Rapporto Tecnico N. 43
Novembre 2012

treatrew: a user-written STATA routine
for estimating Average Treatment Effects
by reweighting on propensity score

Giovanni Cerulli
treatrew: a user-written STATA routine for estimating Average Treatment Effects by reweighting on propensity score

Giovanni Cerulli
National Research Council of Italy
Ceris-CNR
Institute for Economic Research on Firms and Growth
Via dei Taurini, 19 - 00185 Roma
Phone: +39.06.4993.7867
E-mail: g.cerulli@ceris.cnr.it

ABSTRACT: Reweighting is a popular statistical technique to deal with inference in presence of a non-random sample. In the literature, various reweighting estimators have been proposed. This paper presents the user-written STATA command treatrew implementing the reweighting on propensity score estimator as proposed by Rosenbaum and Rubin (1983) in their seminal article, where parameters’ standard errors can be obtained either analytically or via bootstrapping. Since an implementation in STATA of this estimator with analytic standard errors was still missing, this paper (and the ADO-file and HELP-file accompanying it) aims at filling this gap by providing the community with an easy-to-use implementation of the reweighting on propensity score method, as a valuable tool for estimating treatment effects under selection on observables.

JEL code: C21, C83, C87

Keywords: treatment models; reweighting; propensity score; Stata program
SUMMARY

1. Introduction 5
2. The reweighting estimator of treatment effects: a brief overview 5
3. Sample estimation and standard errors for ATE, ATET and ATENT 11
4. The STATA routine treatrew: syntax and use 14
5. Conclusion 18
References 19
1. Introduction

treatrew is a STATA routine for estimating Average Treatment Effects by reweighting on propensity score. Depending on the model specified (probit or logit), treatrew provides consistent estimation of Average Treatment Effects under the hypothesis of “selection on observables”. Conditional on a pre-specified set of observable exogenous variables $x$ – thought of as those driving the non-random assignment to treatment – treatrew estimates the Average Treatment Effect (ATE), the Average Treatment Effect on Treated (ATET) and the Average Treatment Effect on Non-Treated (ATENT), as well as these parameters conditional on the observable factors $x$ (i.e., ATE($x$), ATET($x$) and ATENT($x$)).

In the literature, a plethora of reweighting estimators have been proposed. This paper presents the user-written STATA command treatrew implementing the reweighting on propensity score estimator as proposed by Rosenbaum and Rubin (1983) in their seminal article, where parameters’ standard errors can be obtained either analytically (using Wooldridge, 2010, p. 920-930) or via bootstrapping. The command treatrew assumes that the propensity score specified by the user is correct. Thus, it is severely sensitive to propensity score misspecification.

The paper is organized as follows: section 2 provides the statistical description of the reweighting on propensity score estimator; section 3 provides the formulas for calculating the causal parameters and their standard errors; section 4 presents the syntax of treatrew (i.e., its STATA help-file), and an application to real data; section 5 concludes the paper.

2. The reweighting estimator of treatment effects: a brief overview

Reweighting is a valuable approach to estimate (binary) treatment effects in a non-experimental statistical setting, when units’ non-random assignment to treatment is due to selection on observables. The idea behind the reweighting procedure is quite straightforward: when the treatment is not randomly assigned, we expect that the treated and untreated units present very different distributions of their observable characteristics. This may happen either because of units’ self-selection into the experiment (units may ponder the net cost-benefit gain of participating or not), or because of the selection process operated by an external entity (such as, for instance, a public agency managing a subsidization program whose explicit objective is that of selecting beneficiaries with peculiar characteristics to maximize the policy effect). Many examples can be drawn both from social and epidemiological statistical settings.

If this is the case, the distribution of the variables feeding into $x$ could be strongly unbalanced. To establish again a balance in these distributions, a suitable way could be that of “reweighting” the observations using their “probability of becoming treated”, that is, according to their propensity score. A possible reweighting estimation protocol is as follows:

1. estimate the propensity score (based on $x$) by a Logit or a Probit getting the predicated probability $p_i$;
2. build weights as $1/p_i$ for the treated observations, and $1/(1-p_i)$ for the untreated observations;
3. calculate the ATE simply by a comparison of the weighted means of the two groups (for instance, by a weighted regression).

