The Stata Journal (2011)
11, Number 4, pp. 1–11

Treatment interactions with nonexperimental data in Stata

Graham K. Brown
Centre for Development Studies
University of Bath
Bath, UK
g.k.brown@bath.ac.uk

Thanos Mergoupis
Department of Economics
University of Bath
Bath, UK
a.mergoupis@bath.ac.uk

Abstract. Treatment effects may vary with the observed characteristics of the treated, often with important implications. In the context of experimental data, a growing literature deals with the problem of specifying treatment interaction terms that most effectively capture this variation. Some results of this literature are now implemented in Stata. With nonexperimental (observational) data, and in particular when selection into treatment depends on unmeasured factors, treatment effects can be estimated using Stata's `treatreg` command. Though not originally designed for this purpose, `treatreg` can be used to consistently estimate treatment interaction parameters. With interactions, however, adjustments are required to generate predicted values and estimate the average treatment effect. In this article, we introduce commands that perform this adjustment for multiplicative interactions, and we show the required adjustment for more complicated interactions.

Keywords: st0001, itreatreg, treatment-effects models, interaction terms

1 Introduction

Treatment effects may vary with the observed characteristics of the treated, often with important implications (Royston and Sauerbrei 2008). In the context of experimental data, a growing literature deals with the problem of specifying treatment interaction terms that most effectively capture this variation (see Sauerbrei, Royston, and Zapien [2007], for references). Some results of this literature are now implemented in Stata (Royston and Sauerbrei 2009). With nonexperimental (observational) data, and in particular when selection into treatment depends on unmeasured factors, treatment effects can be estimated using the Stata `treatreg` (see [R] **treatreg**) command. Though not originally designed for this purpose, `treatreg` can be used to consistently estimate treatment interaction parameters. With interactions, however, adjustments are required to generate predicted values and to estimate the average treatment effect (ATE). In this article, we introduce commands that perform this adjustment for multiplicative interactions, and we show the required adjustment for more complicated interactions.

2 Treatment interactions and treatreg

Consider an example where selection into the treatment Y_2 is a function of ϵ_2, which is correlated with ϵ_1, the error term in the equation of the outcome Y_1:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \beta_2 Y_2 X_1 + \delta Y_2 + \epsilon_1 \\ Y_2^* &= \gamma_0 + \gamma_1 X_2 + \epsilon_2 \\ Y_2 &= \begin{cases} 1 & \text{if} \quad Y_2^* > 0 \\ 0 & \text{if} \quad Y_2^* \leq 0 \end{cases} \end{aligned} \tag{1}$$

We observe X_1, X_2, Y_1, and Y_2; $E(\epsilon_i) = 0$; $\text{Var}(\epsilon_i) = \sigma_i^2$ for $i = 1, 2$; and we assume that $\sigma_2^2 = 1$. Assuming that ϵ_1 and ϵ_2 follow a bivariate normal distribution with correlation ρ, the parameters β_0, β_1, β_2, δ, γ_0, γ_1, σ_1, and ρ can be consistently estimated using either the maximum likelihood (ML) or the two-stage estimation procedure of `treatreg`. Using `treatreg` to fit models similar to (1) but with $\beta_2 = 0$ was first discussed in Cong and Drukker (2000). When $\beta_2 \neq 0$, we have an additional endogenous variable, but this does not change the underlying random structure of the model; the identification conditions remain the same as when $\beta_2 = 0$ (Wooldridge 2010, 265–266).

For the purpose of estimating the above parameters, it is irrelevant whether `treatreg` recognizes the term $\beta_2 Y_2 X_1$ as an interaction term between the treatment and an exogenous variable. What matters is that the likelihood function (for ML estimation) and the estimating equations (for two-stage estimation) are correctly specified and therefore the estimates are consistent.

