Blinder-Oaxaca as a Reweighting Estimator

Patrick Kline

UC Berkeley

ASSA Meetings, 2011

Regression vs. Reweighting

Several common approaches to adjusting for covariates:

- Regression based approaches (OLS, Blinder-Oaxaca)
- Propensity score methods (matching, reweighting)
- Doubly robust methods (Robins, Rotnitzky, and Zhao, 1994; Egel, Graham, and Pinto, 2009)

Today

- Study estimators of counterfactual mean

$$\mu_0^1 \equiv E\left[Y_i^0 | D_i = 1\right]$$

  - Key input to identification of

$$ATT \equiv E\left[Y_i^1 - Y_i^0 | D_i = 1\right]$$

- Show that a classic regression based approach, Blinder-Oaxaca estimation, is a DR estimator.
  - Under misspecification B-O provides MMSE approxiation to appropriate propensity score based weights.

Blinder-Oaxaca

Exogenous regime switching setup:

$$Y_i = Y_i^1 D_i + (1 - D_i)\, Y_i^0$$
$$Y_i^1 = X_i' \beta^1 + \varepsilon_i^1$$
$$Y_i^0 = X_i' \beta^0 + \varepsilon_i^0$$
$$E[\varepsilon_i^1 | X_i, D_i] = 0,\ E[\varepsilon_i^0 | X_i, D_i] = 0$$

Original application (Oaxaca, 1973): $\left(Y_i^1, Y_i^0\right)$ male/female wages and $\left(\beta^1, \beta^0\right)$ latent skill prices. Different prices imply discrimination.

Blinder-Oaxaca

B-O model allows identification of counterfactual means but not (without further assumptions) distributions. Mean independence of errors implies:

$$\begin{aligned}\mu_0^1 &\equiv E\left[Y_i^0 | D_i = 1\right] \\ &= E\left[X | D_i = 1\right]' \beta^0\end{aligned}$$

Likewise,

$$\beta^0 = E\left[X_i X_i' | D_i = 0\right]^{-1} E\left[X_i Y_i | D_i = 0\right]$$

given that $E\left[X_i X_i' | D_i = 0\right]$ is full rank. Hence,

$$\begin{aligned}\mu_0^1 &= E\left[X | D_i = 1\right]' \\ &\quad \times E\left[X_i X_i' | D_i = 0\right]^{-1} E\left[X_i Y_i | D_i = 0\right] \\ &\equiv \delta^{BO}\end{aligned}$$

Blinder-Oaxaca

B-O estimator simply replaces population quantity δ^{BO} with sample analogue – predicted values from a regression among the controls. Several advantages of this approach:

- Estimation simply requires less than perfect multicollinearity among X_i in the $D_i = 0$ sample. Useful in a number of evaluation designs where many more controls are available than treated units.
- Easy to conduct inference.
- Weakness: linear model may provide a poor fit at points far from $E[X_i]$.

Reweighting Approach

- Alternative approach: reweight controls by

$$\frac{dF_{X|D=1}(x)}{dF_{X|D=0}(x)}$$

  so that distribution of covariates among two samples is identical.

- By balancing distribution, the influence of these covariates will be removed.
- Then form estimate of counterfactual mean as $\int E[Y|X=x, D=0]\, dF_{X|D=1}(x)$.

Unconfoundedness:

$$Y_i^1, Y_i^0 \perp D_i | X_i$$

Stronger than earlier mean independence, but nonparametric about dependence of $\left(Y_i^1, Y_i^0\right)$ on X_i.

Unconfoundedness in B-O framework would require

$$E\left[g\left(\varepsilon_i^d\right)|X_i, D_i\right] = 0 \; d \in \{0,1\}$$

for any continuous function $g(.)$ not vanishing outside a finite interval.

Propensity Score

Propensity score (Rosenbaum and Rubin, 1983):

$$e(X_i) \equiv P(D_i = 1|X_i)$$

Overlap condition

$$e(X_i) < 1$$

Not directly testable without further assumptions.

