Some Finite Sample Properties and Assumptions of Methods for Determining Treatment Effects

Ordinary Least Squares Regression, Propensity Score Matching, and Inverse Probability Weighing Compared

Petrovski, Erik

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Petrovski, E. (2016). *Some Finite Sample Properties and Assumptions of Methods for Determining Treatment Effects: Ordinary Least Squares Regression, Propensity Score Matching, and Inverse Probability Weighing Compared*.

Some Finite Sample Properties and Assumptions of Methods for Determining Treatment Effects:

Ordinary Least Squares Regression, Propensity Score Matching, and Inverse Probability Weighing Compared

Erik Petrovski[1], 2016

Introduction

There is a growing interest in determining the exact effects of policies, programs, and other social interventions within the social sciences. In order to do so, researchers have a variety of econometric techniques at their disposal. However, the choice between them may be obscure.

In this paper, I will compare assumptions and properties of select methods for determining treatment effects with Monte Carlo simulation. The comparison will highlight the pros and cons of using one method over another and the assumptions that researchers need to make for the method they choose.

To limit the scope of this paper, three popular methods for determining treatment effects were chosen: ordinary least squares regression, propensity score matching, and inverse probability weighting. The assumptions and properties tested across these methods are: unconfoundedness, differences in average treatment effects and treatment effects on the treated, overlap, and robustness.

Methods

The treatment effect (τ_i) for the person, i, is defined as what a person's potential outcome (y_i) would have been under treatment (w) versus what their outcome would have been had they not been treated:

$$\tau_i = E[y_i|w_i = 1] - E[y_i|w_i = 0]$$

In practice, this problem cannot be solved since we only observe y_i for w=1 or w=0—i.e. the counterfactual is unobservable and we therefore face a problem of missing data. This forces us to estimate treatment effects alternatively and one appealing solution is to simply compare different individuals with and without treatment.

Under experimental circumstances—where treatment assignment is random—this comparison is mathematically straightforward, since:

$$E[x|w = 1] - E[x|w = 0] = 0,$$

where x represents pretreatment characteristics of individuals that could possibly confound the relationship between w and y. In this case, the average treatment effect (ATE) can simply be calculated as a difference in means of y, given w:

[1] MSc in Sociology & Ph.D. Candidate at Roskilde University, Denmark

$$ATE_{randomized} = E[y|w = 1] - E[y|w = 0]$$

Treatment effects models are relevant for observational studies in which such randomized circumstances are not present and we therefore need to control for confounding factors, such that:

$$ATE(x) = E[y|\boldsymbol{x}, w = 1] - E[y|\boldsymbol{x}, w = 0]$$

I will now account for how these models address the issue of confounding factors in order to determine treatment effects.

In practice, the most widely used method for determining a treatment effect in observational studies—such as the effect of higher education on income for a population—is OLS regression. Using OLS, the effect of a given treatment can simply be determined as the addition (β_1) of a binary treatment variable (w) to the slope (β_2) between a dependent variable (y) and confounding factors (x):

$$y = \alpha + \beta_1 w + \beta_2 x + u,$$

where α is the constant and u is the error term.

Other methods, such as matching and inverse probability weighing, have furthermore been developed for estimating treatment effects. These methods more closely emulate the actual process of treatment assignment by using propensity scores. With propensity score methods, we first estimate the propensity for receiving treatment ($\hat{p}(x)$), given pre-treatment characteristics (x):

$$\hat{p}(x) = prob\ (w = 1 \mid x),$$

which could be stated as a logit model—which I do in this study, rather than the alternative probit model. The resulting propensity scores are then assigned to all units.

With propensity score methods, we can then define the ATE as:

$$ATE \mid p(x) = E[y \mid p(x), w = 1] - E[y \mid p(x), w = 0],$$

With propensity score matching *(or just matching)*, the ATE is empirically estimated as:

$$\widehat{ATE}_{match} = N^{-1} \sum_{i=1}^{N} [\hat{y}_{i1} - \hat{y}_{i0}],$$

where $\hat{y}_{i1}$ and $\hat{y}_{i0}$ are retrieved by solving $\min \| p_{i1} - p_{i0} \|$ over all treated observations, thereby matching each treated observations with—in this case—a single most similar non-treated observation, since I will use nearest neighbor matching.