However, when estimating the ATE, results computed with `treatreg` postestimation must be corrected. In the context of (1), the ATE is given by $E(Y_1|Y_2 = 1) - E(Y_1|Y_2 = 0)$ (Wooldridge 2010, 905). `treatreg` postestimation provides the command `predict` *newvar*`, yctrt` to estimate $E(Y_1|X_1, X_2, Y_2 = 1)$ and `predict` *newvar*`, ycntrt` to estimate $E(Y_1|X_1, X_2, Y_2 = 0)$. These estimated conditional expectations are then averaged across the sample and differenced to obtain an estimate of the ATE. This is appropriate when there is no treatment interaction term. When a treatment interaction term is present, however, the `predict` commands do not condition the treatment interaction term according to the conditioning value of the treatment; the sample value of the treatment is used instead.

It is instructive for what follows to derive the deviation between the two processes in the context of (1). In the population, the conditional expectations of the outcome are given by[1]

$$E(Y_1|X_1, X_2, Y_2 = 1) = \beta_0 + (\beta_1 + \beta_2)X_1 + \delta + \sigma_1 \rho \frac{\phi(\gamma_0 + \gamma_1 X_2)}{\Phi(\gamma_0 + \gamma_1 X_2)} \tag{2}$$

$$E(Y_1|X_1, X_2, Y_2 = 0) = \beta_0 + \beta_1 X_1 - \sigma_1 \rho \frac{\phi(\gamma_0 + \gamma_1 X_2)}{1 - \Phi(\gamma_0 + \gamma_1 X_2)} \tag{3}$$

1. Equations (2) and (3) follow the formula for deriving moments of the incidentally truncated bivariate normal distribution (see theorem 19.5 in Greene [2012, 873]).

where $\phi(\cdot)$ is the standard normal density and $\Phi(\cdot)$ is the standard normal cumulative distribution function. The effect of the treatment on one observation is just their difference:

$$\begin{aligned} &E\left(Y_1|X_1, X_2, Y_2 = 1\right) - E\left(Y_1|X_1, X_2, Y_2 = 0\right) = \\ &\beta_2 X_1 + \delta + \sigma_1\rho\frac{\phi(\gamma_0 + \gamma_1 X_2)}{\Phi(\gamma_0 + \gamma_1 X_2)\left\{1 - \Phi(\gamma_0 + \gamma_1 X_2)\right\}} \end{aligned} \tag{4}$$

The ATE, the treatment effect across the whole population, is then

$$\begin{aligned} &E\left(Y_1|Y_2 = 1\right) - E\left(Y_1|Y_2 = 0\right) = \\ &\beta_2 E(X_1) + \delta + \sigma_1\rho E\left[\frac{\phi(\gamma_0 + \gamma_1 X_2)}{\Phi(\gamma_0 + \gamma_1 X_2)\{1 - \Phi(\gamma_0 + \gamma_1 X_2)\}}\right] \end{aligned} \tag{5}$$

where (5) follows from (4) by the law of iterated expectations (Wooldridge 2010, 31), and where the expectations on the right-hand side (RHS) are over X_1 and X_2, respectively. An estimator of the ATE can then take the general form

$$\widehat{\beta}_2\overline{X}_1 + \widehat{\delta} + \widehat{\sigma}_1\widehat{\rho}\overline{\left[\frac{\phi(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2)}{\Phi(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2)\{1 - \Phi(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2)\}}\right]} \tag{6}$$

where the expectation terms of the RHS of (5) are estimated by the corresponding sample means, and the parameters β_2, δ, σ_1, ρ, γ_0, and γ_1 can be estimated using ML or another method.

