The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, STATA, Stata Press, Mata, MATA, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2014)
14, Number 3, pp. 541–561

treatrew: A user-written command for estimating average treatment effects by reweighting on the propensity score

Giovanni Cerulli
Ceris-CNR
National Research Council of Italy
Institute for Economic Research on Firms and Growth
Rome, Italy
g.cerulli@ceris.cnr.it

Abstract. Reweighting is a popular statistical technique to deal with inference in the presence of a nonrandom sample, and various reweighting estimators have been proposed in the literature. This article presents the user-written command `treatrew`, which implements reweighting on the propensity-score estimator as proposed by Rosenbaum and Rubin (1983, *Biometrika* 70: 41–55) in their seminal article. The main contribution of this command lies in providing analytical standard errors for the average treatment effects in the whole population, in the subpopulation of the treated, and in that of the untreated. Standard errors are calculated using the approximation suggested by Wooldridge (2010, 920–930, *Econometric Analysis of Cross Section and Panel Data* [MIT Press]), but bootstrapped standard errors can also be easily computed. Because an implementation of this estimator with analytic standard errors and nonnormalized weights is missing in Stata, this article and the accompanying ado-file aim to provide the community with an easy-to-use method for reweighting on the propensity-score. The estimator proves to be a valuable tool for estimating average treatment effects under selection on observables.

Keywords: st0350, treatrew, treatment models, reweighting, propensity score, average treatment effects, ATE, ATET, ATENT

1 Introduction

`treatrew` is a user-written command for estimating average treatment effects (ATEs) by reweighting (REW) on the propensity score. Depending on the specified model (probit or logit), `treatrew` provides consistent estimation of ATEs under the hypothesis of selection on observables. Conditional on a prespecified set of observable exogenous variables $\mathbf{x}$—thought of as those driving the nonrandom assignment to treatment—`treatrew` estimates the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the average treatment effect on the nontreated (ATENT); it also estimates these parameters conditional on the observable factors $\mathbf{x}$ (that is, ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$)).

In program evaluations and the epidemiological literature, a plethora of REW estimators have been proposed. This article presents the user-written command `treatrew`, which implements REW on the propensity-score estimator as proposed by Rosenbaum and Rubin (1983) in their seminal article.

The main contribution of this command lies in providing analytical standard errors for the estimation of the ATE, ATET, and ATENT using the approximation suggested by Wooldridge (2010, 920–930). However, bootstrapped standard errors can also be easily computed. `treatrew` assumes that the propensity score specified by the user is correct. Thus it is sensitive to propensity-score misspecification.

The article is organized as follows: Section 2 provides the statistical description of REW on the propensity-score estimator as implemented by `treatrew`. Section 3 provides the formulas for calculating the causal parameters of interest and their standard errors. Section 4 presents the syntax of `treatrew` and an application to real data. Section 5 shows the relation between `treatrew` and the recent Stata 13 command `teffects ipw` for implementing the inverse-probability weighting (IPW) estimator. Section 6 concludes the article. Finally, two appendixes are reported at the end of the article.

2 The REW estimator of treatment effects: A brief overview

Reweighting is a valuable approach to estimate (binary) treatment effects in a nonexperimental statistical setting when subjects' nonrandom assignment to treatment is due to selection on observables. The idea behind the REW procedure is straightforward: when the treatment is not randomly assigned, treated and untreated subjects may present different distributions of their observable characteristics. This may happen either because of the subjects' self-selection into the experiment (subjects may consider the net benefit of participation) or because of the selection process operated by an external entity (such as a public agency managing a subsidization program whose explicit objective is selecting beneficiaries with peculiar characteristics to maximize policy effect). Many examples can be drawn from both social and epidemiological statistical settings.

In nonrandomized experiments, the distribution of the variables feeding into $\mathbf{x}$ could be strongly unbalanced. To establish a balance in their distributions, one could implement REW on observations, using their probability of becoming treated, that is, according to subjects' propensity scores. A possible REW estimation protocol is as follows:

1. Estimate the propensity score (based on $\mathbf{x}$) using a logit or a probit regression, thus obtaining the predicated probability p_i.
2. Build weights as $1/p_i$ for treated observations and $1/(1-p_i)$ for untreated observations.
3. Calculate ATEs by comparing the weighted means of the two groups (for instance, with a weighted least-squares [WLS] regression).

This weighting scheme is based on inverse-probability regression (Robins, Hernán, and Brumback 2000; Brunell and DiNardo 2004)—that is, the idea that penalizing (advantaging) treated subjects with higher (lower) probability to be treated and advantaging (penalizing) untreated subjects with higher (lower) probability to be treated make the two groups as similar as possible. In other words, weights eliminate a confounding component induced by the extent of the nonrandom assignment to a program.

Alternative weighting schemes have been proposed in the literature[1], and some authors have shown that various matching methods can also be seen as specific REW estimators (Lunceford and Davidian 2004; Morgan and Harding 2006). As in matching, these estimators have different properties, but the main limit resides in the specification of the propensity score because measurement errors in this specification could produce severe bias. In what follows, we focus on REW on propensity-score inverse probability as proposed by Rosenbaum and Rubin (1983). Here we start with the following assumptions about the data-generating process:

i. $y_1 = g_1(\mathbf{x}) + \varepsilon_1, \quad E(\varepsilon_1) = 0$

ii. $y_0 = g_0(\mathbf{x}) + \varepsilon_0, \quad E(\varepsilon_0) = 0$

iii. $y = wy_1 + y_0(1-w)$

iv. Conditional mean independence (CMI) holds; therefore, $E(y_1|w,\mathbf{x}) = E(y_1|\mathbf{x})$ and $E(y_0|w,\mathbf{x}) = E(y_0|\mathbf{x})$

v. $\mathbf{x}$ exogenous

y_1 and y_0 are the subject's outcome when treated and untreated, respectively; $g_1(\mathbf{x})$ and $g_0(\mathbf{x})$ are the subject's reaction function to the confounder $\mathbf{x}$ when the subject is treated and untreated, respectively; w is the treatment binary indicator taking value 1 for treated and 0 for untreated subjects; ε_0 and ε_1 are two error terms with unconditional zero mean; and $\mathbf{x}$ is a set of observable and exogenous confounding variables assumed to drive the nonrandom assignment into treatment. In short, the CMI assumption states that it is sufficient to control only for $\mathbf{x}$ to restore random assignment conditions. When assumptions i–v hold,

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

$$\text{ATET} = E\left[\frac{\{w-p(\mathbf{x})\}y}{p(w=1)\{1-p(\mathbf{x})\}}\right] \tag{2}$$

$$\text{ATENT} = E\left[\frac{\{w-p(\mathbf{x})\}y}{p(w=0)p(\mathbf{x})}\right] \tag{3}$$

Appendix A shows the mathematical steps to get these formulas.