Another propensity based method for determining treatment effects is inverse propensity weighting *(IPW)*, which simply utilizes regression weights (w) on the basis of estimated propensity scores $\hat{p}$, for which $w = \frac{1}{\hat{p}}$ for treated and $w = \frac{1}{1-\hat{p}}$ for untreated individuals. Individuals who receive a treatment status that they have a low probability of receiving are thus given greater weight and this creates comparative treatment and control groups.

With IPW, the estimation of the ATE is defined as:

$$ATE = E\left[\frac{\big(w - p(x)\big)y}{p(x)\big(1 - p(x)\big)}\right],$$

which is estimated by:

$$\widehat{ATE} = N^{-1}\sum_{i=1}^{N}\left[\frac{\big(w_i - \hat{p}(x_i)\big)y_i}{\hat{p}(x_i)\big(1 - \hat{p}(x_i)\big)}\right].$$

The principal difference between matching and IPW then becomes that IPW is a weighted estimate over all relevant data points, whereas matching is looking for pairs of single observations to calculate differences *(i.e. treatment effects)* from.

Unconfoundedness

Treatment effects models—propensity based or not—fundamentally rest on our ability to control for confounding factors. This requirement can be defined strictly as the *conditional independence assumption:*

$$y_0, y_1 \perp w|x,$$

which states that conditional on x, w and potential outcome (y_0 & y_1) are independent.

However, it suffices that we use the milder *conditional mean independence assumption* in order to solve for the treatment effect*:*

$$E[y_0|x,w] = E[y_0|x] \; and \; E[y_1|x,w] = E[y_1|x].$$

This assumption asserts that potential outcome is the same for a treated and non-treated observation when all confounding pretreatment characteristics (x) have been controlled for.

Conditional mean independence (or unconfoundedness) cannot be asserted in the very likely case where y, in addition to being determined by x, is also determined by unobserved factors (z). Whether this is the case is fundamentally untestable when we only observe (y, w, x).

Lack of unconfoundedness is defined in the following Monte Carlo simulation, where selection into treatment is determined as:

$$w = 3x + z - 2 + \mathrm{e},$$

and treatment is given as:

$$y = 0 + \mathbf{0}\boldsymbol{w} + x + 0.5z + 0.2u,$$

where *e* and *u* are error terms.

Treatment is in this case worthless and any observed effect of w on y may be entirely attributed to selection of individuals for treatment on the basis of x *and* z, both of which are correlated with y.

In the real world, this setup could mirror something like the Spence argument for education as social signaling. According to Spence, higher education does not attribute to productivity but since productive individuals go for higher education, higher educated

individuals have higher earnings (Spence 1973). In our case, y could measure earnings, w higher education, and x and z aspects of individual productivity.

I use Monte Carlo simulation to create 1,000 finite samples of (n=1,000) according to the definitions of w and y above. In scenario 1, I run OLS regression, matching and IPW with all relevant variables (x & z) included in order to fully satisfy the assumption of mean independence. The estimated average treatment effects from each model are shown in table 1 below:

Table 1:	**Mean**	**Std. Dev.**	**Min**	**Max**
Regression	0.000	0.015	-0.051	0.052
Matching (ATE)	0.009	0.020	-0.058	0.078
IPW (ATE)	0.008	0.023	-0.071	0.074

Clearly, all three methods provide identical and correct estimates for the ATE equal to zero. This is due to the fact that we have controlled for—or in the case of matching, balanced the treatment and control group—on x and z.

In scenario 2, I have run the models on the same simulated datasets but the z variable has been dropped from the matching and regression models, thus emulating an omitted variable bias and thereby a breach of the conditional mean independence assumption. The results are shown in table 2 below:

Table 2:	**Mean**	**Std. Dev.**	**Min**	**Max**
Regression	0.068	0.018	0.013	0.121
Matching (ATE)	0.069	0.023	-0.008	0.137
IPW (ATE)	0.073	0.021	0.006	0.134

The results in table 2 highlight that all estimation methods are biased when the conditional independence assumption does not hold. Not surprisingly, there is nothing to be gained from matching and IPW since these methods rest on similar assumptions of unconfoundedness as OLS (Heckman 2005).

