

Intro 2 — The models that ERMs fit

[Description](#)[Remarks and examples](#)[Also see](#)

Description

The ERM commands fit linear regressions, interval regressions, probit regressions, and ordered probit regressions. These models are described below.

Remarks and examples

[stata.com](#)

Remarks are presented under the following headings:

Linear regression models
Interval regression models
Probit regression models
Ordered probit regression models

In what follows, the expression

$$\beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \cdots + \beta_k \mathbf{x}_k$$

arises so often that we will write it as

$$\mathbf{x}_i \boldsymbol{\beta}$$

$\mathbf{x}_1, \mathbf{x}_2, \dots$ are variables in your data. They are the explanatory variables—the covariates—of the models that you fit. $\mathbf{x}_{1i}, \mathbf{x}_{2i}, \dots$ are the values of the variables in observation i .

Linear regression models

Linear regression is for use with continuous dependent variables. To fit a linear regression, type

```
. eregress y x1 x2 ... xk
```

The model fit is

$$y_i = \beta_0 + \mathbf{x}_i \boldsymbol{\beta} + e_i$$

where e_i is the error and is assumed to be normally distributed with mean 0 and variance σ^2 .

The fitted parameters are β_0 , $\boldsymbol{\beta}$, and σ^2 .

When you make predictions based on linear regressions, what is predicted is the expected value of y given \mathbf{x} .

Interval regression models

Interval regression is for use with continuous dependent variables. To fit an interval regression, type

```
. eintreg y1 y2 x1 x2 ... xk
```

The model fit is the same as that for linear regression except that y is not a variable in the dataset:

$$y_i = \beta_0 + \mathbf{x}_i\beta + e_i.y$$

The assumptions are the same as for linear regression too. $e_i.y$ is assumed to be normally distributed with mean 0 and variance σ^2 .

The fitted parameters are β_0 , β , and σ^2 .

When you use `eintreg`, rather than specify y , the value of the dependent variable, you specify $y1$ and $y2$, where

$$y1_i \leq y_i \leq y2_i$$

Variables $y1$ and $y2$ specify the interval in which y is known to lie. For instance, if subject 1's blood pressure were not precisely recorded but instead a box was checked reporting that the blood pressure was in the range 110 to 139, then $y1_1$ would equal 110 and $y2_1$ would equal 139.

If $y1_i = y2_i$ in all observations, `eintreg` is the same as linear regression. All values are precisely observed.

If $y1_i = y2_i$ in some observations, those observations are precisely observed.

$y1_i$ may contain a missing value and that means $y1_i = -\infty$. In such observations, all that is known is that $y_i \leq y2_i$. The observation is left-censored. If the box was checked for subject 2's blood pressure being below 120, then $y1_2$ would equal . (missing value) and $y2_2$ would equal 119.

$y2_i$ may contain a missing value and that means $y2_i = +\infty$. In such observations, all that is known is that $y_i \geq y1_i$. The observations are right-censored. If the box was checked that subject 3's blood pressure was above 160, then $y1_3$ would equal 161 and $y2_3$ would equal . (missing value).

If both $y1_i$ and $y2_i$ contain missing values, then all that is known is that $-\infty \leq y_i \leq \infty$, and the observation is ignored when fitting the model.

`eintreg` can be used to fit tobit models. Assume that you have data in which y is left-censored at 0. To fit a tobit model, type

```
. generate y1 = cond(y=0, ., y)
. generate y2 = y
. eintreg y1 y2 x1 x2 ... xk
```

When you make predictions based on interval regressions, `predicted` is the expected value of the dependent variable, the unobserved y , conditioned on the covariates.

Probit regression models

Probit regression is for use with binary dependent variables. To fit a probit regression, type

```
. eprobit y x1 x2 ... xk
```

Variable y in theory should contain the values 0 and 1, but `eprobit` does not require that. It treats all nonzero (and nonmissing) values as if they were 1, which means a positive outcome, such as “subject was hired” or “subject tested positive”. The positive result can be a negative event, such as “subject died”.

The model is

$$p_i = \Pr(\text{positive outcome in obs. } i) = \Pr(\beta_0 + \mathbf{x}_i\beta + e_i.y > 0)$$

where e_i is assumed to be normally distributed with mean 0 and variance 1. With that assumption, the probability of a positive outcome is

$$p_i = \text{normal}(\beta_0 + \mathbf{x}_i\boldsymbol{\beta})$$

The fitted parameters are β_0 and $\boldsymbol{\beta}$.

When you make predictions based on probit regressions, predicted is the probability of a positive outcome conditional on the covariates.

Ordered probit regression models

Ordered probit regression is for use with ordinal dependent variables. To fit an ordered probit regression, type

```
. eoprobit y x1 x2 ... xk
```

Variable y is expected to contain 1, 2, ..., M indicating category number although, just like `oprobit`, `eoprobit` is less demanding. y could contain values 2, 3, 5, and 8 to indicate four ordered categories. What is important is that the categories have a natural ordering and that the numbers used to represent them order the categories in the same way. `eoprobit` could be used with the ordered categories 1) not ambulatory, 2) partially ambulatory, and 3) fully ambulatory. Or the order of the categories could be reversed: 1) fully ambulatory, 2) partially ambulatory, and 3) not ambulatory. Reversing the order reverses the signs of the fitted coefficients but does not substantively change the model.

The model fit is

$$\begin{aligned} p_{m,i} &= \Pr(\text{outcome } m \text{ in obs. } i) \\ &= \Pr(c_{m-1} \leq \mathbf{x}_i\boldsymbol{\beta} + e_i \cdot y \leq c_m) \end{aligned}$$

where e_i is assumed to be normally distributed with mean 0 and variance 1. Thus, the probability that the outcome is m is

$$p_{m,i} = \text{normal}(c_m - \mathbf{x}_i\boldsymbol{\beta}) - \text{normal}(c_{m-1} - \mathbf{x}_i\boldsymbol{\beta})$$

where c_0 and c_M are $-\infty$ and $+\infty$, and c_1, \dots, c_{M-1} are fit from the data. The c values play the role of intercepts and are called cutpoints.

The fitted parameters are $\boldsymbol{\beta}$ and c_1, \dots, c_{M-1} .

When $M = 2$, the ordered probit model reduces to the probit model with $c_0 = -\beta_0$.

When you make predictions based on ordered probit regressions, predicted are the probabilities of the dependent variable equaling each category conditional on the covariates.

Also see

[ERM] `eintreg` — Extended interval regression

[ERM] `eoprobit` — Extended ordered probit regression

[ERM] `oprobit` — Extended probit regression

[ERM] `eregress` — Extended linear regression