This particular weighting scheme is based on the inverse probability regression (Robins et al., 2000; Brunell and Dinardo, 2004) and the intuitive idea is that of penalizing (advantaging) treated units with higher (lower)
probability to be treated, and advantaging (penalizing) untreated units with higher (lower) probability to be treated, thus rendering the two groups as more similar as possible. In other words, for each observation the weight wipes out a component induced by the extent of the non-random assignment to the program (a confounding element).

Various alternative weighting schemes have been proposed in the literature\(^1\) and some authors have shown as the various Matching methods can be also seen as specific reweighting estimators (Lunceford and Davidian, 2004; Morgan and Harding, 2006). As in the case of Matching, these estimators have different properties but the main limits reside in the specification of the propensity score, as measurement errors in this specification could produce severe bias. In what follows we focus on reweighting on propensity-score inverse probability as proposed by Rosenbaum and Rubin (1983) in their seminal paper. In this case, we start by the following assumptions about the data generating process (DGP):\(^2\)

i. \( y_1 = g_0(x) + e_0 \), \( E(e_0)=0 \)
ii. \( y_0 = g_1(x) + e_1 \), \( E(e_1) = 0 \)
iii. \( y = y_1 + y_0 (1-w) \)
iv. Conditional Mean Independence (CMI) holds, so that:
\[
E(y_1|w, x) = E(y_1| x) \quad \text{and} \quad E(y_0|w, x) = E(y_0| x)
\]
v. \( x \) exogenous,

where \( y_1 \) and \( y_0 \) is the unit’s outcome when it is treated and untreated respectively; \( g_1(x) \) and \( g_0(x) \) is the unit’s reaction function to the confounder \( x \) when the unit is treated and untreated respectively; \( e_0 \) and \( e_1 \) are two errors with unconditional zero mean; \( x \) is a set of observable exogenous confounding variables assumed to drive the non-random assignment into treatment. Very concisely, the CMI assumption states that it is sufficient to control only for \( x \) in order to restore random assignment conditions. We now prove that, when assumptions i-v hold, then:

\[
\text{ATE} = E \left\{ \frac{[w - p(x)]y}{p(x)[1 - p(x)]} \right\}
\]
\[
\text{ATET} = E \left\{ \frac{[w - p(x)]y}{p(w = 1)[1 - p(x)]} \right\}
\]
\[
\text{ATENT} = E \left\{ \frac{[w - p(x)]y}{p(w = 0)p(x)} \right\}
\]

To this purpose, observe first that: \( wy = w[y_1 + y_0 (1-w)] = wy_1 + w y_0 - w^2 y_0 = w y_1 \), since \( w^2 = w \).

---

1 Another possible weighting scheme could that of assuming \( p_i/(1-p_i) \) for the untreated units and 1 for the treated ones (Austin, 2007)

2 As reminder, we consider the following versions of the Law of Iterated Expectations: LIE 1: \( E_r(y) = \mu_r = E_x[E_r(y|x)] \); LIE 2: \( E_r(y|x) = \mu_x = E_x[E_r(y|x, z)|x] = E_x[\mu_1(x, z)|x] \); LIE 3: \( E(h) = p_1 \cdot E(h|x_1) + p_2 \cdot E(h|x_2) + \ldots + p_M \cdot E(h|x_M) \).
Therefore:

\[
E \left\{ \frac{wy}{p(x)} \mid x \right\} = E \left\{ \frac{wy_i}{p(x)} \mid x \right\} \overset{\text{LIE2}}{=} E \left\{ \frac{wy_i}{p(x)} \mid x, w \right\} = E \left\{ wE(y_i|x, w) \mid x \right\} \overset{\text{CMI}}{=}
\]

\[
E \left\{ \frac{wE(y_i|x)}{p(x)} \mid x \right\} = E \left\{ \frac{wg_i(x)}{p(x)} \mid x \right\} = g_1(x) \cdot E \left\{ \frac{w}{p(x)} \mid x \right\} = \frac{g_1(x)}{p(x)} \cdot p(x) = g_1(x)
\]

because: \( E(w \mid x) = p(x) \). Similarly, we can show that:

\[
E \left\{ \frac{(1-w)y}{[1-p(x)]} \mid x \right\} = g_0(x)
\]