In empirical studies, researchers investigate complex social phenomena that are influenced by almost countless factors. Scenario 2 is therefore much more likely than scenario 1. This is important for researchers to be aware of since any omitted variable bias on the basis of unobservables is untestable and we therefore need to account for the extent that we expect unconfoundedness to have been broken.

Treatment effects

Up until now, I have limited my focus to the ATE. However, there are two types of treatment effects that are of interest to researchers: the average treatment effect (ATE) and the average treatment effect on the treated (ATT). In the previous simulation, by design ATT=ATE *(See appendix 1)*. This was due to the fact that benefit of treatment was not conditional on the

mechanism used for treatment assignment. In other words, it is the same stochastic process, which determines the potential outcome of treatment given x for individuals who have been treated as the process that determines potential outcome given x for all individuals who are not treated, so that:

$$ATE = ATT = E[y|x, w = 1] - E[y|x, w = 0]$$

This assumption is true by design in randomized studies but rarely true for observational studies since it is often the case that individuals participating in treatment have either been selected or have self-selected into treatment on the basis of an increased expectation of benefit. In such cases, the magnitude of the treatment effect is dependent on pretreatment characteristics and therefore ATT≠ATE.

The distinguish between the two treatment effects, the following formal definitions may be used:

$$\text{ATE} = E[y_1 - y_0|x]$$

$$\text{ATT} = E[y_1 - y_0|x, w = 1]$$

From these definitions, it becomes apparent that even though the average difference on the outcome variable for the two groups are estimated in both the ATE and ATT, only the ATT conditions this on individuals for which treatment is actually observed.

To empirically explore the difference between the ATE and ATT, I proceed to define a treatment selection process where:

$$w = -\mathbf{4}\boldsymbol{x} + 2 + \text{e},$$

and response to treatment is defined as:

$$y = 0 + (\mathbf{1} - \boldsymbol{x})\boldsymbol{w} + 0.5z + 0.4u$$

The above creates a process where, when the value of x is low, the more likely an individual is of receiving treatment and the lower x, the larger the benefit of w. In this case:

$$ATT = ATE + E[v1 - v0|w = 1],$$

which formally determines that the two treatment effects differ by the expected person specific gain (v_1-v_0) from treatment for those who have participated.

In such circumstances matching and IPW have clear benefits over standard OLS, since they make it possible to calculate the average treatment effects for the treated as well as the average treatment effects for the entire sample.

In the case of matching this is done by empirically estimating:

$$\widehat{ATT}_{match} = \sum_{i=1}^{N} w_i[\hat{y}_i - \hat{y}_{i0}].$$

In the case of IPW, this is done by estimating:

$$\widehat{ATT} = \mathrm{N}^{-1} \sum_{i=1}^{N} \frac{(w_i - \hat{p}(x_i))y_i}{\hat{\rho}(1 - \hat{p}(x_i))},$$

with $\hat{\rho} = N^{-1} \sum_i w_i$.

Naturally, the ability of IPW and matching to correctly calculate the ATT only holds under the assumption that we are actually able to control entirely for all factors (x) that determine both treatment effect size and selection into treatment. If not, we run into a similar omitted variable bias as discussed in the previous chapter.

In my case, I have controlled entirely for all relevant variables. The correct difference between ATT and ATE is therefore recovered in the following Monte Carlo simulation of 1,000 datasets of n=1,000 according to the definitions of w and y above. Results are shown in table 3 below:

Table 3:	**Mean**	**Std. Dev.**	**Min**	**Max**
Regression	0.499	0.036	0.366	0.611
Matching (ATT)	0.695	0.060	0.472	0.892
Matching (ATE)	0.508	0.046	0.357	0.691
IPW (ATT)	0.694	0.054	0.537	0.889
IPW (ATE)	0.507	0.045	0.344	0.664
True ATT	0.692	0.010	0.660	0.730
True ATE	0.500	0.009	0.471	0.528

In table 3, it is shown that regression provides a treatment effect of .5 whereas matching and IPW both provide correct estimates of ATE=.5 and ATT=.69.

This is due to the fact that whereas treatment on the treated estimators, such as matching and IPW, are able to put most weight on covariates of those who are most likely to receive treatment, regression on the other hand puts most weight on covariates where the conditional variance of treatment status is greatest (Angrist & Pischke 2008).