1. Another possible weighting scheme could be assuming $p_i/(1-p_i)$ for untreated subjects and 1 for treated ones (Nichols 2007). The literature distinguishes between normalized and nonnormalized weighting schemes depending on whether the weights sum to one or to a different value, respectively (Busso, DiNardo, and McCrary 2008).

3 Sample estimation and standard errors for ATE, ATET, and ATENT

Assuming that the propensity score is correctly specified, we can estimate previous parameters by using the sample equivalent of the population parameters; that is,

$$\text{ATE} = \frac{1}{N}\sum_{i=1}^{N} \frac{\{w_i - \widehat{p}(\mathbf{x}_i)\}y_i}{\widehat{p}(\mathbf{x}_i)\{1-\widehat{p}(\mathbf{x}_i)\}}$$

$$\text{ATET} = \frac{1}{N}\sum_{i=1}^{N} \frac{\{w_i - \widehat{p}(\mathbf{x}_i)\}y_i}{\widehat{p}(w=1)\{1-\widehat{p}(\mathbf{x}_i)\}}$$

$$\text{ATENT} = \frac{1}{N}\sum_{i=1}^{N} \frac{\{w_i - \widehat{p}(\mathbf{x}_i)\}y_i}{\widehat{p}(w=0)\widehat{p}(\mathbf{x}_i)}$$

Estimation follows in two steps: i) estimate the propensity score $p(\mathbf{x}_i)$, thus obtaining $\widehat{p}(\mathbf{x}_i)$; and ii) substitute $\widehat{p}(\mathbf{x}_i)$ into previous formulas to get parameters. Consistency is guaranteed because these estimators are M-estimators.

But how do we get standard errors for previous estimators? We can exploit some results when the first step is a maximum likelihood (ML) estimation and the second step is an M-estimation. In our case, the first step is an ML based on logit (or probit), and the second step is a standard M-estimator. For such cases, Wooldridge (2007; 2010, 922–924) proposed a straightforward procedure to get analytical standard errors provided that the propensity score is correctly specified. In what follows, we demonstrate Wooldridge's (2007; 2010, 922–924) procedure and formulas for obtaining these standard errors.

3.1 Standard-error estimation for ATE

First, define the estimated ML score of the first step (probit or logit). It is, by definition, equal to

$$\widehat{\mathbf{d}}_i = \widehat{\mathbf{d}}(w_i, \mathbf{x}_i, \widehat{\boldsymbol{\gamma}}) = \frac{\{\nabla_\gamma \widehat{p}(\mathbf{x}_i, \widehat{\boldsymbol{\gamma}})\}' \times \{w_i - \widehat{p}(\mathbf{x}_i, \widehat{\boldsymbol{\gamma}})\}}{\widehat{p}(\mathbf{x}_i, \widehat{\boldsymbol{\gamma}})\{1-\widehat{p}(\mathbf{x}_i, \widehat{\boldsymbol{\gamma}})\}}$$

Observe that $\mathbf{d}$ is a row vector of the $R-1$ parameters $\boldsymbol{\gamma}$ and represents the gradient of the function $p(\mathbf{x}, \boldsymbol{\gamma})$.

Second, define the generic estimated summand of ATE as

$$\widehat{k}_i = \frac{\{w_i - \widehat{p}(\mathbf{x}_i)\}y_i}{\widehat{p}(\mathbf{x}_i)\{1-\widehat{p}(\mathbf{x}_i)\}}$$

Third, calculate ordinary least-squares (OLS) residuals from this regression,

$$\widehat{k}_i \quad \text{on} \quad \left(1, \widehat{\mathbf{d}}'_i\right) \quad \text{with} \quad i = 1, \ldots, N$$

and call them $\widehat{e}_i$ $(i = 1, \ldots, N)$. The asymptotic standard error for ATE is equal to

$$\frac{\left(\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \widehat{e}_i^2\right)^{1/2}}{\sqrt{N}} \tag{4}$$

and we can use it to test the significance of ATE. Of course, $\mathbf{d}$ will have a different expression according to the probability model adopted. Here we consider the logit and probit cases.

Case 1: Logit

Suppose that the correct probability follows a logistic distribution. This means that

$$p(\mathbf{x}_i, \boldsymbol{\gamma}) = \frac{\exp(\mathbf{x}_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}_i \boldsymbol{\gamma})} = \Lambda(\mathbf{x}_i \boldsymbol{\gamma}) \tag{5}$$

Thus, by simple algebra, we see that

$$\underbrace{\widehat{\mathbf{d}}'_i}_{1 \times R} = \mathbf{x}_i (w_i - \widehat{p}_i)$$

Case 2: Probit

Suppose that the right probability follows a normal distribution. This means that

$$p(\mathbf{x}_i, \boldsymbol{\gamma}) = \Phi(\mathbf{x}_i \boldsymbol{\gamma})$$

Thus, by simple algebra, we see that

$$\widehat{\mathbf{d}}'_i = \frac{\phi(\mathbf{x}_i, \widehat{\boldsymbol{\gamma}}) \mathbf{x}_i \times \{w_i - \Phi(\mathbf{x}_i \boldsymbol{\gamma})\}}{\Phi(\mathbf{x}_i \boldsymbol{\gamma})\{1 - \Phi(\mathbf{x}_i \boldsymbol{\gamma})\}}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the normal cumulative distribution and density function, respectively. One can also add functions of $\mathbf{x}$ to estimate previous formulas. This reduces standard errors if these functions are partially correlated with $\widehat{k}_i$.