Combining (1) and (2) we have that:

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

provided that \( 0 < p(x) < 1 \). To get the ATE, it is sufficient to take the expectation on \( x \):

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

It is interesting to show that the previous formula for ATE is equal to the famous Horvitz-Thompson estimator of the population mean. Indeed:

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

\[
= E \left\{ \frac{p(x)wy}{p(x)[1-p(x)]} + \frac{wy[1-p(x)]}{p(x)[1-p(x)]} - \frac{p(x)y}{p(x)[1-p(x)]} \right\} =
\]

\[
= E \left\{ \frac{wy}{1-p(x)} + \frac{wy}{1-p(x)} - \frac{y}{1-p(x)} \right\} = E \left\{ \frac{wy - y}{1-p(x)} + \frac{wy}{p(x)} \right\} =
\]

\[
= E \left\{ \frac{wy}{p(x)} - \frac{y-wy}{1-p(x)} \right\} = E \left\{ \frac{wy}{p(x)} - \frac{(1-w)y}{1-p(x)} \right\} =
\]

\[
= E \left\{ \frac{wy}{p(x)} \right\} - E \left\{ \frac{(1-w)y}{1-p(x)} \right\}
\]
Summing-up:

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

whose sample equivalent is:

\[ \hat{\text{ATE}} = \frac{1}{N} \sum_{i=1}^{N} \frac{w_i y_i}{p(x_i)} - \frac{1}{N} \sum_{i=1}^{N} \frac{(1-w_i) y_i}{1-p(x_i)} \]  

An example can provide the proof. Suppose to get a dataset \( \{y, w, x\} \) as in Table 1.

### Table 1. Dataset coming from a non-experimental statistical setting.

<table>
<thead>
<tr>
<th>id</th>
<th>y</th>
<th>w</th>
<th>Inclusion probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y_1</td>
<td>1</td>
<td>( \pi_1 = p_1(x) )</td>
</tr>
<tr>
<td>2</td>
<td>y_2</td>
<td>0</td>
<td>( \pi_2 = 1 - p_2(x) )</td>
</tr>
<tr>
<td>3</td>
<td>y_3</td>
<td>1</td>
<td>( \pi_3 = p_3(x) )</td>
</tr>
<tr>
<td>4</td>
<td>y_4</td>
<td>1</td>
<td>( \pi_4 = p_4(x) )</td>
</tr>
<tr>
<td>5</td>
<td>y_5</td>
<td>0</td>
<td>( \pi_5 = 1 - p_5(x) )</td>
</tr>
</tbody>
</table>

If we define the **inclusion probability** of unit \( i \) into the sample \( S \) as:

\[ \pi_i = \Pr\{i \Rightarrow S\} \]

it is immediate to see that:

- for **treated** units the inclusion probability is equal to the propensity score: \( p(w=1 \mid x) \);
- for **untreated** units the inclusion probability is equal to: \( p(w=0 \mid x) = 1 - \Pr(w=1 \mid x) \).

Thus, by applying to this example the last formula for ATE, we get:

\[
\hat{\text{ATE}} = \frac{1}{5} \left[ \frac{y_1}{p(x_1)} + \frac{y_3}{p(x_3)} + \frac{y_4}{p(x_4)} \right] - \frac{1}{5} \left[ \frac{y_2}{1-p(x_2)} + \frac{y_5}{1-p(x_5)} \right] = \\
= \frac{1}{5} \left[ \frac{y_1}{p(x_1)} + \frac{y_3}{p(x_3)} + \frac{y_4}{p(x_4)} + \frac{y_2}{1-p(x_2)} + \frac{y_5}{1-p(x_5)} \right] = \\
= \frac{1}{5} \left[ \frac{y_1}{\pi_1} + \frac{y_2}{\pi_2} + \frac{y_3}{\pi_3} + \frac{y_4}{\pi_4} + \frac{y_5}{\pi_5} \right] = \frac{1}{5} \sum_{i=1}^{5} \frac{y_i}{\pi_i} 
\]
Therefore, we have proved that:

\[ \hat{\text{ATE}} = \hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i}{\pi_i} \]

The Inverse Probability Reweighting estimation of ATE is thus equivalent to the Horvitz–Thompson estimator, due to Daniel G. Horvitz and Donovan J. Thompson in 1952. In sampling theory it is a method for estimating the total and mean of a super-population in a stratified sample. Inverse probability weighting is applied to account for “different proportions of observations within strata in a target population”. The Horvitz-Thompson estimator is frequently applied in survey analyses and can be used also to account for missing data.