To put the importance of this distinction into practical terms, it may be appropriate with a short example. Say that a company wanted to implement a motivational program (w) and x measures pre-treatment motivation level. The ATT would be the benefit to expect from the program had it been administered to the most demotivated employees only, whereas ATE would have been the average expected benefit for all employees had the program been administered to everyone. In a cost-benefit situation, the difference in ATT and ATE may be crucial in determining whether to administer a motivational program to select employees only.

Finally, we should also comment on the consistency of the models. The standard deviation of the estimates for OLS is noticeably smaller than that of both matching and IPW. This is due to the fact that OLS uses x of all observations as a regressor, whereas propensity based methods use p(x) of treated observations and *n* most similar non-treated observations. In order to reduce bias, we match on small numbers of *n* but this results in greater variance, which translates into larger standard errors and therefore larger risk of type-I errors in singular

empirical applications. This is even more evident when looking at the estimation of the ATT, which comes with the least consistency due to the fact that the ATT makes the strictest limits on which observations to include in its calculation.

Overlap

It follows from the procedures for estimating the ATE, which were accounted for in the "Methods" chapter that we need to observe treated and non-treated observation on a common overlap of x with $x \in \mathfrak{I}$, where $\mathfrak{I}$ is the support of the covariates. Formally, Wooldridge (2010) states this assumption as ATE.2:

$$0 < P(w = 1|x) < 1 \; for \; all \; x \in \mathfrak{I} \, ,$$

which explicates that there needs to be a positive probability of treatment for all values of x.

I will now explore the overlap assumption further by creating a scenario of insufficient overlap by defining a dataset where selection into treatment depends strongly on x:

$$w = 9x - 4 + e,$$

and the effect treatment is further dependent upon the value of w, thus making treatment effects non-constant over x:

$$y = 0 + (1 - x) * w + z + 0.4u.$$

In order to show that the overlap assumption has been clearly violated for large parts of x, I have plotted treatment status (w) against its determining variable (x) for a single randomly generated dataset of n=1,000:

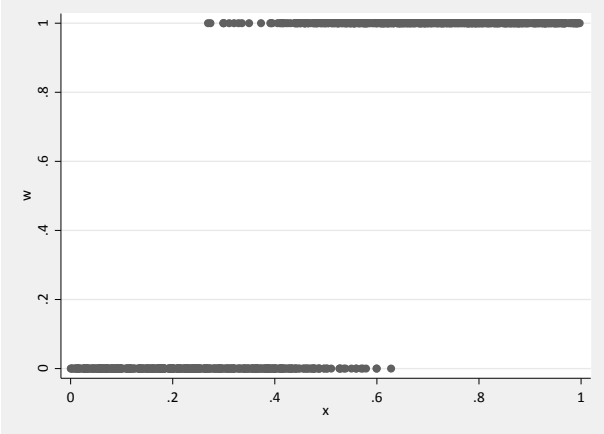

Figure 1

In figure 1, it is clearly seen that the overlap assumption is only fully satisfied on the small stretch of x from circa .2 to .6, leaving a large span of the variable without full overlap.

In the following, I choose to ignore the fact that ATE.2 had been violated and run OLS, matching, and IPW on 1000 simulated datasets of n=1,000. The results of which are shown in table 4 below:

Table 4:	Mean	Std. Dev.	Min	Max
Regression	0.592	0.046	0.459	0.747
Matching (ATE)	0.464	0.092	0.168	0.754
IPW (ATE)	0.427	0.094	-0.136	1.219
True ATE	0.501	0.009	0.472	0.527

Table 4 shows that had a researcher attempted to run an OLS regression model, on average a treatment effect for w at .59 would have been estimated. However, the true ATE≈.5. The misestimating of the treatment effect is due to the fact that the effect of w on y has been extrapolated out of bounds by OLS to the full span of x from the small stretch on which overlap is observed (Wooldridge, 2010). A similar biased estimate of the ATE is provided by the propensity score models.

