may be rewritten

$$\phi^*(Y, A, X) = (E(Y^*(1)))^{-1}(A - p)\left(\frac{e^{(1-A)\psi}}{p(1-p)}Y - E(\tilde{Y} \mid X)\right) + o_P(1).$$

In case $p = 1/2$, $\tilde{Y} = 4e^{(1-A)\hat{\psi}_n}Y$, and

(8)

$$\phi^*(Y, A, X) = 2(E(Y^*(1)))^{-1}(2A - 1)[e^{(1-A)\psi}Y - E(e^{(1-A)\hat{\psi}_n}Y \mid X)] + o_P(1).$$

1.1.1. *Two-step regression.* Let

$$Z = Y - e^{(A-1)\psi_0}[E(e^{(1-A)\psi_0}Y \mid X) - E(Y \mid A = 1)]$$

and consider the log-linear regression model

(9) $$\log(E(Z \mid A)) = \beta_0 + \beta_1 A.$$

Then

$$E(e^{(1-A)\psi_0}Y) = (1/2)[e^{\psi_0}E(Y \mid A = 0) + E(Y \mid A = 1)] = E(Y \mid A = 1)$$

implies

$$\begin{aligned}
E(Z \mid A = 1) &= E(Y \mid A = 1) - E[E(e^{(1-A)\psi_0}Y \mid X) - E(Y \mid A = 1) \mid A = 1] \\
&= E(Y \mid A = 1) - E(e^{(1-A)\psi_0}Y) + E(Y \mid A = 1) \\
&= E(Y \mid A = 1),
\end{aligned}$$

and similarly

$$\begin{aligned}
E(Z \mid A = 0) &= E(Y \mid A = 0) - E[E(e^{(1-A)\psi_0}Y \mid X) - E(Y \mid A = 1) \mid A = 0] \\
&= E(Y \mid A = 0).
\end{aligned}$$

Therefore, under model (9),

$$\begin{aligned}
\beta_0 &= \log(E(Z \mid A = 0)) = \log(E(Y \mid A = 0)), \\
\beta_0 + \beta_1 &= \log(E(Z \mid A = 1)) = \log(E(Y \mid A = 1)), \\
\beta_1 &= \frac{\log(E(Y \mid A = 1))}{\log(E(Y \mid A = 0))} = \psi_0.
\end{aligned}$$

An estimator $(\hat{\beta}_0, \hat{\beta}_1)$ under the log-linear regression model (9) is given by the estimating equations

$$0 = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ A_i \end{pmatrix} (Z_i - e^{\hat{\beta}_0 + \hat{\beta}_1}).$$

The influence function of $\beta_1 = \psi_0$ is computed to be

$$\begin{aligned}
\phi_{\beta_1}(Y, A, X; \beta) &= 2(2A - 1)(e^{-\beta_0 - \beta_1 A}Z - 1) \\
&= 2(2A - 1)\left(\frac{Z}{E(Z \mid A)} - 1\right) \\
&= 2(2A - 1)\left(\frac{e^{(1-A)\beta_1}Z}{E(Z \mid A = 1)} - 1\right) \\
&= 2(2A - 1)\frac{e^{(1-A)\beta_1}Y - E(e^{(1-A)\psi_0}Y \mid X) + E(Y \mid A = 1) - E(Z \mid A = 1)}{E(Z \mid A = 1)} \\
&= 2(2A - 1)\frac{e^{(1-A)\beta_1}Y - E(e^{(1-A)\psi_0}Y \mid X)}{E(Y \mid A = 1)} \\
&= 2(E(Y^*(1)))^{-1}(2A - 1)[e^{(1-A)\psi_0}Y - E(e^{(1-A)\psi_0}Y \mid X)].
\end{aligned}$$

This influence function is the same as the influence function (8), so the estimator $\hat{\beta}_1$ under the log-linear regression is asymptotically equivalent to the efficient estimator $\hat{\psi}$.

1.2. $\text{logit}(P(Y = 1 \mid A)) = \psi_0 + \psi_1 A$. The target is the slope in the logistic model

$$\text{logit}(P(Y = 1 \mid A)) = \psi_0 + \psi_1 A.$$

This model is also a type of restricted moment model. Let $\sigma$ denote the sigmoid function $x \mapsto e^x/(1 + e^x)$ and $\sigma' = \sigma(1 - \sigma)$ its derivative. The coefficients may be estimated as the root of

$$\sum_{i=1}^{n} \begin{pmatrix} 1 \\ A_i \end{pmatrix} [Y_i - \sigma(\psi_0 + \psi_1 A_i)].$$

The influence function is computed in the usual way as

$$U(Y, A, X; \psi) = \frac{A - p}{p(1 - p)\sigma'(\psi_0)} \begin{pmatrix} A - 1 \\ [\sigma'(\psi_0)/\sigma'(\psi_0 + \psi_1)]^A \end{pmatrix} (Y - \sigma(\psi_0 + \psi_1 A)).$$

To minimize the asymptotic variance we subtract out $h^*(X)$ given by

$$\begin{aligned} (A - p)^{-1} h^*(X) &= E(U \mid A = 1, X) - E(U \mid A = 0, X) \\ &= \begin{pmatrix} -\frac{E(Y|A=0,X)-\sigma(\psi_0)}{(1-p)\sigma'(\psi_0)} \\ \frac{E(Y|A=1,X)-\sigma(\psi_0+\psi_1)}{p\sigma'(\psi_0+\psi_1)} + \frac{E(Y|A=0,X)-\sigma(\psi_0)}{(1-p)\sigma'(\psi_0)} \end{pmatrix}. \end{aligned}$$

The optimized influence function for the slope, $\phi^*_{\psi_1}$, is then given by

$$\begin{aligned} \frac{p(1 - p)}{A - p} \phi^*_{\psi_1}(Y, A, X; \psi) &= \frac{Y}{\sigma'(\psi_0 + \psi_1)^A \sigma'(\psi_0)^{1-A}} \\ &\quad - \left( \frac{(1 - p)E(Y \mid A = 1, X)}{\sigma'(\psi_0 + \psi_1)} + \frac{pE(Y \mid A = 0, X)}{\sigma'(\psi_0)} \right) \\ &\quad - (A - (1 - p)) \left( \frac{1}{1 - \sigma(\psi_0 + \psi_1)} - \frac{1}{1 - \sigma(\psi_0)} \right). \end{aligned}$$

Under the transformation

$$Y = \left( \frac{p}{1 - p} \right)^{2A-1} \sigma'(\psi_0 + \psi_1)^A \sigma'(\psi_0)^{1-A} \tilde{Y},$$

the optimized influence function may be rewritten as

$$\begin{aligned} \frac{p(1 - p)}{A - p} \phi^*_{\psi_1}(Y, A, X; \psi) &= \left( \frac{p}{1 - p} \right)^{2A-1} \tilde{Y} - E(\tilde{Y} \mid X) \\ &\quad - (A - (1 - p)) \left( \frac{1}{1 - \sigma(\psi_0 + \psi_1)} - \frac{1}{1 - \sigma(\psi_0)} \right). \end{aligned}$$

In case $p = 1/2$,

$$\phi^*_{\psi_1}(Y, A, X; \psi) = \frac{2(2A - 1)}{\sigma'(\psi_0 + \psi_1)^A \sigma'(\psi_0)^{1-A}} [Y - E(Y \mid X)] - \left( \frac{1}{1 - \sigma(\psi_0 + \psi_1)} - \frac{1}{1 - \sigma(\psi_0)} \right).$$