Similarly, we can also calculate the ATET by considering that:

\[
[w - p(x)] y = [w - p(x)] \cdot [y_0 + w \cdot (y_1 - y_0)] = [w - p(x)] \cdot y_0 + w \cdot [w - p(x)] \cdot (y_1 - y_0) =
\]

\[= [w - p(x)] \cdot y_0 + w \cdot [1 - p(x)] \cdot (y_1 - y_0), \text{ since } w^2 = w.\]

Thus, by dividing the previous expression by \([1 - p(x)]\):

\[
\frac{[w - p(x)] y}{[1 - p(x)]} = \frac{[w - p(x)] y_0}{[1 - p(x)]} + w (y_1 - y_0) \tag{4}
\]

Consider now the quantity \([w - p(x)] y_0\) in the RHS of (4). We have that:

\[
[w - p(x)] y_0 = E([w - p(x)] y_0 | x) = E(E([w - p(x)] y_0 | x, w) | x) = E([w - p(x)] \cdot E[y_0 | x, w] | x) = E([w - p(x)] \cdot g_0(x) | x) = g_0(x) \cdot E([w - p(x)] | x) - E(p(x) | x) = g_0(x) \cdot [p(x) - p(x)] = 0.
\]

Taking relation (3), and applying the expectation conditional on \(x\):

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

since we proved that \([w - p(x)] y_0\) is zero. By LIE 1, we can get that:

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

\]
that is:
\[
E \left\{ \frac{w - p(x)}{1 - p(x)} \right\} = E\{w(y_i - y_0)\}
\]

From the LIE 3 we know that, if \( x \) is a generic discrete variable assuming values \( x = (x_1, x_2, \ldots, x_M) \) with probabilities \( p = (p_1, p_2, \ldots, p_M) \), then:
\[
E(h) = p_1 \cdot E(h \mid x_1) + p_2 \cdot E(h \mid x_2) + \ldots + p_M \cdot E(h \mid x_M).
\]

Thus, by assuming \( h = w(y_1 - y_0) \), we get that:
\[
E(h) = E[w(y_1 - y_0)] = p(w=1) \cdot E[w(y_1 - y_0) \mid w=1] + p(w=0) \cdot E[w(y_1 - y_0) \mid w=0] =
\]
\[
= p(w=1) \cdot E[(y_1 - y_0) \mid w=1] = p(w=1) \cdot ATET.
\]

It means that:
\[
E \left\{ \frac{w - p(x)}{1 - p(x)} \right\} = E\{w(y_i - y_0)\} = p(w=1) \cdot ATET
\]

proving that:
\[
ATET = E \left\{ \frac{w - p(x)}{p(w=1)(1 - p(x))} \right\}
\]

(proving that:
\[
ATET = E \left\{ \frac{w - p(x)}{p(w=1)(1 - p(x))} \right\}
\]

Now, by remembering that \( ATE = p(w=1) \cdot ATET + p(w=0) \cdot ATENT \), we have that:
\[
ATENT = \frac{ATE}{p(w=0)} - \frac{p(w=1)}{p(w=0)} \cdot ATET =
\]
\[
= \frac{1}{p(w=0)} E \left\{ \frac{w - p(x)}{p(x)(1 - p(x))} \right\} - \frac{p(w=1)}{p(w=0)} \cdot E \left\{ \frac{w - p(x)}{1 - p(x)} \right\} =
\]
\[
= \frac{1}{p(w=0)} E \left\{ \frac{w - p(x)}{p(x)(1 - p(x))} \right\} - \frac{1}{p(w=0)} \cdot E \left\{ \frac{w - p(x)}{1 - p(x)} \right\} =
\]
\[
= \frac{1}{p(w=0)} E \left\{ \frac{w - p(x)}{p(x)(1 - p(x))} \right\} = \frac{1}{p(w=0)} E \left\{ \frac{w - p(x)}{p(x)} \right\} =
\]
\[
= E \left\{ \frac{w - p(x)}{p(x)} \right\}
\]
This implies, finally, that:

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

(6)

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

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

\[
\hat{\text{ATE}} = \frac{1}{N} \sum_{i=1}^{N} \frac{[w_i - \hat{p}(x_i)]y_i}{\hat{p}(x_i)[1 - \hat{p}(x_i)]}
\]

\[
\hat{\text{ATET}} = \frac{1}{N} \sum_{i=1}^{N} \frac{[w_i - \hat{p}(x_i)]y_i}{\hat{p}(w = 1)[1 - \hat{p}(x_i)]}
\]

\[
\hat{\text{ATENT}} = \frac{1}{N} \sum_{i=1}^{N} \frac{[w_i - \hat{p}(x_i)]y_i}{\hat{p}(w = 0)\hat{p}(x_i)}
\]

Estimation follows in two steps: (i) estimate the propensity score \( p(x) \) getting \( \hat{p}(x) \); (ii) substitute \( \hat{p}(x) \) into the formulas to get the parameter. Observe that consistency is guaranteed by the fact that these estimators are M-estimators. How to get the standard errors for previous estimations? We can exploit some results from the case in which the first step is a ML-estimation and the second step is a M-estimation. In our case, the first step is a ML based on logit or probit and the second step is a standard M-estimator. For case like this, Wooldridge (2007; 2010, p. 922-924) has proposed a straightforward procedure to get standard errors, provided that the propensity score is correctly specified. In what follows, we set out the Wooldridge’s procedure and formulas for obtaining these (analytical) standard errors.

(i) Standard Errors estimation for ATE

First: define the estimated ML-score of the first step (Probit or Logit). It is, by definition equal to:

\[
\hat{d}_i = \hat{d}(w, x, \gamma) = \frac{[\nabla \hat{p}(x, \gamma)]' [w_i - \hat{p}(x, \gamma)]}{\hat{p}(x, \gamma)[1 - \hat{p}(x, \gamma)]}
\]

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

Second: define the generic estimated summand of ATE as:

\[
\hat{k}_i = \frac{[w_i - \hat{p}(x_i)]y_i}{\hat{p}(x_i)[1 - \hat{p}(x_i)]}
\]
Third: calculate the OLS residuals from this regression:

\[ \hat{k}_i \quad \text{on} \quad (1, \hat{d}_i') \quad \text{with} \quad i = 1, \ldots, N \]

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

\[
\left[ \frac{1}{N} \sum_{i=1}^{N} \hat{e}_i^2 \right]^{1/2}
\]

and we can use it to test the significance of ATE. Of course, \( d \) will have a different expression according to the probability model considered. Here, we consider the logit and probit case.

**Case 1: Logit**
Suppose that the correct probability follows a logistic distribution. It means that:

\[
p(x_i, \gamma) = \frac{\exp(x_i \gamma)}{1 + \exp(x_i \gamma)} = \Lambda(x_i, \gamma)
\]

Thus, by simple algebra, we get that:

\[
\hat{d}_i' = x_i (w_i - \hat{p}_i)
\]

**Case 2: Probit**
Suppose that the right probability follows a Normal distribution. It means that:

\[
p(x_i, \gamma) = \Phi(x_i, \gamma)
\]

Thus, by simple algebra, we get that:

\[
\hat{d}_i' = \frac{\phi(x_i, \hat{\gamma}) x_i \cdot [w_i - \Phi(x_i, \gamma)]}{\Phi(x_i, \gamma)[1 - \Phi(x_i, \gamma)]}
\]

Observe that one can add also functions of \( x \) to estimate previous formulas. This reduces the standard errors if these functions are partially correlated with \( k \).
Finally, observe that the previous procedure produces standard errors that are lower than those produced by ignoring the first step (i.e., the propensity score estimation via ML). Indeed, the naïve standard error:

\[
\left[ \frac{1}{N} \sum_{i=1}^{N} (\hat{k}_i - \hat{ATE})^2 \right]^{1/2}
\]

\sqrt{N}

is higher than the one produced by the previous procedure.