However, matching and IPW do allow for a correct estimation of the ATT when only the following milder overlap assumption ATT.2 in (Wooldridge 2010) holds:

$$P(w = 1|x) < 1 \; for \; all \; x \in \Im$$

ATT.2 simply states that there may be parts of the population which are never likely to be treated, but not parts of the population that are always treated.
To simulate this circumstance, I define treatment selection as:

$$w = 5x - 4 + e$$

And response as:

$$y = 0 + (1 - x)w + 0.5z + 0.4u$$

Which provides the following overlap between w and x, where overlap is lacking only in the lower end and not the high end of the specter of the x variable:

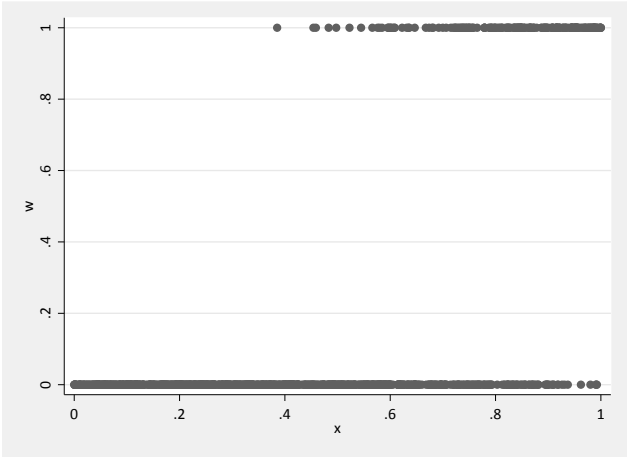

Figure 2

Figure 2 illustrates that the weaker assumption for estimating ATT should hold in this case, and I may therefore proceed with estimating the ATT by propensity score methods. In table 4 below, I have provided the results from the Monte Carlo simulation of 1,000 datasets with n=1,000:

Table 5:	Mean	Std. Dev.	Min	Max
Matching, ATT	0.178	0.057	-0.016	0.362
IPW, ATT	0.176	0.049	0.021	0.341
True ATT	0.178	0.009	0.144	0.210

The results in table 5 show that both matching and IPW provide correct estimates of the true ATT, even when only the weaker overlap assumption holds.

The ability to estimate treatment effects under ATT.2 is an attractive feature of propensity score methods since it is often the case that targeted social policies, which we may wish to estimate the effect of, are only relevant for a particular subset of the population, thus leaving large parts untreated, which is in breach of ATE.2 and thus hinders the reliable estimation of the ATE.

Robustness

In econometrics, results equal data plus assumptions. Though assumptions do enable many efficient calculations, it is also generally the case that the fewer assumptions made, the more likely results will reflect the underlying data structure rather than the assumptions made by the researcher. One issue with OLS is that it enforces strict assumptions on the linear functional form of the relationship between y and its covariates and deviations from this form must be explicated in the model (Cameron & Trivedi 2005). Matching and IPW methods, on the other hand, do not make such assumptions since the relationship between w and y is treated as one-dimensional (Wooldridge 2010). This follows from the already stated conditional mean independence, in which it is assumed that:

$$E(y_0 \,|x, w) \;=\; E(y_0 \,|x) \; and \; E(y_1 \,|x, w) \;=\; E(y_1 \,|x).$$

Robustness is achieved due to the fact that the definition above does not preclude higher moments of x to depend on w.

I test the robustness of matching and IPW by defining datasets where treatment selection is given by a linear function:

$$w = 2x - 2 + \mathrm{e},$$

but the distribution of y is given by a non-linear function of the form:

$$y = 0 + 1w + \mathbf{14}(\boldsymbol{x} - \mathbf{0.5})^2 + 0.5z + 0.4u$$

The effect of the non-linearity of the x control variable on the estimation of the treatment effect of w, which is 1, can be seen in table 6 below, which are results from the models from Monte Carlo simulation of 1,000 datasets of n=1,000:

Table 6:	Mean	Std. Dev.	Min	Max
Regression	1.245	0.099	0.957	1.565
Matching (ATT)	1.010	0.049	0.868	1.180
Matching (ATE)	0.931	0.066	0.703	1.106
IPW (ATT)	0.958	0.086	0.691	1.269
IPW (ATE)	0.939	0.127	0.603	1.447

In this case, regression has produced an upward bias of ~20 % for the effect of w since the effect of x on y is non-linear and w is absorbing this nonlinear relationship due to the fact that w is partly determined by x. As expected, both matching and IPW generally provided good estimates of the true ATT and ATE≈1, thereby underlining their robust properties.

If researchers have a good understanding of how treatment is assigned — but not of how response arises — propensity score methods therefore provide a good robust estimation approach.