In general, the ATE can be estimated using the `treatreg predict` commands, but using the `predict` command is not straightforward when interaction terms are present. With one treatment interaction term, the predicted value of one observation conditional on the observed variables and on being treated is given in (2). Because the `predict` commands in the interaction term use the sample value instead of the conditioning value of the treatment, using the `predict` command the RHS of (2) is estimated as

$$\widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 Y_2 X_1 + \widehat{\delta} + \widehat{\sigma}_1\widehat{\rho}\frac{\phi(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2)}{\Phi(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2)}$$

instead of

$$\widehat{\beta}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2\right) X_1 + \widehat{\delta} + \widehat{\sigma}_1\widehat{\rho}\frac{\phi\left(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2\right)}{\Phi\left(\widehat{\gamma}_0 + \widehat{\gamma}_1 X_2\right)}$$

The difference between the two is

$$\widehat{\beta}_2 X_1 - \widehat{\beta}_2 Y_2 X_1$$

Averaging across the sample, we have

$$\begin{aligned}
\widehat{\beta}_2\overline{X}_1 - \widehat{\beta}_2\overline{Y_2X_1} &= \widehat{\beta}_2\frac{1}{N}\left(\sum_i X_{1i} - \sum_i Y_{2i}X_{1i}\right) \\
&= \widehat{\beta}_2\frac{1}{N}\left(\sum_{i:\ Y_2=1} X_{1i} + \sum_{i:\ Y_2=0} X_{1i} - \sum_{i:\ Y_2=1} Y_{2i}X_{1i} - \sum_{i:\ Y_2=0} Y_{2i}X_{1i}\right) \\
&= \widehat{\beta}_2\frac{1}{N}\left(\sum_{i:\ Y_2=1} X_{1i} + \sum_{i:\ Y_2=0} X_{1i} - \sum_{i:\ Y_2=1} X_{1i}\right) \\
&= \widehat{\beta}_2\frac{1}{N}\sum_{i:\ Y_2=0} X_{1i} \qquad (7)
\end{aligned}$$

Similarly, the predicted value of one observation conditional on the observed variables and on not being treated is given in (3). As before, using the `treatreg predict` command, the RHS of (3) is estimated as

$$\widehat{\beta}_0 + \widehat{\beta}_1X_1 + \widehat{\beta}_2Y_2X_1 - \widehat{\sigma}_1\widehat{\rho}\frac{\phi\left(\widehat{\gamma}_0 + \widehat{\gamma}_1X_2\right)}{\Phi\left(\widehat{\gamma}_0 + \widehat{\gamma}_1X_2\right)}$$

instead of

$$\widehat{\beta}_0 + \widehat{\beta}_1X_1 - \widehat{\sigma}_1\widehat{\rho}\frac{\phi(\widehat{\gamma}_0 + \widehat{\gamma}_1X_2)}{\Phi(\widehat{\gamma}_0 + \widehat{\gamma}_1X_2)}$$

The difference between the two is now

$$-\widehat{\beta}_2Y_2X_1$$

Averaging across the sample gives

$$\begin{aligned}
-\widehat{\beta}_2\overline{Y_2X_1} &= -\widehat{\beta}_2\frac{1}{N}\sum_i Y_{2i}X_{1i} \\
&= -\widehat{\beta}_2\frac{1}{N}\left(\sum_{i:\ Y_2=1} Y_{2i}X_{1i} + \sum_{i:\ Y_2=0} Y_{2i}X_{1i}\right) \\
&= -\widehat{\beta}_2\frac{1}{N}\sum_{i:\ Y_2=1} X_{1i} \qquad (8)
\end{aligned}$$

Subtracting (8) from (7) gives the difference between the estimator in (6) and the quantity computed on the basis of the `predict` commands:

$$\begin{aligned}\widehat{\beta}_2 \frac{1}{N} \sum_{i:\ Y_2=0} X_{1i} + \widehat{\beta}_2 \frac{1}{N} \sum_{i:\ Y_2=1} X_{1i} &= \widehat{\beta}_2 \frac{1}{N} \left(\sum_{i:\ Y_2=0} X_{1i} + \sum_{i:\ Y_2=1} X_{1i} \right) \\ &= \widehat{\beta}_2 \frac{1}{N} \sum_{i} X_{1i} \\ &= \widehat{\beta}_2 \overline{X}_1 \end{aligned} \tag{9}$$

This is the first term of the estimator in (6). Therefore the quantity computed with the predict commands is the sum of the remaining two terms in (6), namely, the coefficient of the treatment indicator $\widehat{\delta}$ and the selection bias term. From (9), we conclude that the estimator in (6) will be greater than the quantity computed with the `predict` commands if the sample mean of the interaction variable and the interaction coefficient have the same sign; otherwise, the estimator will be smaller.