(ii) Standard error for ATET

It follows a route similar to ATE. In fact, define:

\[
\hat{q}_i = \frac{[w_i - \hat{p}(x_i)]y_i}{\hat{p}(w = 1)[1 - \hat{p}_i(x)]}
\]

and calculate:

\[
\hat{r}_i = \text{residuals form the regression of } "\hat{q}_i \text{ on 1, } \hat{d}"
\]

Then, the asymptotic standard error for ATET is given by:

\[
\left[ \hat{p}(w = 1) \right]^{-1} \cdot \left[ \frac{1}{N} \sum_{i=1}^{N} (\hat{r}_i - w_i \cdot \hat{ATE})^2 \right]^{1/2}
\]

\sqrt{N}

(iii) Standard error for ATENT

In this case, define:

\[
\hat{b}_i = \frac{[w_i - \hat{p}_i(x_i)]y_i}{\hat{p}(w = 0) \hat{p}(x_i)]}
\]

and then calculate:

\[
\hat{s}_i = \text{residuals form the regression of } "\hat{b}_i \text{ on 1, } \hat{d}"
\]
The asymptotic standard error for ATENT is:

\[
\left[ \hat{p}(w = 0)^{-1} \cdot \frac{1}{N} \sum_{i=1}^{N} (\hat{s}_i - (1 - w_i) \cdot \text{ATENT})^2 \right]^{1/2}
\]

Previous standard errors are correct as soon as the probit or the logit are the correct probability rules in the DGP. If not, then a measurement error is present and previous estimation might be inconsistent. The literature has provided more flexible non-parametric estimation of previous standard errors, such as in Hirano, Imbens, and Ridder (2003) or in Li, Racine, and Wooldridge (2009). 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 STATA routine treatrew: syntax and use

treatrew estimates the previous parameters with either analytical or bootstrapped standard errors. The syntax is rather simple and follows the typical STATA command syntax. The help of the command is reported in Table 2. The user has to declare: (a) the outcome variable, i.e. the variable over which the treatment is expect 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” sets the type of model, probit or logit, that has to be used in the estimation of the propensity score; the option “graphic” produces a chart where the distribution of ATE(x), ATET(x) and ATENT(x) are jointly reported.

As e-class command, treatrew provides a ereturn list of objects (such as scalars and matrices) to be used in 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. Observe that, by default, treatrew provides analytical standard errors.

To show a practical application of treatrew, we employ an instructional dataset called FERTIL2.DTA accompanying the manual “Introductory Econometrics: A Modern Approach”, by Wooldridge (2000) collecting 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 various woman 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 or not the family owns a TV (tv), whether or not the woman lives in a city (urban), and so forth. In order to inquiry into the relation between education and fertility and according to Wooldridge (2002, example 18.3, p. 624) specification we estimate ATE, ATET and ATENT (as well as ATE(x), ATET(x) and ATENT(x)) by “reweighting” using treatrew. We also compare reweighting results with those from a regression analysis based on a random-coefficient specification with “heterogeneous reaction to confounders” (see the user-written STATA command ivtreatreg for an implementation of this method (Cerulli, 2012)), and the usual “Difference-In-Mean” (DIM) as benchmark. All the results are reported in Table 3.
Table 2. The `treatrew` help file.

**Title**

`treatrew` - Estimation of Average Treatment Effects by reweighting on propensity score

**Syntax**

```bash
treatrew outcome treatment [varlist] [if] [in] [weight], model(modeltype) [Graph conf(number) vce(robust)]
```

fweights, iweights, and pweights are allowed; see weight.

**Description**

`treatrew` estimates Average Treatment Effects by reweighting on propensity score. Depending on the model specified, `treatrew` provides consistent estimation of Average Treatment Effects under the hypothesis of "selection on observables". Conditional on a pre-specified set of observable exogenous variables $x$ - thought of as those driving the non-random assignment to treatment - `treatrew` estimates the Average Treatment Effect (ATE), the Average Treatment Effect on Treated (ATET) and the Average Treatment Effect on Non-Treated (ATENT), as well as the estimates of these parameters conditional on the observable factors $x$ (i.e., $ATE(x)$, $ATET(x)$ and $ATENT(x)$). Parameters standard errors are provided either analytically (following Wooldridge, 2010, p. 920-930) and via bootstrapping. `treatrew` assumes that the propensity score specification is correct.