Finally, observe that the previous procedure produces standard errors that are lower than those produced by ignoring the first step (that is, the propensity-score estimation via ML). Indeed, the naïve standard error

$$\frac{\left\{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \left(\widehat{k}_i - \widehat{\text{ATE}}\right)^2\right\}^{1/2}}{\sqrt{N}}$$

is higher than the one produced by the previous procedure.

3.2 Standard-error estimation for ATET

This follows a route similar to ATE. Define the generic estimated summand of ATET as

$$\widehat{q}_i = \frac{\{w_i - \widehat{p}(\mathbf{x}_i)\}y_i}{\widehat{p}(w=1)\{1-\widehat{p}(\mathbf{x}_i)\}}$$

and calculate

$$\widehat{r}_i = \text{residuals from the regression of “}\widehat{q}_i \text{ on } 1, \widehat{\mathbf{d}}_i'\text{”}$$

The asymptotic standard error for ATET is

$$\frac{\{\widehat{p}(w=1)\}^{-1} \times \left\{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left(\widehat{r}_i - w_i \times \widehat{\text{ATET}}\right)^2\right\}^{1/2}}{\sqrt{N}}$$

3.3 Standard-error estimation for ATENT

In this case, define the generic estimated summand of ATENT as

$$\widehat{b}_i = \frac{\{w_i - \widehat{p}(\mathbf{x}_i)\}y_i}{\widehat{p}(w=0)\widehat{p}(\mathbf{x}_i)}$$

and then calculate

$$\widehat{s}_i = \text{residuals from the regression of “}\widehat{b}_i \text{ on } 1, \widehat{\mathbf{d}}_i'\text{”}$$

The asymptotic standard error for ATENT is

$$\frac{\{\widehat{p}(w=0)\}^{-1} \times \left[\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left\{\widehat{s}_i - (1-w_i) \times \widehat{\text{ATENT}}\right\}^2\right]^{1/2}}{\sqrt{N}}$$

The standard errors presented in this section are correct when the actual data-generating process follows the probit or the logit probability rules. If not, then a measurement error is present, and the estimations might be inconsistent. Authors such as Hirano, Imbens, and Ridder (2003) and Li, Racine, and Wooldridge (2009) have suggested more flexible nonparametric estimation of the standard errors. Under correct specification, a straightforward alternative is to use bootstrapping, where the binary response estimation and the averaging are included in each bootstrap iteration.

4 The treatrew command: Syntax and use

`treatrew` estimates ATE, ATET, and ATENT parameters with either analytical or bootstrapped standard errors. The syntax is rather simple and follows the typical Stata command syntax. The user has to declare: a) the outcome variable, that is, the variable over which the treatment is expected to have an impact (*outcome*); b) the binary treatment variable (*treatment*); c) a set of confounding variables (*varlist*); and, finally, d) a series of options. Two options are important: the option `model(`*modeltype*`)` sets the type of model, `probit` or `logit`, that has to be used in estimating the propensity score; the option `graphic` and the related option `range(`*a b*`)` produce a chart where the distribution of ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$) are jointly plotted within the interval [a; b].

As an e-class command, `treatrew` provides an `ereturn list` of objects (such as scalars and matrices) to be used in the next elaborations. In particular, the values of ATE, ATET, and ATENT are returned in the scalars `e(ate)`, `e(atet)`, and `e(atent)`, and they can be used to get bootstrapped standard errors. By default, `treatrew` provides analytical standard errors.

4.1 Syntax

`treatrew` *outcome treatment* [*varlist*] [*if*] [*in*] [*weight*], `model(`*modeltype*`)`
[`graphic range(`*a b*`) conf(`#`) vce(robust)`]

outcome is the target variable for measuring the impact of the treatment.

treatment is the binary treatment variable taking 1 for treated and 0 for untreated subjects.

varlist is the set of pretreatment (or observable confounding) variables.

`fweight`s, `iweight`s, and `pweight`s are allowed; see [U] **11.1.6 weight**.

4.2 Description

`treatrew` estimates ATEs by REW on the propensity score. Depending on the specified model, `treatrew` provides consistent estimation of ATEs under the hypothesis of selection on observables. Conditional on a prespecified set of observable exogenous variables $\mathbf{x}$—thought of as those driving the nonrandom assignment to treatment—`treatrew` estimates the ATE, the ATET, the ATENT, and these parameters conditional on the observable factors $\mathbf{x}$ (that is, ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$)). Parameters' standard errors are provided either analytically (following Wooldridge [2010, 920–930]) or via bootstrapping. `treatrew` assumes that the propensity-score specification is correct.

`treatrew` creates several variables:

- `ATE_x` is an estimate of the idiosyncratic ATE.
- `ATET_x` is an estimate of the idiosyncratic ATET.
- `ATENT_x` is an estimate of the idiosyncratic ATENT.

4.3 Options

`model(`*modeltype*`)` specifies the model for estimating the propensity score, where *modeltype* must be one of `probit` or `logit`. `model()` is required.

`graphic` allows for a graphical representation of the density distributions of ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$) within their whole support.

`range(`*a b*`)` allows for a graphical representation of the density distributions of ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$) within the support $[a; b]$ specified by the user. `range()` must be specified with the `graphic` option.

`conf(`#`)` sets the confidence level of probit or logit estimates equal to the specified #. The default is `conf(95)`.

`vce(robust)` allows for robust regression standard errors in the probit or logit estimates.