Furthermore, covariate based (not propensity score) matching is said to be doubly robust, which means that it provides unbiased estimates even when the functional form of the treatment model is not correctly specified, however, the researcher does need to correctly specify either the treatment or response model. In order to achieve doubly robustness with IPW, we need to combine IPW with regression adjustment (Wooldridge 2010) but this technique is beyond the scope of this paper.

Conclusion

This paper shows that OLS, matching, and IPW provide identical results in cases where only the ATE is of interest (or ATE=ATT), the functional form has been correctly specified, and the strong overlap assumption holds. If such assumptions can forcefully be made, researchers could just apply an OLS model—and thereby furthermore reap its efficiency gains. However, in likely variations of these circumstances, both matching and IPW have provided more correct and/or desirable results.

However, no noticeable difference in the estimated treatment effects were detected between the two propensity score based techniques, IPW and matching, in these admittedly very limited simulations.

References

Angrist, J. D., & **Pischke**, J.-S. (2008): *Mostly Harmless Econometrics: An Empiricists Companion* (pp. 1–290).

Heckman, James J. (2005): "The Scientific Model of Causality". *Sociological Methodology*, 35:1-97.

Spence, Michael (1973): "Job Market Signaling". *Quarterly Journal of Economics* 87 (3): 355–374.

Cameron, Colin & **Trivedi**, Pravin. (2005): *Microeconometrics*, 1–1058.

Wooldridge, J. M. (2010)*: Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Appendix 1

Table a.1:	**Mean**	**Std. Dev.**	**Min**	**Max**
Regression	0.000	0.015	-0.051	0.052
Matching (ATT)	0.008	0.026	-0.067	0.103
Matching (ATE)	0.009	0.020	-0.058	0.078
IPW (ATT)	0.007	0.031	-0.108	0.094
IPW (ATE)	0.008	0.023	-0.071	0.074

Table a.2:	**Mean**	**Std. Dev.**	**Min**	**Max**
Regression	0.068	0.018	0.013	0.121
Matching (ATT)	0.069	0.028	-0.021	0.146
Matching (ATE)	0.069	0.023	-0.008	0.137
IPW (ATT)	0.072	0.027	-0.013	0.148
IPW (ATE)	0.073	0.021	0.006	0.134

