Title

Intro 5 — Treatment assignment features

Description

ERMs can fit treatment-effect models. Treatment can be binary (not treated or treated) or ordinal (not treated or treated or treated extremely).

Option `extreat()` specifies exogenous treatment effects.

Option `entreat()` specifies endogenous treatment effects.

ERM's treatment-effect features are explained below.

Remarks and examples

Remarks are presented under the following headings:

What are treatment-effect models?

Let's consider a simple binary treatment-effect problem. A treatment is applied to some patients, and we want to measure its effect. We start by imagining that patients are assigned randomly to the treated group. We observe a continuous outcome `y`, such as blood pressure, and we think the treatment affects `y`. We think the treatment's effect varies with patients' age, `x1`.

Here is one way we could fit the model:

```
. eregress y x1 i.treated i.treated#c.x1
```

Variable `treated` specifies which patients were treated. It contains 1 or 0. The model we just fit is

$$\mathtt{y}_i = \beta_0 + \beta_1\mathtt{x1}_i + \beta_2\mathtt{treated}_i + \beta_3\mathtt{treated}_i\mathtt{x1}_i + e_i.\mathtt{y}$$

This model says that the outcome for patients who are not treated is

$$\mathtt{y}_i = \beta_0 + \beta_1\mathtt{x1}_i + e_i.\mathtt{y} \tag{1}$$

For those treated, the outcome is

$$\mathtt{y}_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)\mathtt{x1}_i + e_i.\mathtt{y} \tag{2}$$

β_2 and β_3x1 measure the effect of being treated. That effect varies observation by observation with the patient's age. Many researchers would stop here, satisfied to have the fitted coefficients.

Researchers who fit treatment-effect models, however, usually want to know the average treatment effect (ATE). We could obtain it.

We would calculate the average outcome when being treated over the entire dataset, the average outcome when not being treated over the entire dataset, and subtract the two results to obtain an ATE for this group of patients.

In an equivalent method of obtaining the ATE, we would calculate new variable `if_not_treated` equal to

$$\texttt{if\_not\_treated}_i = \widehat{\beta}_0 + \widehat{\beta}_1\texttt{x1}_i$$

We would calculate new variable `if_treated` equal to

$$\texttt{if\_treated}_i = \texttt{if\_not\_treated}_i + \widehat{\beta}_2 + \widehat{\beta}_3\texttt{x1}_i$$

Then, we would subtract them:

$$\texttt{diff}_i = \texttt{if\_treated}_i - \texttt{if\_not\_treated}_i$$

We would finally calculate the mean of `diff`:

```
. summarize diff
```

Stata's `margins` command is wonderful at doing things like this, and it even reports the standard error! And we do not even have to calculate `if_not_treated`, `if_treated`, and `diff`. Instead, we just type

```
. margins r.treated
```

Treatment-effect models and potential outcomes

We can write the equations for the untreated and the treated, (1) and (2), in another way. For the untreated, the first potential outcome, we write

$$\mathrm{y}_{0i} = \gamma_{00} + \gamma_{01}\texttt{x1} + e_i.\mathrm{y}_0 \tag{3}$$

For the treated, the second potential outcome, we write

$$\mathrm{y}_{1i} = \gamma_{10} + \gamma_{11}\texttt{x1} + e_i.\mathrm{y}_1 \tag{4}$$

We can also write the ATE in terms of these potential outcomes as

$$\text{ATE} = E\left(\mathrm{y}_{1i} - \mathrm{y}_{0i}\right)$$

Instead of using the `eregress` and `margins` commands above, we can type

```
. eregress y x1, extreat(treated)
```

The coefficients reported are now the γ coefficients from (3) and (4). Because we fit the model using the `extreat()` option, we can now type the following to estimate the ATE:

```
. estat teffects
```

Endogenous and exogenous treatment effects

ERMs can fit models far more complicated than the model we just fit. That is good because the story we told above is too simple. For instance, we said that patients were assigned randomly. Ethical considerations sometimes prevent that.

There are two types of treatment effects: exogenous and endogenous. What distinguishes them is the same thing that distinguished them in [ERM] **Intro 4**, where we discussed exogenous and endogenous selection effects. It matters whether the error in the selection (treatment assignment) equation is correlated with the error in the main equation.

Here are four examples of treatment assignment.