4.4 Stored results

`treatrew` stores the following in `e()`:

Scalars

`e(N)`	number of observations	`e(ate)`	value of the ATE
`e(N1)`	number of (used) treated subjects	`e(atet)`	value of the ATET
`e(N0)`	number of (used) untreated subjects	`e(atent)`	value of the ATENT

4.5 Examples

To show a practical application of `treatrew`, we use an instructional dataset called `fertil2.dta`, which is included in Wooldridge (2013) and collects cross-sectional data on 4,361 women of childbearing age in Botswana. This dataset is freely downloadable at http://fmwww.bc.edu/ec-p/data/wooldridge/fertil2.dta. It contains 28 variables on women and family characteristics.

Using `fertil2.dta`, we are interested in evaluating the impact of the variable `educ7` (taking value 1 if a woman has more than or exactly seven years of education and 0 otherwise) on the number of family children (`children`). Several conditioning (or confounding) observable factors are included in the dataset, such as the age of the woman (`age`), whether the family owns a television (`tv`), whether the woman lives in a city (`urban`), and so forth. To inquire into the relation between education and fertility according to Wooldridge's (2010, ex. 21.3, 940) specification, we estimate the ATE, ATET, and ATENT (as well as ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$)) by REW using `treatrew`. We also compare REW results with other popular program evaluation methods: i) the difference in mean (DIM), taken as benchmark; ii) the OLS regression-based random-coefficient model with heterogeneous reaction to confounders, estimated through the user-written command `ivtreatreg`, provided by Cerulli (2011); and iii) a one-to-one nearest-neighbor matching, computed by the command `psmatch2`, provided by Leuven and Sianesi (2003). Because matching estimators can be seen as specific REW procedures (Busso, DiNardo, and McCrary 2008), comparing REW with matching is worthwhile. By taking just the case of ATET, we can prove that

$$\begin{aligned}
\text{ATET}_{\text{Matching}} &= \frac{1}{N_i} \sum_{i \in (w=1)} \left\{ y_i - \sum_{j \in C(i)} h(i,j) y_j \right\} \\
&= \frac{1}{N_1} \sum_{i=1}^{N} w_i y_i - \sum_{j=1}^{N} (1 - w_j) y_j \frac{1}{N_1} \sum_{i=1}^{N} w_i h(i,j) \\
&= \frac{1}{N_1} \sum_{i=1}^{N} w_i y_i - \frac{1}{N_0} \sum_{j=1}^{N} (1 - w_j) y_j \omega(j) = \text{ATET}_{\text{Reweighting}}
\end{aligned}$$

where $\omega(j) = N_0/N_1 \sum_{i=1}^{N} w_i h(i,j)$ are REW factors, $C(i)$ is the untreated subject's neighborhood for the treated subject i, and $h(i,j)$ are matching weights that—once opportunely specified—produce different types of matching methods. Results from all of these estimators are reported in table 1.

Table 1. Comparison of ATE, ATET, and ATENT estimation among DIM, CF-OLS, REW, and MATCH

	1	2	3	4	5	6	7
	DIM	CF-OLS	REW (probit) analytical standard errors	REW (logit) analytical standard errors	REW (probit) bootstrapped standard errors	REW (logit) bootstrapped standard errors	MATCH(a)
ATE	−1.77 ***	−0.374 ***	−0.43 ***	−0.415 ***	−0.434 ***	−0.415 ***	−0.316 ***
	0.062	0.051	0.068	0.068	0.070	0.071	0.080
	−28.46	−7.35	−6.34	−6.09	−6.15	−5.87	−3.93
ATET		−0.255 ***	−0.355 **	−0.345 ***	−0.355 ***	−0.345 ***	−0.131
		0.048	0.15	0.104	0.0657	0.054	0.249
		−5.37	−2.37	−3.33	−5.50	−6.45	−0.52
ATENT		−0.523 ***	−0.532 ***	−0.503 **	−0.532 ***	−0.503 ***	−0.549 ***
		0.075	0.19	0.257	0.115	0.119	0.135
		−7.00	−2.81	−1.96	−4.61	−4.21	−4.07

Note: b/se/t; DIM; CF-OLS: control-function OLS; REW; MATCH. (a) Standard errors for ATE and ATENT are computed by bootstrapping. *** = 1%, ** = 5%, * = 10% of significance.

Results in column 1 refer to the DIM and are obtained by typing

```
. regress children educ7
```

Results in column 2 refer to CF-OLS and are obtained by typing

```
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
> hetero(age agesq evermarr urban electric tv) model(cf-ols)
```

For CF-OLS, standard errors for ATET and ATENT are obtained via bootstrap and can be obtained in Stata by typing

```
. bootstrap atet=r(atet) atent=r(atent), rep(200):
> ivtreatreg children educ7 age agesq evermarr urban electric tv,
> hetero(age agesq evermarr urban electric tv) model(cf-ols)
```

Results set out in columns 3–6 refer to the REW estimator. In columns 3 and 4, standard errors are computed analytically, whereas in columns 5 and 6, they are computed via bootstrap for the logit and probit models, respectively. These results can be retrieved by typing sequentially

```
. treatrew children educ7 age agesq evermarr urban electric tv, model(probit)
. treatrew children educ7 age agesq evermarr urban electric tv, model(logit)
. bootstrap e(ate) e(atet) e(atent), reps(200):
> treatrew children educ7 age agesq evermarr urban electric tv, model(probit)
. bootstrap e(ate) e(atet) e(atent), reps(200):
> treatrew children educ7 age agesq evermarr urban electric tv, model(logit)
```

Finally, column 7 presents an estimation of ATEs obtained by implementing a one-to-one nearest-neighbor matching on the propensity score (MATCH). Here the standard error for ATET is obtained analytically, whereas those for ATE and ATENT are computed by bootstrapping. Matching results can be obtained by typing

```
. psmatch2 educ7 age agesq evermarr urban electric tv, ate out(children) common
. bootstrap r(ate) r(atu): psmatch2 educ7 $xvars, ate out(children) common
```

where the option `common` restricts the sample to subjects with common support. To test the balancing property for such a matching estimation, we provide a DIM on the propensity score before and after matching treated and untreated subjects using the `psmatch2` postestimation command `pstest`:

```
. pstest _pscore, both
```

Variable	Unmatched Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
_pscore	Unmatched	.65692	.42546	111.7		37.05	0.000
	Matched	.65692	.65688	0.0	100.0	0.01	0.994

(*output omitted*)

This test suggests that with regard to the propensity score, the matching procedure implemented by `psmatch2` is balanced, so we can trust matching results (the propensity score was unbalanced before matching, and it becomes balanced after matching).