According to the syntax:

- `outcome`: is the target variable over which measuring the impact of the treatment
- `treatment`: is the binary treatment variable taking 1 for treated, and 0 for untreated units
- `varlist`: is the set of pre-treatment (or observable confounding) variables

**Options**

`model(modeltype)` specifies the model for estimating the propensity score, where modeltype must be one out of these two: "probit" or "logit". It is always required to specify one model.

`graphic` allows for a graphical representation of the density distributions of $ATE(x)$, $ATET(x)$ and $ATENT(x)$.
vce(robust) allows for robust regression standard errors in the probit or logit estimates.

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

<table>
<thead>
<tr>
<th>modeltype_options</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>probit</td>
<td>The propensity score is estimated by a probit regression</td>
</tr>
<tr>
<td>logit</td>
<td>The propensity score is estimated by a logit regression</td>
</tr>
</tbody>
</table>

**treatrew** creates a number of variables:

ATE_x is an estimate of the idiosyncratic Average Treatment Effect.

ATET_x is an estimate of the idiosyncratic Average Treatment Effect on treated.

ATENT_x is an estimate of the idiosyncratic Average Treatment Effect on Non-Treated.

**treatrew** returns the following scalars:

e(N) is the total number of (used) observations.

e(N1) is the number of (used) treated units.

e(N0) is the number of (used) untreated units.

e(ate) is the value of the Average Treatment Effect.

e(atet) is the value of the Average Treatment Effect on Treated.

e(atent) is the value of the Average Treatment Effect on Non-treated.
Results on the column DIM are obtained by typing:

```
.xi: reg children educ7
```

whereas, results on the column CF-OLS are obtained by typing:

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

*
\ Bootstrapping std. err. for "atet" and "atent":

```
.xi: 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)
```

Finally, results on the column REW are obtained by typing:

```
.xi: treatrew children educ7 age agesq evermarr urban electric tv , ///
model(probit)
```

<table>
<thead>
<tr>
<th></th>
<th>DIM</th>
<th>CF-OLS</th>
<th>REW</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATE</td>
<td>-1.77 ***</td>
<td>-0.374 ***</td>
<td>-0.434 ***</td>
</tr>
<tr>
<td></td>
<td>0.062</td>
<td>0.051</td>
<td>0.068</td>
</tr>
<tr>
<td>ATET</td>
<td>-28.46</td>
<td>-7.35</td>
<td>-6.34</td>
</tr>
<tr>
<td>ATENT</td>
<td>-0.523 ***</td>
<td>-0.355 **</td>
<td>-0.532</td>
</tr>
<tr>
<td></td>
<td>0.075</td>
<td>0.19 ***</td>
<td>-2.81</td>
</tr>
<tr>
<td></td>
<td>-7.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: b/se/t; DIM: Difference-in-Mean; CF-OLS: Control-function OLS; REW: Reweighting on propensity score.

Unlike DIM, results from CF-OLS and REW are fairly comparable both in terms of coefficients’ size and significance: ATE, ATET and ATENT values obtained using reweighting on propensity score are just a little higher than those obtained by CF-OLS. It means that the linearity of the outcome equations assumed by the CF-OLS is an acceptable approximation. According to the value of ATET in the REW column of Table 3, 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 to have around 0.5 children less. Observe that, if confounding variables are not considered as it happens using DIM, this negative effect would appear dramatically higher and 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. Finally, Figure 1 sets out the estimated kernel density for the distribution of ATE(x), ATET(x) and ATENT(x) when treatrew is used with the option “graphic”.
Figure 1. Estimation of the distribution of $\text{ATE}(x)$, $\text{ATET}(x)$ and $\text{ATENT}(x)$ by reweighting on propensity score.

5. Conclusion

This paper provides a simple STATA routine, treatrew, for estimating average treatment effects by reweighting on propensity score as proposed by Rosenbaum and Rubin (1983). Although reweighting is a popular and long-standing statistical technique to deal with the bias induced by drawing inference in presence of a non-random sample, its implementation in STATA with parameters’ analytic standard errors (as proposed by Wooldridge, 2010, p. 920-930) was still missing. This routine fills this gap by providing an easy-to-use implementation of the reweighting method, thought of as a valuable tool for estimating causal effects under selection on observables.
References