1. Assignment is random (as above). In this case, the assignment equation contains only an error, e.`treated`, and it is uncorrelated with e.y. Treatment is exogenous.
2. Assignment is determined by hard-and-fast rules. There is no e.`treated`, or if you prefer, it is 0. Either way, it is uncorrelated with e.y. Treatment is exogenous.
3. Assignment is determined in part by hard-and-fast rules, but if the patient meets them, a coin is flipped to determine whether the patient is treated or untreated. e.`treated` is the coin flip, and it is uncorrelated with e.y. Treatment is exogenous.
4. Assignment is by whatever rules, if any, plus unobserved judgment. Thus, judgment appears in e.`treated`, and we must consider the possibility that it is correlated with e.y. Treatment is endogenous.

ERMs can fit models with exogenous or endogenous treatment assignment. You specify the `extreat()` or `entreat()` option. In the four examples above, you would specify

1. `extreat(treated)`
2. `extreat(treated)`
3. `extreat(treated)`
4. `entreat(treated = ...)`

You could fit

```
. eregress y x1 x2 x3, extreat(treated)
```

or

```
. eregress y x1 x2 x3, entreat(treated = x1 z1 z2)
```

These models estimate distinct intercepts and distinct coefficients on `x1`, `x2`, and `x3` in the equation for `y`. They also estimate the variance of e.`y`, but it is assumed to be equal across treatment groups. In the potential-outcomes framework, this means that the variance of $e_i.\mathrm{y}_0$ and that of $e_i.\mathrm{y}_1$ are assumed to be equal. This may not be reasonable. Perhaps the variance of the error for the treated group is larger than the variance of the error for the untreated group. The `povariance` suboption relaxes this constraint so that distinct error variances are estimated for each potential outcome. We can type

```
. eregress y x1 x2 x3, extreat(treated, povariance)
```

and

```
. eregress y x1 x2 x3, entreat(treated = x1 z1 z2, povariance)
```

We might want to allow the correlation between error terms to differ across potential outcomes. The `pocorrelation` suboption specifies that distinct correlations are estimated. For this, we type

```
. eregress y x1 x2 x3, entreat(treated = x1 z1 z2, pocorrelation)
```

More likely, we would want to let both error variances and correlations vary across potential outcomes by typing

```
. eregress y x1 x2 x3, entreat(treated = x1 z1 z2, povariance pocorrelation)
```

Whichever you type, you can obtain the ATE by typing

```
. estat teffects
```

Binary and ordinal treatment effects

We have been assuming that treatment is binary. ERMs can also fit ordinal treatment models. Think of these models as all being the same treatment but of different intensities. For instance,

1. A rehabilitative exercise program might be attended not at all, once a week, or twice a week.
2. A drug might be administered in different dosages.
3. A jobs program might be attended not at all, once a week, or twice a week.
4. The amount of post-secondary education could be none, some college, graduated, or graduated plus postgraduate.

When treatment is ordinal, variable `treated` contains more than two values. The variable might contain 0, 1, or 2; or 1, 2, or 3; or even 2, 3, or 5. If there are four ordered treatments, the variable contains four different values. The particular values do not matter as long as the numeric values order the treatments in the way they should be ordered.

When the treated variable takes on more than two values, `entreat()` fits the endogenous treatment equation by using ordered probit instead of binary probit.

Sample versus population standard errors

Researchers who fit treatment models usually want population standard errors.

When we fit the treatment model by hand, not using the `extreat()` or `entreat()` option, we typed

```
. eregress y x1 i.treated i.treated#c.x1
. margins r.treated
```

When we used the `extreat()` option to fit the same model, we used `estat teffects` to obtain the ATE. We typed

```
. eregress y x1, extreat(treated)
. estat teffects
```

In both cases, the standard errors reported for the ATE were the same. The data were treated as fixed and not as a draw from the underlying population.

The standard error would also be treated that way if we fit a model with endogenous instead of exogenous treatment assignment.

```
. eregress y x1 x2 x3, entreat(treated = x1 z1 z2)
. estat teffects
```

Researchers fitting treatment-effect models often want standard errors for ATEs suitable for predicting to the entire population and not just this particular sample. If you want population-based standard errors, you must fit the model by using the `vce(robust)` option:

```
. eregress y x1, extreat(treated) vce(robust)
. estat teffects
```

Do that and `estat teffects` will report population-based standard errors.

Returning to the `eregress` command, when you do not specify `vce(robust)`, it reports OIM standard errors. OIM stands for observed information matrix. The alternative `robust` standard errors assume less and are therefore less efficient. While less efficient, robust standard errors still have correct coverage. The standard errors themselves just have more sampling variability. Robust standard errors are absolutely required if `estat teffects` is to report standard errors for the effect in the population.

Requesting the ATE with population standard errors makes sense only if the sample you are using is an unbiased random draw from the population for which you wish to make predictions. If the sample is not, you need to specify your data's probability sampling weights as well. Type

```
. eregress y x1 [pw = weight], extreat(treated) vce(robust)
. estat teffects
```

In this case, you can omit the `vce(robust)` option because it is assumed when probability sampling weights are specified.