Unlike DIM, results from CF-OLS and REW are fairly comparable in terms of both coefficients' size and significance: the values of ATE, ATET, and ATENT obtained using REW on the propensity score are a little higher than those obtained using CF-OLS. This means that the linearity of the potential-outcome equations assumed by the CF-OLS is an acceptable approximation. According to the value of ATET, as obtained by REW and visible in column 3 of table 1, an educated woman in Botswana would have been—ceteris paribus—significantly more fertile if she had been less educated. We can conclude that education has a negative impact on fertility, leading a woman to have around 0.5 fewer children. If confounding variables were not considered, as it happens using DIM, this negative effect would appear dramatically higher, around 1.77 children: the difference between 1.77 and 0.5 (around 1.3) is an estimation of the bias induced by the presence of selection on observables.

Columns 3 and 4 show REW results using Wooldridge's (2010) analytical standard errors in the case of probit and logit, respectively. As partly expected, these results are similar. But the REW results when standard errors are obtained via bootstrap (columns 5 and 6) are more interesting. Here statistical significance is confirmed when compared with results derived from analytical formulas. However, bootstrapping seems to increase significance for both ATET and ATENT, while the standard error for ATE is in line with the analytical one.

Some differences in results emerge when applying the one-to-one nearest-neighbor matching (column 7) on this dataset. In this case, ATET becomes insignificant with a magnitude that is around one-third lower than that obtained by REW. As said above, the standard errors of ATE and ATENT are here obtained via bootstrap because `psmatch2` does not provide analytical solutions for these two parameters. Nevertheless, as proved by Abadie and Imbens (2008), bootstrap performance is generally poor in the case of matching, so these results have to be taken with some caution.

Finally, figure 1 sets out the estimated kernel density for the distribution of ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$) when `treatrew` is used with options `graphic` and `range(-30 30)`. It is evident that the distribution of ATET($\mathbf{x}$) is a bit more concentrated around its mean (equal to ATET) than the distribution of ATENT($\mathbf{x}$) is; this indicates that more educated women respond more homogeneously to a higher level of education. On the contrary, less educated women react more heterogeneously to a potential higher level of education.

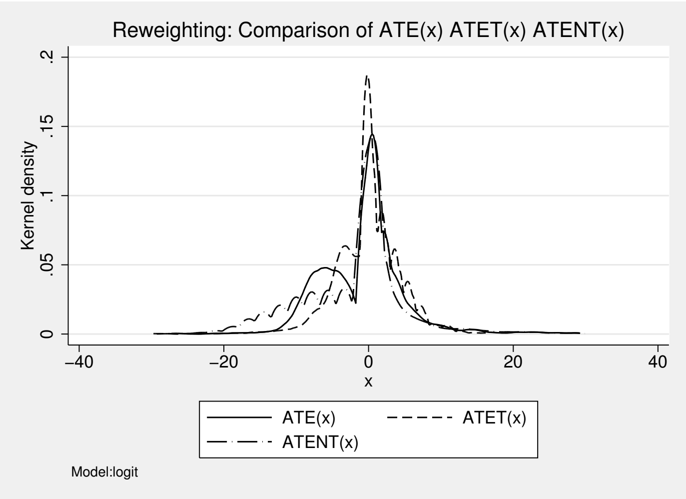

Figure 1. Estimation of the distribution of ATE($\mathbf{x}$), ATET($\mathbf{x}$), and ATENT($\mathbf{x}$) by REW on the propensity score with range equal to $(-30; 30)$

5 Relation between treatrew and Stata 13's teffects ipw

Stata 13 provides a new command, `teffects`, for estimating treatment effects for observational data. Among the many estimation methods provided by this command, `teffects ipw` implements a REW estimator based on IPW.

`teffects ipw` estimates the parameters ATE and ATET and the mean potential outcomes using a WLS regression where weights are a function of the propensity score estimated in the first step. To see the equivalence between IPW and WLS, we apply the `teffects ipw` command to our previous dataset by computing ATE.

```
. use fertil2
. teffects ipw (children) (educ7 $xvars, probit), ate
Iteration 0:   EE criterion =  6.624e-21
Iteration 1:   EE criterion =  4.722e-32
Treatment-effects estimation                   Number of obs      =      4358
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: probit
```

children	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
ATE						
educ7						
(1 vs 0)	-.1531253	.0755592	-2.03	0.043	-.3012187	-.0050319
POmean						
educ7						
0	2.208163	.0689856	32.01	0.000	2.072954	2.343372

In this estimation, we see that the value of ATE is -0.153 with a standard error of 0.075, which results in a moderately significant effect of `educ7` on `children`.