Appendix 2

This appendix contains a Stata script for running all Monte Carlo simulations presented in the paper. The script was written in Stata 13.1.

```
ssc install labsumm


**********************************
*SELECTION INTO WORTHLESS TREATMENT*
**********************************

clear all

program define sim1, rclass

drop _all

set obs 1000

gen x=runiform()
gen z=runiform()
gen w=rbinomial(1,normal(3*x+z-2))
gen u=rnormal()
gen y=0+0*w+x+0.5*z+0.2*u

reg y w x z
return scalar a=_b[w]

teffects psmatch (y) (w x z), atet
return scalar b=[ATET]_b[r1vs0.w]

teffects psmatch (y) (w x z), ate
```

```
return scalar c=[ATE]_b[r1vs0.w]

teffects ipw (y) (w x z), atet
return scalar d=[ATET]_b[r1vs0.w]

teffects ipw (y) (w x z), ate
return scalar e=[ATE]_b[r1vs0.w]

reg y w x
return scalar f=_b[w]

teffects psmatch (y) (w x), atet
return scalar g=[ATET]_b[r1vs0.w]

teffects psmatch (y) (w x), ate
return scalar h=[ATE]_b[r1vs0.w]

teffects ipw (y) (w x), atet
return scalar i=[ATET]_b[r1vs0.w]

teffects ipw (y) (w x), ate
return scalar j=[ATE]_b[r1vs0.w]

end

simulate a=r(a) b=r(b) c=r(c) d=r(d) e=r(e) f=r(f) g=r(g) h=r(h)
i=r(i) j=r(j), reps(1000): sim1

label variable a "Regression"
label variable b "Matching (ATT)"
label variable c "Matching (ATE)"
label variable d "IPW (ATT)"
label variable e "IPW (ATE)"

label variable f "Regression"
label variable g "Matching (ATT)"
label variable h "Matching (ATE)"
label variable i "IPW (ATT)"
label variable j "IPW (ATE)"

labsumm a c e
labsumm a b c d e
labsumm f h j
labsumm f g h i j

*********************************************************
*SELECTION INTO TREATMENT FOR THOSE WHO BENEFIT THE MOST*
*********************************************************

clear all

program define sim2, rclass

drop _all

set obs 1000

gen x=runiform()
gen z=runiform()
gen w=rbinomial(1,normal(-4*x+2))
gen u=rnormal()
gen y=0+(1-x)*w+0.5*z+0.4*u
```

```
gen atet=1-x if w==1
sum atet
return scalar atet=r(mean)

gen ate=1-x
sum ate
return scalar ate=r(mean)

reg y w x z
return scalar a=_b[w]

teffects psmatch (y) (w x z), atet
return scalar b=[ATET]_b[r1vs0.w]

teffects psmatch (y) (w x z), ate
return scalar c=[ATE]_b[r1vs0.w]

teffects ipw (y) (w x z), atet
return scalar d=[ATET]_b[r1vs0.w]

teffects ipw (y) (w x z), ate
return scalar e=[ATE]_b[r1vs0.w]

end

simulate a=r(a) b=r(b) c=r(c) d=r(d) e=r(e) atet=r(atet) ate=r(ate),
reps(1000): sim2

label variable a "Regression"
label variable b "Matching (ATT)"
label variable c "Matching (ATE)"
label variable d "IPW (ATT)"
label variable e "IPW (ATE)"
label variable atet "True ATT"
label variable ate "True ATE"

labsumm a b c d e atet ate

******************************
*SEVERE LACK OF COMMON SUPPORT*
******************************

clear all

program define sim3, rclass

drop _all

set obs 1000

gen x=runiform()
gen z=runiform()
gen w=rbinomial(1,normal(9*x-4))
gen u=rnormal()
gen y=0+(1-x)*w+0.5*z+0.4*u

plot w x

gen ate=1-x
sum ate
return scalar ate=r(mean)

reg y w x z
```

```
return scalar a=_b[w]

teffects psmatch (y) (w x z), ate
return scalar c=[ATE]_b[r1vs0.w]

teffects ipw (y) (w x z), ate
return scalar e=[ATE]_b[r1vs0.w]

end

simulate a=r(a) c=r(c) e=r(e) ate=r(ate), reps(1000): sim3

label variable a "Regression"
label variable c "Matching (ATE)"
label variable e "IPW (ATE)"
label variable ate "True ATE"

labsumm a c e ate

****************************
*SOME LACK OF COMMON SUPPORT*
****************************


clear all

program define sim4, rclass

drop _all

set obs 1000

gen x=runiform()
gen z=runiform()
gen w=rbinomial(1,normal(5*x-4))
gen u=rnormal()
gen y=0+(1-x)*w+0.5*z+0.4*u

gen atet=1-x if w==1
sum atet
return scalar atet=r(mean)

teffects psmatch (y) (w x z), atet
return scalar b=[ATET]_b[r1vs0.w]

teffects ipw (y) (w x z), atet
return scalar d=[ATET]_b[r1vs0.w]

end

simulate b=r(b) d=r(d) atet=r(atet), reps(1000): sim4

label variable b "Matching (ATT)"
label variable d "IPW (ATT)"
label variable atet "True ATT"

labsumm b d atet

*************************
*NONLINEARITY OF RESPONSE*
*************************

clear all
```

```
program define sim5, rclass

drop _all

set obs 1000

gen x=runiform()
gen z=runiform()
gen w=rbinomial(1,normal(2*x-2))
gen u=rnormal()
gen y=0+1*w+14*((x-0.5)^2)+0.5*z+0.4*u

reg y w x z

return scalar a=_b[w]

teffects psmatch (y) (w x z), atet
return scalar b=[ATET]_b[r1vs0.w]

teffects psmatch (y) (w x z), ate
return scalar c=[ATE]_b[r1vs0.w]

teffects ipw (y) (w x z), atet
return scalar d=[ATET]_b[r1vs0.w]

teffects ipw (y) (w x z), ate
return scalar e=[ATE]_b[r1vs0.w]

end

simulate a=r(a) b=r(b) c=r(c) d=r(d) e=r(e), reps(1000): sim5

label variable a "Regression"
label variable b "Matching (ATT)"
label variable c "Matching (ATE)"
label variable d "IPW (ATT)"
label variable e "IPW (ATE)"

labsumm a b c d e
```