Propensity Score Reweighting

Define

$$\pi \equiv P(D_i = 1)$$

$$w(X_i) \equiv \frac{1-\pi}{\pi} \frac{e(X_i)}{1-e(X_i)}$$

By Bayes' Rule

$$w(x) = \frac{dF_{X|D=1}(x)}{dF_{X|D=0}(x)}$$

Although $w(X_i)$ is distributed on $[0,\infty)$, refer to $w(X_i)$ as propensity score "weights" because

$$\begin{aligned} E\left[w(X_i)\,|\,D_i=0\right] &= \int w(x)\,dF_{X|D=0}(x) \\ &= \int dF_{X|D=1}(x) \\ &= 1 \end{aligned}$$

A useful result

Unconfoundedness and overlap imply:

$$
\begin{aligned}
\mu_0^1 &= E\left[\frac{e(X_i)}{\pi}\frac{1-D_i}{1-e(X_i)}Y_i\right] \\
&= E\left[w(X_i)\frac{1-D_i}{1-\pi}Y_i\right] \\
&= E\left[w(X_i)\,Y_i | D_i = 0\right]
\end{aligned}
$$

Hence, a weighted average of untreated outcomes identifies the counterfactual mean of interest μ_0^1.

Estimation

- Identification result motivates plug-in estimators where, typically, $e(X_i)$ is estimated via a flexible logit or probit model and π is chosen to ensure $E[w(X_i)|D_i = 0] = 1$ (Imbens, 2004; Hirano, Imbens, and Ridder, 2003).
- Useful in cases where researcher knows more about assignment mechanism than process generating outcomes.
- May be difficult to estimate propensity score in small samples or with unbalanced design (perfect prediction problem)
    - Problems may arise when estimated $e(X_i)$ is near one since lots of weight given to a few observations. (e.g. Kang and Schaeffer, 2007; Huber, Lechner, and Wunsch, 2010)

Equivalence

Given the overlap condition, it is straightforward to show that $E[X|D_i = 1] = E[w(X_i)X|D_i = 0]$ and hence that:

$$\begin{aligned}
\delta^{BO} &= E[w(X_i)X|D_i = 0]' \\
&\quad \times E\left[X_i X_i'|D_i = 0\right]^{-1} E[X_i Y_i|D_i = 0] \\
&= E[\widetilde{w}(X_i)Y_i] \\
\widetilde{w}(X_i) &\equiv X_i' E\left[X_i X_i'|D_i = 0\right]^{-1} E\left[X_i \frac{1-\pi}{\pi}\frac{e(X_i)}{1-e(X_i)}|D_i = 0\right]
\end{aligned}$$

Interpretation:

- B-O weights provide MMSE approximation to true nonparametric weights $w(X_i)$
- Approximation is exact if $\frac{e(X_i)}{1-e(X_i)} = X'\gamma$ (log-logistic) as opposed to logistic model which assumes $\frac{e(X_i)}{1-e(X_i)} = exp(X'\gamma)$

Double Robustness

- Result implies B-O estimator is "doubly robust" (Robins, Rotnitzky, and Zhao, 1994) – consistent if *either* log-logistic model for propensity score or linear model for $E\left[Y_i^0|X_i\right]$ is correct.
- Propensity score model justified by latent variable model of the form

$$D_i = 1\left[X_i'\gamma + v_i\right]$$

where $v_i \sim F_v(.)$ and $F_v(z) = \frac{z}{1+z}$.

Misspecification

- In practice, neither the outcome nor the propensity score model is likely to hold globally. Simply convenient local approximations.
- Bias in B-O estimator is:

$$\mu_0^1 - \delta^{BO} = E\left[\left(w\left(X_i\right) - \widetilde{w}\left(X_i\right)\right) Y_i | D_i = 0\right]$$

- Can show that $E\left[w\left(X_i\right) - \widetilde{w}\left(X_i\right)\right] = 0$, so bias emerges from correlation of specification errors with $E\left[Y_i^0 | X_i\right]$.

Misspecification

- B-O approximates the weights $w(X_i)$ directly, while typical plugin estimators approximate $e(X_i)$ and then form implied weights. Best approximation to $e(X_i)$ will not guarantee best approximation to $w(X_i)$.
- A very poor approximation to the weights will avoid bias provided the approximation errors are uncorrelated with control outcomes.
- Conversely, a very good approximation may perform poorly if the errors are strongly correlated with outcomes.
- Relative performance of the two approaches will ultimately depend on process generating outcomes.