Variable `weight` contains inverse probabilities that the observations were sampled from the population. For instance, if some observations were sampled with probability 0.001 and others with 0.0001, then `weight` contains 1,000 and 10,000. For our purposes here, the scale of weights does not matter, so `weight` could just as well contain 1 and 10. Scale of weights matters when you request totals, which `estat teffects` does not produce.

Using treatment effects with other ERMs

The outcome variable `y` need not be continuous. It can be interval, binary, or ordinal, meaning that you can use the `eintreg`, `eprobit`, or `eoprobit` command to fit the model.

If we had a binary outcome variable, we would type

```
. eprobit y x1 x2 x3, entreat(treated = x1 z1 z2)
```

If we planned on obtaining the ATE with population standard error, we would type

```
. eprobit y x1 x2 x3, entreat(treated = x1 z1 z2) vce(robust)
. estat teffects
```

Using treatment effects with other features of ERMs

`extreat()` and `entreat()` can be used with `endogenous()` and `select()`. Said differently, treatment models can contain endogenous covariates and be adjusted to handle endogenous sample selection. Treatment models can also include random effects to account for within-panel or within-group correlation; we will discuss this in [ERM] **Intro 6**. Here we will focus on combining treatment with endogenous covariates and sample selection.

By now, you are familiar with the `endogenous()` option. Some examples of `eregress` used with `extreat()` and `entreat()` are

```
. eregress y x1 x2 w1, extreat(treated) endogenous(w1 = x1 z1 z2, nomain)
. eregress y x1 x2 w1, entreat(treated = z3 w1)  ///
      endogenous(w1 = x1 z1 z2, nomain)
. eregress y x1 x2, entreat(treated = z3 w1) endogenous(w1 = x1 z1 z2, nomain)
```

We used the `nomain` suboption and explicitly included `w1` in the main equation if we wanted it there. In those cases, we could have omitted the explicit mention and deleted option `nomain`. Equivalent to the first example is

```
. eregress y x1 x2, extreat(treated) endogenous(w1 = x1 z1 z2)
```

So far, we have not included the `povariance` and `pocorrelation` suboptions. We can add these to estimate potential-outcome specific variances of *e*.y and potential-outcome specific correlations between *e*.y and *e*.`w1` and between *e*.y and *e*.`treated`. For instance, we could extend the second example to include variances and correlations that vary across treatment groups by typing

```
. eregress y x1 x2 w1, ///
      entreat(treated = z3 w1, povariance pocorrelation) ///
      endogenous(w1 = x1 z1 z2, nomain)
```

Next, we consider use of `entreat()` and `extreat()` with `select()` to account for endogenous and exogenous sample selection.

We wish to fit a treatment-effect model but there is a problem. The treatment-effect model we want to fit is

```
. eregress y x1 x2, entreat(treated = x1 z3)
```

The problem is that the information on `y` was collected at the end of the study, and some patients never showed up—they dropped out along the way. To fit the desired model with the complication, we type

```
. generate selected = !missing(y)
. eregress y x1 x2, entreat(treated = x1 z3) select(selected = x1 z4 z5)
```

To obtain the ATE, we type

```
. estat teffects
```

The model reported by `eregress` and the ATE reported by `estat teffects` will be adjusted for both the endogenous treatment assignment and the endogenous selection effects. The latter adjusts for the censored observations in which the final outcome `y` was not observed.

Reported were sample statistics. If we had wanted population statistics, we would have typed

```
. eregress y x1 x2, entreat(treated = x1 z3) select(selected = x1 z4 z5)  ///
      vce(robust)
. estat teffects
```

If treatment assignment had been exogenous, we would have specified `extreat(treated)` instead of `entreat(treated = x1 z3)`.

If we wanted to allow the variance of *e*.y to vary across potential outcomes, we could add the `povariance` suboption to the `entreat()` or `extreat()` option. If we wanted to allow the correlations between *e*.y and other error terms in the model to vary across potential outcomes, we could add the `pocorrelation` suboption to `entreat()` or `extreat()`.

Note in the above example that treatment can have one arm as in the story we told or it can have multiple arms. In other words, the treatment can be binary or ordinal. Nothing we typed would need to change.

Using treat() and select() to handle lost to follow-up

The important feature of the above example is that the censored observations were lost to follow-up. By that, we mean that the patients did not report for the final meeting, and thus y was unobserved. Specifying the select() equation allowed the error in selection (that is, the unobserved reasons that subjects showed up or did not show up) to be correlated with the error in the main outcome equation (the error in the benefit of the treatment). It also allowed the error in selection to be correlated with the error in the treatment assignment. Said statistically, all endogeneity issues were handled.