This value of ATE can also be obtained using a simple WLS regression of y on w and a constant, with weights h_i designed in this way:

$$h_i = h_{i1} = 1/p(\mathbf{x}_i) \quad \text{if } w_i = 1 \tag{6}$$

$$h_i = h_{i0} = 1/\{1 - p(\mathbf{x}_i)\} \quad \text{if } w_i = 0 \tag{7}$$

The Stata code for computing such a WLS regression is as follows:

```
. global xvars age agesq evermarr urban electric tv
. probit educ7 $xvars, robust // estimate the probit regression
  (output omitted)
. predict _ps,  p  // call the estimated propensity score as _ps
(3 missing values generated)
. generate H=(1/_ps)*educ7+1/(1-_ps)*(1-educ7) // weighing function H for w=1
> and w=0
(3 missing values generated)
. regress children educ7 [pw=H], vce(robust) // estimate ATE by a WLS regression
(sum of wgt is   9.1714e+03)
Linear regression                                      Number of obs =    4358
                                                       F(  1,  4356) =    2.00
                                                       Prob > F      =  0.1576
                                                       R-squared     =  0.0013
                                                       Root MSE      =  2.1324
```

children	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ7	-.1531253	.1083464	-1.41	0.158	-.3655393	.0592887
_cons	2.208163	.0867265	25.46	0.000	2.038135	2.378191

This table shows that the results of the commands calculating IPW and WLS for ATE are identical. A difference, however, appears in the estimated standard errors, which are quite divergent: 0.075 for IPW against 0.108 for WLS. Moreover, observe that ATE calculated by WLS becomes nonsignificant.

Why are these standard errors different? The answer resides in a different approach used for estimating the variance of ATE (and, possibly, ATET): WLS regression uses the usual OLS variance–covariance matrix adjusted for the presence of a matrix of weights, let's say $\mathbf{\Omega}$; however, WLS does not consider the presence of a generated regressor, namely, the weights computed through the propensity scores estimated in the first step. On the contrary, IPW accounts for the variability introduced by the generated weights by exploiting a generalized method of moments approach for estimating the correct variance–covariance matrix (see StataCorp [2013, 68–88]). In this sense, IPW is a more robust approach than a standard WLS regression.

As implemented in Stata, both WLS and IPW by default use normalized weights, that is, weights that add up to one. `treatrew`, on the contrary, uses nonnormalized weights, which is why the ATE values obtained from `treatrew` (see the previous section) are numerically different from those obtained from WLS and IPW. As proved by Busso, DiNardo, and McCrary (2008, 7), a general formula for estimating ATE by REW is

$$\widehat{\mathrm{ATE}} = \frac{1}{N}\sum_{i=1}^{N} w_i y_i h_{i1} - \frac{1}{N}\sum_{i=1}^{N}(1-w_i)y_i h_{i0} \tag{8}$$

`treatrew` uses nonnormalized inverse-probability weights defined as above; that is

$$h_{i1} = 1/p(\mathbf{x})$$
$$h_{i0} = 1/\{1-p(\mathbf{x}_i)\}$$

Such weights do not sum up to one. In this case, analytical standard errors cannot be retrieved by a weighted regression, and the method suggested by Wooldridge (2010)—and implemented through `treatrew`—for getting correct analytical standard errors for ATE, ATET, and ATENT is thus needed because a generated regressor from the first-step estimation is used in the second step.

The normalized weights used in WLS and IPW are instead

$$h_{i1} = \frac{1/p(\mathbf{x}_i)}{\frac{1}{N_1}\sum_{i=1}^{N} w_i/p(\mathbf{x}_i)}$$
$$h_{i0} = \frac{1/\{1-p(\mathbf{x}_i)\}}{\frac{1}{N_0}\sum_{i=1}^{N}(1-w_i)/\{1-p(\mathbf{x}_i)\}}$$

Appendix B shows that if the formula of ATE implemented in `treatrew` using normalized (rather than nonnormalized) weights was adopted, then the `treatrew`'s ATE estimation would become numerically equivalent to the value of ATE obtained by the commands used to calculate WLS and IPW.

Thus we can assert that both `teffects ipw` and `treatrew` lead to correct analytical standard errors because both take into account that the propensity score is a generated regressor from a first-step (probit or logit) regression. The different values of ATE and ATET obtained in the two approaches reside only in the different weighting scheme (normalized versus nonnormalized).

In short, `treatrew` is useful when considering nonnormalized weights, that is, when a pure IPW scheme is used. Moreover, compared with `teffects ipw`, `treatrew` provides an estimation of ATENT, though it does not by default provide an estimation of the mean potential outcomes.

6 Conclusion

This article provides a command, `treatrew`, for estimating ATEs by REW on the propensity score as proposed by Rosenbaum and Rubin (1983). Although REW is a popular and long-standing statistical technique to deal with the bias induced by drawing inference in the presence of a nonrandom sample, its implementation in Stata with parameters' analytic standard errors (as proposed by Wooldridge [2010, 920–930]) and a nonnormalized weighting scheme was still missing. This article and the accompanying ado-file fill this gap by providing an easy-to-use implementation of the REW method, which can be used as a valuable tool for estimating causal effects under selection on observables.

7 References

Abadie, A., and G. W. Imbens. 2008. On the failure of the bootstrap for matching estimators. *Econometrica* 76: 1537–1557.

Brunell, T. L., and J. DiNardo. 2004. A propensity score reweighting approach to estimating the partisan effects of full turnout in American presidential elections. *Political Analysis* 12: 28–45.

Busso, M., J. DiNardo, and J. McCrary. 2008. Finite sample properties of semiparametric estimators of average treatment effects. http://elsa.berkeley.edu/users/cle/laborlunch/mccrary.pdf.

Cerulli, G. 2011. ivtreatreg: A new Stata routine for estimating binary treatment models with heterogeneous response to treatment under observable and unobservable selection. 8th Italian Stata Users Group meeting proceedings. http://www.stata.com/meeting/italy11/abstracts/italy11_cerulli.pdf.

Hirano, K., G. W. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161–1189.

Horvitz, D. G., and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47: 663–685.

Leuven, E., and B. Sianesi. 2003. psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s432001.html.

Li, Q., J. S. Racine, and J. M. Wooldridge. 2009. Efficient estimation of average treatment effects with mixed categorical and continuous data. *Journal of Business and Economic Statistics* 27: 206–223.

Lunceford, J. K., and M. Davidian. 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* 23: 2937–2960.

Morgan, S. L., and D. J. Harding. 2006. Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research* 35: 3–60.

Nichols, A. 2007. Causal inference with observational data. *Stata Journal* 7: 507–541.