It is straightforward to extend this result to contexts of treatment interactions with more independent variables. For a treatment interaction of the general form $f(X_1, Y_2)$, where $f(\cdot)$ is any function, the adjustment term corresponding to (9) is

$$\widehat{\beta}_2 \left\{ \overline{f(X_1, Y_2 = 1)} - \overline{f(X_1, Y_2 = 0)} \right\} \tag{10}$$

The results of this section can be extended easily to several interactions.

3 The itreatreg command

The `itreatreg` command can be used when multiplicative treatment interactions enter the outcome equation in a model such as (1). In a model with nonexperimental data and selection on the basis of unobservables, as in (1), multiplicative treatment interactions are interactions of the form $Y_2 f(X_1)$, where $f(\cdot)$ can be any function of X_1.

The `itreatreg` and `treatreg` commands produce the same parameter estimates of the model. In addition to these estimates, `itreatreg` uses the adjustment described in the previous section to evaluate the estimator in (6). The computational heart of the commands calls `treatreg` internally, and the adjustments are made from the estimates provided by `treatreg` and stored in two new variables. `itreatreg` also displays and returns the adjusted ATE and the standard deviation of the treatment effect.

3.1 Syntax

The syntax of the `itreatreg` command is

`itreatreg` *depvar* [*indepvars_ni*] [*if*] [*in*] ,
`treat(`*depvar_t* `=` *indepvars_t* [`,` `noconstant`]`)` `x(`*xvars* [`=` *indepvars_i*]`)`
`gen(`*stubname*`)` [`oos twostep`]

where *depvar* is the dependent variable of interest in the outcome equation. *indepvars_ni* is the list of predictors in the outcome equation that are not interacted with the treatment variable. This is optional in so far as predictor variables that are interacted with the treatment variable are specified in the `x()` option, so if all the predictor variables are included with interaction terms, then this list will be empty.

3.2 Options

`treat(`*depvar_t* `=` *indepvars_t* [`, noconstant`]`)` specifies the equation for the treatment selection, where *depvar_t* is the treatment variable and *indepvars_t* is the list of predictor variables for the treatment, in a manner identical to the specification in the `treatreg` command. It is an integral part of specifying a treatment-effects model and is required. The `noconstant` option suppresses the constant in the treatment equation.

`x(`*xvars* [`=` *indepvars_i*]`)` specifies the treatment interaction variables *xvars* and, optionally, the original variables *indepvars_i* that were interacted with the treatment. `x()` is required. The inclusion of *indepvars_i* is optional if one wishes to include only the interaction term. At least one *xvar* must be specified, otherwise `treatreg` itself is appropriate. Moreover, if the original variables are included, then they must be specified correctly in `x()` and not included in the list of independent variables *indepvars_ni* directly after the dependent variable. For example, `itreatreg y1, treat(y2=x1) x(y2x2) gen(pr)` would fit a simple model in which an interaction between the treatment variable `y2` and an independent variable `x2`—that is, `y2x2`—is the sole predictor of `y1`, aside from the treatment variable itself. Inclusion of the original independent variable `x2` in the model must be specified: `itreatreg y1 x2, treat(y2=x1) x(y2x2=x2) gen(pr)`.