This was all possible because the treatment arm was assigned even for the censored observations. Variable treated was not missing. It contained a treatment-arm value just as it does in all the other observations.

What if that is not true? What if censoring occurred before the treatment arm was assigned? Then we have an issue we need to discuss. First, here is how you determine whether your data have this issue. Type

```
. assert !missing(treat) if selected==0
```

If assert reports that the assertion is false, your data have this issue. You have censored observations for which the treatment arm is unassigned.

ERMs handle this issue differently for exogenous and endogenous treatment assignment. If you are fitting an exogenous treatment model,

```
. eregress y x1 x2, extreat(treated) select(selected = x1 z4 z5)
```

ERMs do not care that the treatment arm is missing. Endogenous selection will be fully handled just as if the treatment arm had been observed. That is, ERMs handle the issue as long as the treatment arm does not appear as an explanatory variable in your selection equation.

It would not be unreasonable to fit a model such as

```
. eregress y x1 x2, extreat(treated) select(selected = treated x1 z4 z5)
```

If you are fitting this model, it should be obvious that the treatment arm must be assigned to the censored observations. Your selection equation says that the treatment arm itself will affect whether observations are censored.

Let's put that case aside and return to the usual case of missing treatment with exogenous treatment assignment. The ERM commands fit the model without problem. estat teffects will report the ATE. estat teffects has options for reporting other statistics, all of which will be fine except ATET—the average treatment effect among the treated. ATET is defined to include all treated observations. Because treated is sometimes missing because of selection, the computed ATET will exclude those observations for which treatment assignment is missing.

Now, let's consider endogenous treatment assignment. We want to fit the model

```
. eregress y x1 x2, entreat(treated = x1 z3) select(selected = x1 z4 z5)
```

What makes us hesitate is that some of or all the censored observations have treated equal to missing, meaning that treatment was evidently not assigned for them. If we typed the command and fit the model, ERM would fit it omitting those observations. This is equivalent to assuming that the observations were censored completely at random. That could be reasonable. Perhaps most of the censored observations were lost to follow-up—for them, the treatment arm is observed—and only a few were lost before treatment was assigned because of misplaced paperwork.

On the other hand, if all the censored observations were censored before the treatment arm was assigned, the model cannot be fit. Omitting those with missing treatment omits the censored observations, and there is simply no selection equation left to fit. After dropping the observations containing missing values, everyone left in the estimation sample is not censored.

The bottom line is that ERMs cannot fit this model. ERMs place selection after treatment assignment because lost to follow-up is the common case.

You might be able to salvage the situation. Is it just that the treatment-arm values are not in your dataset because the data were not entered? If so, retrieve the data. If that is not the case but the experiment is still ongoing, run the censored observations through the treatment-assignment process.

Treatment statistics reported by estat teffects

`estat teffects` reports ATEs, which are the average effects of the treatment if it had been applied to the entire sample or to the entire underlying population. `estat teffects` reports its value and standard error. The standard error is for the sample if `vce(robust)` was not specified when the model was fit and for the population if `vce(robust)` was specified when the model was fit.

Sometimes ERMs assume `vce(robust)` even when you do not type it. This happens when you specify features that themselves require `vce(robust)`. Option `vce(cluster)` requires `vce(robust)`. Actually, it is a variation on `vce(robust)`, but that is not important for this problem. If you specify `pweights`, then `vce(robust)` is used too.

The types of standard errors reported will be clearly labeled on the output `eregress`, `eintreg`, `eprobit`, or `eoprobit` produces. `estat teffects` indicates clearly in its output, too, whether output is for the sample or the population.

`estat teffects` reports

> ATE, the average treatment effect for each treatment for the entire sample/population.

`estat teffects, atet` reports

> ATET, the average treatment effect for each treatment for the treated sample or population.

`estat teffects, pomeans` reports

> POMEANS, the potential-outcome means for each treatment arm, meaning means for the untreated and means for each of the treated.
>
> Outcomes here are the values of the dependent variable in the main equation, or `y`.
>
> Potential outcomes are the values, observation by observation, of y_i that would be observed if each was treated and untreated.
>
> Potential-outcome means are the means of treated and, separately, of untreated. The difference between them is the ATE.

All of these statistics can also be reported for subsamples or subpopulations.

Video example

Extended regression models: Nonrandom treatment assignment

Also see

[ERM] **Intro 9** — Conceptual introduction via worked example