# Title

stata.com

**eintreg predict —** predict after eintreg and xteintreg



## Description

In this entry, we show how to create new variables containing observation-by-observation predictions after fitting a model with `eintreg` or `xteintreg`.

## Syntax

You previously fit the model

```
eintreg yl yu x1 ..., ...
```

The equation specified immediately after the `eintreg` command is called the main equation. It is

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

where $\mathtt{yl}_i \leq y_i \leq \mathtt{yu}_i$.

Or perhaps you had panel data and you fit the model with `xteintreg` by typing

```
xteintreg yl yu x1 ..., ...
```

Then the main equation would be

$$\mathtt{y}_{ij} = \beta_0 + \beta_1 \mathtt{x1}_{ij} + \cdots + u_i.\mathtt{y} + v_{ij}.\mathtt{y}$$

where $\mathtt{yl}_{ij} \leq y_{ij} \leq \mathtt{yu}_{ij}$.

In either case, `predict` calculates predictions for `y` in the main equation. The other equations in the model are called auxiliary equations or complications. Our discussion follows the cross-sectional case with a single error term, but it applies to the panel-data case when we collapse the random effects and observation-level error terms, $e_{ij}.\mathtt{y} = u_i.\mathtt{y} + v_{ij}.\mathtt{y}$.

All predictions after `xteintreg` assume the panel-level random effects ($u_i.\mathtt{y}$) are zero. Put another way, predictions condition on random effects being set to their mean.

The syntax of `predict` is

`predict` [*type*] *newvar* [*if*] [*in*] [`,` *statistic* *asfmethod* *counterfactual*]

| *statistic* | Description |
|---|---|
| Main | |
| `mean` | linear prediction; the default |
| `xb` | linear prediction excluding all complications |
| `expmean` | expected value of exponential of the `mean`; $E\{\exp(\texttt{mean})\}$ |
| `ystar(`*a*`,`*b*`)` | $E(y*_j), y*_j = \max\{a, \min(y_j, b)\}$ |

*a* and *b* are numeric values, missing (`.`), or variable names.

| *asfmethod* | Description |
|---|---|
| Main | |
| `asf` | average structural function; the default |
| `fixedasf` | fixed average structural function |
| `noasf` | no average structural function adjustment |

| *counterfactual* | Description |
|---|---|
| Main | |
| `target(`*valspecs*`)` | specify counterfactuals |

*valspecs* specify the values for variables at which predictions are to be evaluated. Each *valspec* is of the form

*varname* `=` *#*

*varname* `=` `(`*exp*`)`

*varname* `=` *othervarname*

For instance, `target(`*valspecs*`)` could be `target(w1=0)` or `target(w1=0 w2=1)`.

Notes:

(1) `predict` can also calculate treatment-effect statistics. See [ERM] **predict treatment**.

(2) `predict` can also make predictions for the other equations in addition to the main-equation predictions discussed here. It can also compute some rarely used statistics. See [ERM] **predict advanced**.

## Options for statistics

Main

`mean`, the default, specifies that the linear prediction be calculated. In each observation, the linear prediction is the expected value of the dependent variable $y$ conditioned on the covariates. Results depend on how complications are handled, which is determined by the *asfmethod* and *counterfactual* options.

`xb` specifies that the linear prediction be calculated ignoring all complications. This prediction corresponds to what would be observed in data in which all the covariates in the main equation were exogenous.

`expmean` calculates the expected value of the exponential of the mean. This is particularly useful when the dependent variable is estimated in the log metric but you want to express results in the natural metric of the dependent variable. `expmean` accounts for integrating over the error when forming the expected value of the exponential of the mean. That expectation is not zero.

As with the nonexponentiated mean, results depend on how complications are handled, which is determined by the *asfmethod* and *counterfactual* options. So, by default, the exponential mean has a structural interpretation because the default `asf` option has computed the average structural function of the exponential mean.

`ystar(`*a*`,` *b*`)` specifies that the linear prediction be censored between *a* and *b*. If *a* is missing (`.`), then `a` is treated as $-\infty$. If *b* is missing (`.`), then *b* is treated as $+\infty$. *a* and *b* can be specified as numeric values, missing (`.`), or variable names.

`ystar(`*a*`,` *b*`)` is often useful when calculating predictions or using `margins` to form inferences because it gives you predictions from the censored distribution rather than treating the censoring as a nuisance and calculating predictions from the uncensored distribution.

To obtain predictions using the censored distribution from your data, you can use the same censoring limits you used when fitting the model. For example, if you typed

```
. eintreg yl yu ...
```

you can then type

```
. predict ystar, ystar(yl, yu)
```

to obtain predictions based on the censored distribution corresponding with your observed data.

## Options for asfmethod

Main

`asf`, `fixedasf`, and `noasf` determine whether and how the average structural function (ASF) of the specified statistic is computed. These options are not allowed with `xb`.

`asf`, the default, calculates the ASF of the statistic. Thus, the default when no *statistic* is specified is the ASF of the linear prediction.

`asf` computes the statistic conditional on the errors of the endogenous variable equations. Put another way, it is the statistic accounting for the correlation of the endogenous covariates with the errors of the outcome equation. Derivatives and contrasts based on `asf` have a structural interpretation. See `margins` for computing derivatives and contrasts.

`fixedasf` calculates a fixed ASF. It is the specified statistic computed using only the coefficients and variables of the outcome equation. `fixedasf` does not condition on the errors of the endogenous variable equations. Contrasts between two fixed counterfactuals averaged over the whole sample have a potential-outcome interpretation. Intuitively, it is as if the values of the covariates were fixed at a value exogenously by fiat. See `margins` for computing derivatives and contrasts.

To be clear, derivatives and contrasts between two fixed counterfactuals using the default `asf` option also have a potential-outcome interpretation. And, unlike `fixedasf`, they retain that interpretation when computed over subpopulations for both linear and nonlinear models.

`noasf` calculates the statistic using the linear prediction with no adjustment. For extended regression models, this is computationally equivalent to `fixedasf`. So `fixedasf` and `noasf` are synonyms.

## Option for counterfactuals

Main

`target(`*valspecs*`)` specifies counterfactual predictions. You specify a list of variables from the main equation and values for them. Those values override the values of the variables calculating $\beta_0 + \beta_1 \texttt{x1}_i + \cdots$. Use of `target()` is discussed in *Remarks and examples* of [ERM] **eregress predict**.

## Remarks and examples

stata.com

Predictions after fitting models with `eintreg` and `xteintreg` are handled the same as they are after fitting models with `eregress` or `xteregress`. The issues are the same. See [ERM] **eregress predict**.

Note that censoring is treated as a nuisance in `eintreg` and `xteintreg` by default. Default predicted values are from the uncensored distribution, not from the censored distribution.

## Methods and formulas

See *Methods and formulas* in [ERM] **eintreg postestimation**.

## Also see

[ERM] **eintreg postestimation** — Postestimation tools for eintreg and xteintreg

[ERM] **eintreg** — Extended interval regression