`gen(`*stubname*`)` is required and specifies the *stubname* for the two new variables created by `itreatreg`, *stubname*`ctrt` and *stubname*`cntrt`, that contain for each observation, respectively, the predicted value of the dependent variable *depvar* in the presence of the treatment and the predicted value in the absence of the treatment. This is analogous to the `predict` *varname*`, yctrt` and `predict` *varname*`, ycntrt` postestimation commands for `treatreg`, but it is corrected for the effect of the interaction variables. Contrary to the usual `predict` syntax, the default in `itreatreg` is to create predictions only for those observations used in the estimation process. Applying the predictions to the entire dataset requires specification of the `oos` option

(see below). If the variable names created by this process are unavailable (for example, if one specifies `gen(pr)` when there already exists a variable named `prctrt`), then `itreatreg` will still produce the estimated coefficients but will not calculate the predicted values or the ATE.

`oos` specifies that the predicted values generated by `treatreg`—and hence the calculation of the ATE—are applied to all observations in the dataset. By default, the predictions are otherwise applied only to those observations included in the estimation of the coefficients.

`twostep` specifies that two-step consistent estimates of the parameters, standard errors, and covariance matrix of the model be produced instead of the default ML estimates.

3.3 Saved results

Though `itreatreg` provides estimation of coefficients, it does so by calling the `treatreg` function internally. `itreatreg` is primarily a postestimation command that creates adjusted predictions for interaction terms. Therefore, normal Stata postestimation commands such as `predict` run after `itreatreg`, they will act on the estimations provided by `treatreg`, and they will not take into account the adjustments for interaction made by `itreatreg`. In addition to the results returned by the `treatreg` command called internally, `itreatreg` returns the following additional results:

Scalars

`r(ate)`	ATE
`r(te_sd)`	standard deviation of the treatment effect
`r(N_ate)`	number of observations used to generate ATE

Macros

`r(varctrt)`	name of new variable containing predicted values in the presence of treatment
`r(varcntrt)`	name of new variable containing predicted values in the absence of treatment

4 Examples

4.1 Multiplicative interactions using itreatreg

This example uses the same data that Cong and Drukker (2000) used in their discussion of the `treatreg` command. It is the same data used in the StataCorp (2011) discussion of the `treatreg` command. The `treatreg` command is used with a dataset of women's wages and other characteristics to explore the possibility that women's college education is endogenous to wage determination (the hypothesis was rejected). The original model is modified to allow for multiplicative interactions between the treatment (college education) with the two exogenous variables in the wage equation, age, and whether they live in a large city.

```
. use http://www.stata-press.com/data/r12/labor
. generate wc = 0
. replace wc = 1 if we > 12
(69 real changes made)
. generate wcXwa = wc*wa
. generate wcXcit = wc*cit
. itreatreg ww, treat(wc=wmed wfed) x(wcXwa=wa wcXcit=cit) gen(padjusted)
Iteration 0:   log likelihood = -706.19914
Iteration 1:   log likelihood = -706.19738
Iteration 2:   log likelihood = -706.19738
Treatment-effects model -- MLE                  Number of obs   =        250
                                                Wald chi2(5)    =       5.91
Log likelihood = -706.19738                     Prob > chi2     =     0.3148
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ww						
wa	.0057609	.0236009	0.24	0.807	-.040496	.0520178
cit	.0720367	.3829244	0.19	0.851	-.6784814	.8225548
wcXwa	-.0542976	.0410126	-1.32	0.186	-.1346807	.0260855
wcXcit	.0980451	.8044176	0.12	0.903	-1.478584	1.674675
wc	3.466534	1.900961	1.82	0.068	-.2592815	7.192349
_cons	1.657002	1.059636	1.56	0.118	-.4198465	3.73385
wc						
wmed	.1197113	.032011	3.74	0.000	.056971	.1824517
wfed	.0964197	.0291015	3.31	0.001	.0393819	.1534576
_cons	-2.633536	.3310894	-7.95	0.000	-3.282459	-1.984613
/athrho	.0435995	.1904776	0.23	0.819	-.3297297	.4169287
/lnsigma	.9210499	.0448669	20.53	0.000	.8331123	1.008988
rho	.0435719	.190116			-.3182779	.3943399
sigma	2.511926	.1127025			2.300467	2.742823
lambda	.1094494	.4779808			-.8273757	1.046274