Robins, J. M., M. A. Hernán, and B. Brumback. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11: 550–560.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

StataCorp. 2013. *Stata 13 Treatment Effects Reference Manual: Potential Outcomes/Counterfactual Outcomes*. College Station, TX: Stata Press.

Wooldridge, J. M. 2007. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 141: 1281–1301.

———. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

———. 2013. *Introductory Econometrics: A Modern Approach*. 5th ed. Mason, OH: South-Western.

About the author

Giovanni Cerulli is a researcher at Ceris-CNR, National Research Council of Italy, Institute for Economic Research on Firms and Growth. He received a degree in statistics and a PhD in economic sciences from Sapienza University of Rome and is editor-in-chief of the *International Journal of Computational Economics and Econometrics*. His research interests are mainly on applied microeconometrics, with a focus on counterfactual treatment-effects models for program evaluation. Stata programming and simulation- and agent-based methods are also among his related fields of study. He has published articles in high-quality, refereed economics journals.

Appendix A

This appendix provides the mathematical steps to get the REW formulas for ATEs as reported in (1)–(3). Observe first that $wy = w\{wy_1 + y_0(1-w)\} = w^2 y_1 + wy_0 - w^2 y_0 = wy_1$ because $w^2 = w$. Therefore,

$$E\left\{\frac{wy}{p(\mathbf{x})}|\mathbf{x}\right\} = E\left\{\frac{wy_1}{p(\mathbf{x})}|\mathbf{x}\right\} \overset{\text{LIE2}}{=} E\left[E\left\{\frac{wy_1}{p(\mathbf{x})}|\mathbf{x},w\right\}|\mathbf{x}\right] = E\left\{\frac{wE(y_1|\mathbf{x},w)}{p(\mathbf{x})}|\mathbf{x}\right\}$$
$$\overset{\text{CMI}}{=} E\left\{\frac{wE(y_1|\mathbf{x})}{p(\mathbf{x})}|\mathbf{x}\right\} = E\left\{\frac{wg_1(\mathbf{x})}{p(\mathbf{x})}|\mathbf{x}\right\} = g_1(\mathbf{x}) \times E\left\{\frac{w}{p(\mathbf{x})}|\mathbf{x}\right\}$$
$$= \frac{g_1(\mathbf{x})}{p(\mathbf{x})} \times E(w|\mathbf{x}) = \frac{g_1(\mathbf{x})}{p(\mathbf{x})} \times p(\mathbf{x}) = g_1(\mathbf{x}) \tag{9}$$

because $E(w|\mathbf{x}) = p(\mathbf{x})$. Similarly, we can show that

$$E\left[\frac{(1-w)y}{\{1-p(\mathbf{x})\}}|\mathbf{x}\right] = g_0(\mathbf{x}) \tag{10}$$

Combining (9) and (10) we see that

$$\text{ATE}(\mathbf{x}) = g_1(\mathbf{x}) - g_0(\mathbf{x}) = E\left\{\frac{wy}{p(\mathbf{x})}|\mathbf{x}\right\} - E\left[\frac{(1-w)y}{\{1-p(\mathbf{x})\}}|\mathbf{x}\right] = E\left[\frac{\{w-p(\mathbf{x})\}y}{p(\mathbf{x})\{1-p(\mathbf{x})\}}|\mathbf{x}\right]$$

provided that $0 < p(\mathbf{x}) < 1$. To get ATE, one needs to take the expectation of $\text{ATE}(\mathbf{x})$ on $\mathbf{x}$,

$$\text{ATE} = E_{\mathbf{x}}\{\text{ATE}(\mathbf{x})\} = E_{\mathbf{x}}E\left[\frac{\{w-p(\mathbf{x})\}y}{p(\mathbf{x})\{1-p(\mathbf{x})\}}|\mathbf{x}\right] = E\left[\frac{\{w-p(\mathbf{x})\}y}{p(\mathbf{x})\{1-p(\mathbf{x})\}}\right]$$

that is, the inverse-probability REW estimation of ATE. Interestingly, it is possible to show that such an estimator is equivalent to the Horvitz–Thompson estimator (Horvitz and Thompson 1952). In sampling theory, it is a method for estimating the total and mean of a super population in a stratified sample. IPW is generally applied to account for different proportions of observations within strata in a target population.

Similarly, we can also calculate ATET by considering that

$$\begin{aligned}\{w-p(\mathbf{x})\}y &= \{w-p(\mathbf{x})\} \times \{y_0 + w \times (y_1 - y_0)\} \\ &= \{w-p(\mathbf{x})\} \times y_0 + w \times \{w-p(\mathbf{x})\} \times (y_1 - y_0) \\ &= \{w-p(\mathbf{x})\} \times y_0 + w \times \{1-p(\mathbf{x})\} \times (y_1 - y_0)\end{aligned}$$

because $w^2 = w$. Thus, by dividing the previous expression by $\{1-p(\mathbf{x})\}$, we get

$$\frac{\{w-p(\mathbf{x})\}y}{\{1-p(\mathbf{x})\}} = \frac{\{w-p(\mathbf{x})\}y_0}{\{1-p(\mathbf{x})\}} + w(y_1 - y_0) \tag{11}$$

Consider now the quantity $\{w-p(\mathbf{x})\}y_0$ in the right-hand side of (11). We see that

$$\begin{aligned}\{w-p(\mathbf{x})\}y_0 &= E[\{w-p(\mathbf{x})\}y_0|\mathbf{x}] = E(E[\{w-p(\mathbf{x})\}y_0|\mathbf{x},w]|\mathbf{x})\\ &= E[\{w-p(\mathbf{x})\} \times E(y_0|\mathbf{x},w)|\mathbf{x}] = E[\{w-p(\mathbf{x})\} \times E\{y_0|\mathbf{x}\}|\mathbf{x}]\\ &= E[\{w-p(\mathbf{x})\} \times g_0(\mathbf{x})|\mathbf{x}] = g_0(\mathbf{x}) \times E[\{w-p(\mathbf{x})\}|\mathbf{x}]\\ &= g_0(\mathbf{x}) \times [E(w|\mathbf{x}) - E\{p(\mathbf{x})|\mathbf{x}\}] = g_0(\mathbf{x}) \times \{p(\mathbf{x})-p(\mathbf{x})\} = 0\end{aligned}$$