Sample Properties

Blinder-Oaxaca estimator:

$$\hat{\delta}^{BO} = \frac{1}{N_1} \boldsymbol{D}' \boldsymbol{X} \left(\boldsymbol{X}' \boldsymbol{W} \boldsymbol{X} \right)^{-1} \boldsymbol{X}' \boldsymbol{W} \boldsymbol{Y}$$
$$= \omega \boldsymbol{Y}$$

where $\boldsymbol{W} = diag \{1 - D_i\}$ and $N_1 = \sum D_i$.

Sample Properties

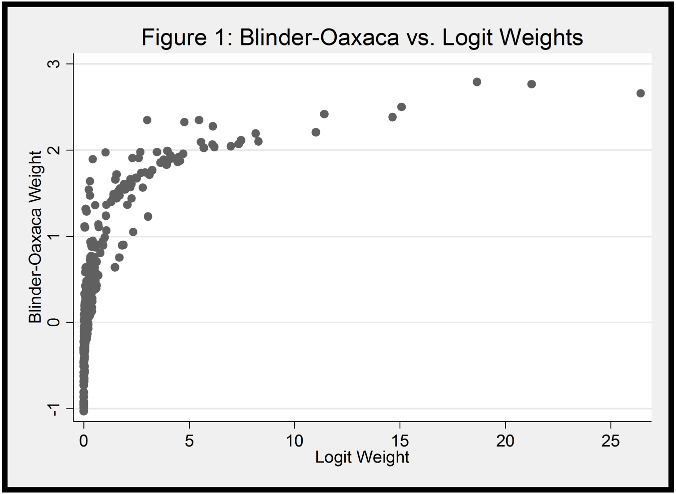
Sample weight vector ω has some interesting properties:

- Weights sum to one – potentially important (Busso, Dinardo, McCrary, 2010)
- Weights are zero for treated observations
- Weights may be negative for some observations (when estimated odds of treatment go negative)

Application

- Revisit Dehejia and Wahba (1999)'s reanalysis of LaLonde's classic 1986 analysis of the National Supported Work (NSW) program.
- Compare three estimators (OLS, B-O, and Logistic reweighting) to experimental benchmark.
- Sample consists of experimental NSW data and observational control sample (CPS-3) of poor and recently unemployed men from the CPS with nonmissing 1975 and 1976 earnings.
- In all cases Y_i is 1978 earnings and X_i contains: an intercept, age, age squared, years of schooling, black, hispanic, married, no degree,1975 earnings, and 1976 earnings.

B-O vs. Logistic Weights



Results

Table 1 - Estimated Impact of NSW on Men's 1978 Earnings		
Estimator/Control Group	CPS-3	NSW
Raw Difference	−$635	$1794
	(677)	(671)
OLS	$1369	$1676
	(739)	(677)
Logistic Reweighting*	$1440	$1808
	(863)	(705)
Blinder-Oaxaca	$1701	$1785
	(841)	(677)
Sample Size	614	445
Note: Heteroscedasticity robust standard errors in parentheses.		
*Reweighting standard errors computed from 1,000 bootstrap replications.		

Conclusion

- Blinder-Oaxaca has dual interpretation as propensity score reweighting estimator
- Provides MMSE approximation to weights without imposing side restriction that weights must be non-negative.
- Performance of B-O relative to conventional reweighting estimators will depend on DGP
  - B-O likely to be of most use in situations with unbalanced design (few treated, many controls) and lots of covariates.
  - Or where estimated propensity scores imply very large weight on a few observations. (Kang and Schaeffer, 2007)

Possible Extensions - Estimands

- If true propensity score is LPM, OLS can be shown to identify

$$\frac{E\left[e\left(X_i\right)\left(1-e\left(X_i\right)\left(Y_i^1-Y_i^0\right)\right]}{E\left[e\left(X_i\right)\left(1-e\left(X_i\right)\right)\right]}$$

  even even if outcome means are not linear in X_i.
- Two-sided B-O is DR for ATE.
- DR B-O decompositions?

Other Extensions

- Dual interpretation to IV-BO?
  - Semiparametric doubly robust estimators of LATE already exist (Tan, 2006; Uysal, 2010)
  - Does IV estimation among the controls provide predictions with a dual interpretation?
- Nonlinear estimators?