```
LR test of indep. eqns. (rho = 0):   chi2(1) =     0.05   Prob > chi2 = 0.8191
Average Treatment Effect (ATE)  = 1.3945965
Standard deviation of Treatment Effect = .44730832
. predict poriginalctrt, yctrt
. predict poriginalcntrt, ycntrt
. generate poriginaldiff = poriginalctrt - poriginalcntrt
. summarize poriginaldiff
```

Variable	Obs	Mean	Std. Dev.	Min	Max
poriginald~f	250	3.663869	.0268047	3.641228	3.790205

This example first generates the necessary interaction terms that are not present in the original dataset and then calls `itreatreg` to estimate the parameters, generate predicted values, and calculate the ATE. After calling `itreatreg`, the example then recalculates the ATE and the standard deviation of the treatment effect on the basis of the unadjusted predicted values generated by the `treatreg` function. The unadjusted ATE is reported as the mean of the `poriginaldiff` variable in the summary table; the standard deviation of the treatment effect is the standard deviation of `poriginaldiff`. Clearly, there is a significant difference in the estimated treatment statistics. The ATE is almost three times higher in the unadjusted calculations than the correct ATE, while the standard deviation of the treatment effect is much smaller.[2]

4.2 Nonmultiplicative interactions

Nonmultiplicative treatment interactions are rarely used. Here we modify the previous example to include a nonmultiplicative interaction between age and the treatment, in addition to the multiplicative interaction between the treatment and living in a large city.

```
. use http://www.stata-press.com/data/r12/labor, clear
. generate wc = 0
. replace wc = 1 if we > 12
(69 real changes made)
. generate wcxcit = wc*cit
. generate wc_wa = 1/(wa^wc)
```

2. As noted in section 2 in the context of one interaction term, the ATE is the sum of the interaction effect ($\widehat{\beta}_2\overline{X}_1$) and the unadjusted ATE. When the interaction effect and the unadjusted ATE have similar magnitudes but opposite signs, the unadjusted ATE will be substantially larger than the ATE (in absolute value). That is easy to verify in this example. Note, however, that in this example, there are two interaction terms and therefore the interaction effect is the sum of the products of the interaction coefficients with the sample means of the corresponding interaction variables.