Taking (11) and applying the expectation conditional on $\mathbf{x}$, we get

$$E\left[\frac{\{w-p(\mathbf{x})\}y}{\{1-p(\mathbf{x})\}}|\mathbf{x}\right] = E\left[\frac{\{w-p(\mathbf{x})\}y_0}{\{1-p(\mathbf{x})\}}|\mathbf{x}\right] + E\{w(y_1-y_0)|\mathbf{x}\} = E\{w(y_1-y_0)|\mathbf{x}\}$$

because we proved that $\{w-p(\mathbf{x})\}y_0$ is 0. By the law of iterated expectations (LIE), we get

$$\begin{cases} E_{\mathbf{x}}E\left[\dfrac{\{w-p(\mathbf{x})\}y}{\{1-p(\mathbf{x})\}}|\mathbf{x}\right] = E\left[\dfrac{\{w-p(\mathbf{x})\}y}{\{1-p(\mathbf{x})\}}\right] \\ E_{\mathbf{x}}E\,\{w(y_1-y_0)|\mathbf{x}\}\,E\{w(y_1-y_0)\} \end{cases} \tag{12}$$

that is,

$$E\left\{\frac{\{w-p(\mathbf{x})\}y}{\{1-p(\mathbf{x})\}}\right\} = E\{w(y_1-y_0)\}$$

Using LIE again, by assuming $h = w(y_1-y_0)$, we get

$$\begin{aligned}E(h) &= E\{w(y_1-y_0)\}\\ &= p(w=1) \times E\{w(y_1-y_0)|w=1\} + p(w=0) \times E\{w(y_1-y_0)|w=0\}\\ &= p(w=1) \times E\{(y_1-y_0)|w=1\}\\ &= p(w=1) \times \text{ATET}\end{aligned}$$

This means that

$$E\left[\frac{\{w-p(\mathbf{x})\}y}{\{1-p(\mathbf{x})\}}\right] = E\{w(y_1-y_0)\} = p(w=1) \times \text{ATET}$$

proving that

$$\text{ATET} = E\left[\frac{\{w-p(\mathbf{x})\}y}{p(w=1)\{1-p(\mathbf{x})\}}\right]$$

Finally, by remembering that $\text{ATE} = p(w=1) \times \text{ATET} + p(w=0) \times \text{ATENT}$, we can also prove that

$$\text{ATENT} = E\left[\frac{\{w-p(\mathbf{x})\}y}{p(w=0)p(\mathbf{x})}\right]$$

Appendix B

In this appendix, we show that if one considers the formula of ATE as implemented in `treatrew` by using normalized rather than nonnormalized weights, then `treatrew`'s ATE estimation becomes numerically equivalent to the ATE obtained by commands used to calculate WLS and IPW. To this purpose, we first calculate the ATE estimator by means of the general formula in (8) by adopting normalized IPW weights:

$$\widehat{\text{ATE}} = \frac{1}{N}\sum_{i=1}^{N} w_i y_i h_{i1} - \frac{1}{N}\sum_{i=1}^{N}(1-w_i) y_i h_{i0}$$

As an intermediary step, we show that normalized weights sum up to one for the weights of both the treated and the untreated subjects.

```
* Weights sum up to one for "treated"
. generate h1 = educ7/_ps        // observe that educ7=w
. summarize h1
. scalar sum_h1 = _N*r(mean)
. summarize educ7 if educ7==1
. scalar mean_h1 = (1/r(N))*sum_h1
. generate H1 = (1/_ps)/mean_h1  // H1 is the normalized weight for treated units
. generate m1 = educ7*H1         // m1 is equal to w*h1 using h1=H1
. summarize m1
. scalar tot_m1 = _N*r(mean)
. summarize educ7 if educ7==1
. scalar N1 = r(N)
. scalar one1 = (1/N1)* tot_m1
. display one1
  1 // ok

* Weights sum up to one for "untreated"
. generate h0 = (1-educ7)/(1-_ps)
. summarize h0
. scalar sum_h0 = _N*r(mean)
. summarize educ7 if educ7==0
. scalar mean_h0 = (1/r(N))*sum_h0
. generate H0 = (1/(1-_ps))/mean_h0 // H0 is the normalized weight for
> untreated units
. generate m0 = (1-educ7)*H0        // m0 is equal to (1-w)*h0 using h0=H0
. summarize m0
. scalar tot_m0 = _N*r(mean)
. summarize educ7 if educ7==0
. scalar N0 = r(N)
. scalar one0 = (1/N0)* tot_m0
. display one0
  1 // ok
```

Second, we compute the estimation of ATE by multiplying the two summands for the treated and untreated units in (8) by the outcome y (equal in this example to the variable `children`):

```
* Average outcome for treated units
. generate s1 = children*educ7*H1  // s1 is the summand y*w*h1 of (8) with h1=H
. summarize s1
. scalar tot_s1 = _N*r(mean)
. summarize educ7 if educ7==1
. scalar N1 = r(N)
. scalar _s1 = (1/N1)* tot_s1      // _s1 is the average outcome for treated units
. display _s1
  2.0550377

* Average outcome for untreated units
. generate s0 = children*(1-educ7)*H0 // s0 is y*(1-w)*h0 of (8) with h0=H0
. summarize s0
. scalar tot_s0 = _N*r(mean)
. summarize educ7 if educ7==0
. scalar N0 = r(N)
. scalar _s0 = (1/N0)* tot_s0  // _s0 is the average outcome for untreated units
. display _s0
  2.208163
```

We see that the ATE is the difference between `_s1` and `_s0`,

```
. display _s1 - _s0  //  ok
 -.15312
```

which is numerically equivalent to the value of the ATE obtained via WLS and IPW.