```
. treatreg ww wa cit wc_wa wcxcit, treat(wc=wmed wfed)
Iteration 0:   log likelihood = -706.17482
Iteration 1:   log likelihood = -706.17325
Iteration 2:   log likelihood = -706.17325
Treatment-effects model -- MLE                  Number of obs   =        250
                                                Wald chi2(5)    =       5.97
Log likelihood = -706.17325                     Prob > chi2     =     0.3094
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ww           |
          wa |    .005609   .0234476     0.24   0.811    -.0403474    .0515654
         cit |   .0724072   .3828214     0.19   0.850    -.6779089    .8227233
       wc_wa |   94.45258   70.37642     1.34   0.180    -43.48267    232.3878
      wcxcit |   .0996757   .8043637     0.12   0.901    -1.476848    1.676199
          wc |   93.29493   68.59617     1.36   0.174    -41.15108    227.7409
       _cons |  -92.79207   70.87472    -1.31   0.190     -231.704    46.11982
-------------+----------------------------------------------------------------
wc           |
        wmed |   .1196905   .0320164     3.74   0.000     .0569394    .1824415
        wfed |   .0964198   .0291069     3.31   0.001     .0393713    .1534683
       _cons |  -2.633293   .3310698    -7.95   0.000    -3.282178   -1.984408
-------------+----------------------------------------------------------------
     /athrho |   .0406697   .1900651     0.21   0.831     -.331851    .4131905
   /lnsigma  |   .9208927   .0448475    20.53   0.000     .8329931    1.008792
-------------+----------------------------------------------------------------
         rho |   .0406473   .1897511                      -.320183    .3911783
       sigma |   2.511531    .112636                      2.300193    2.742287
      lambda |    .102087   .4769336                     -.8326856     1.03686
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =     0.05   Prob > chi2 = 0.8307
. predict wwhat1, yctrt
. predict wwhat0, ycntrt
. generate wwctrt = wwhat1 + (1- wc)*([ww]_b[wc_wa]*(1/wa-1) + [ww]_b[wcxcit]*cit)
. generate wwcntrt = wwhat0 + wc*([ww]_b[wc_wa]*(1-1/wa) - [ww]_b[wcxcit]*cit)
. generate wwatehat = wwctrt - wwcntrt
. generate wwhatdiff = wwhat1 - wwhat0
. summarize wwhat1 wwctrt wwhat0 wwcntrt wwatehat wwhatdiff, sep(0)
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     wwhat1  |       250    69.95679    41.26938    2.468673   95.62572
     wwctrt  |       250    3.277062    .4031218    2.468673   4.157271
     wwhat0  |       250    -23.5222    41.26322   -90.99352   2.051853
    wwcntrt  |       250    1.898402    .0746325    1.687243   2.058028
   wwatehat  |       250     1.37866    .4548583     .5838184   2.281107
  wwhatdiff  |       250    93.47899     .024998    93.45787   93.59681
```

The mean of the variable `wwhatdiff` is the estimate of the ATE produced on the basis of the predict commands without any adjustments. The mean of `wwatehat` is the estimate produced by computing the correct conditional expectations using the adjustments of equations (7) and (8) and following the generalization of (10). The model in this example has the nonmultiplicative interaction term $X_2^{-Y_2}$, but the results are similar to the previous model with the multiplicative interaction term. However,

the absolute values of the estimated coefficients of age and its interaction term, and the constant of the outcome equation are much larger. The estimated ATE is still the same as in the previous example to the first decimal, but the estimated ATE without the necessary adjustment—`wwhatdiff`—is very different.

5 Conclusion

The Stata `treatreg` command can be used to fit models where selection into treatment depends on observed and nonobserved factors. The `treatreg` command gives consistent estimates of the parameters whether treatment interactions are included or not. The `predict` command of `treatreg` postestimation, however, gives the correct conditional predictions only when treatment interactions are not present. In this article, we derived the adjustments that are required to compute the correct conditional predictions and ATE. When the treatment interactions are multiplicative, we introduced the `itreatreg` command, which produces the appropriate estimate of the ATE in addition to the usual output of the `treatreg` command. When treatment interactions are nonmultiplicative, we showed the steps that are required to produce the appropriate estimates of the ATE.

6 References

Cong, R., and D. M. Drukker. 2000. sg141: Treatment effects model. *Stata Technical Bulletin* 55: 25–33. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 159–169. College Station, TX: Stata Press.

Greene, W. H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Royston, P., and W. Sauerbrei. 2008. Interactions between treatment and continuous covariates: A step toward individualizing therapy. *Journal of Clinical Oncology* 26: 1397–1399.

———. 2009. Two techniques for investigating interactions between treatment and continuous covariates in clinical trials. *Stata Journal* 9: 230–251.

Sauerbrei, W., P. Royston, and K. Zapien. 2007. Detecting an interaction between treatment and a continuous covariate: A comparison of two approaches. *Computational Statistics and Data Analysis* 51: 4054–4063.

StataCorp. 2011. *Stata 12 Base Reference Manual*. College Station, TX: Stata Press.

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

About the authors

Graham K. Brown is the director of the Centre for Development Studies and a senior lecturer in international development in the Department of Social and Policy Sciences at the University of Bath.

Thanos Mergoupis is a lecturer in the Department of Economics at the University of Bath.