0.1 `tobit`: Linear Regression for a Left-Censored Dependent Variable

Tobit regression estimates a linear regression model for a left-censored dependent variable, where the dependent variable is censored from below. While the classical tobit model has values censored at 0, you may select another censoring point. For other linear regression models with fully observed dependent variables, see Bayesian regression (Section ??), maximum likelihood normal regression (Section ??), or least squares (Section ??).

Syntax

```
> z.out <- zelig(Y ~ X1 + X2, below = 0, above = Inf,
                 model = "tobit", data = mydata)
> x.out <- setx(z.out)
> s.out <- sim(z.out, x = x.out)
```

Inputs

`zelig()` accepts the following arguments to specify how the dependent variable is censored.

- `below`: (defaults to 0) The point at which the dependent variable is censored from below. If any values in the dependent variable are observed to be less than the censoring point, it is assumed that that particular observation is censored from below at the observed value. (See Section ?? for a Bayesian implementation that supports both left and right censoring.)
- `robust`: defaults to `FALSE`. If `TRUE`, `zelig()` computes robust standard errors based on sandwich estimators (see Huber (1981) and White (1980)) and the options selected in `cluster`.
- `cluster`: if `robust = TRUE`, you may select a variable to define groups of correlated observations. Let `x3` be a variable that consists of either discrete numeric values, character strings, or factors that define strata. Then

  ```
  > z.out <- zelig(y ~ x1 + x2, robust = TRUE, cluster = "x3",
                   model = "tobit", data = mydata)
  ```

  means that the observations can be correlated within the strata defined by the variable `x3`, and that robust standard errors should be calculated according to those clusters. If `robust = TRUE` but `cluster` is not specified, `zelig()` assumes that each observation falls into its own cluster.

Zelig users may wish to refer to `help(survreg)` for more information.

Examples

1. Basic Example
   Attaching the sample dataset:

   ```
   > data(tobin)
   ```

   Estimating linear regression using `tobit`:

   ```
   > z.out <- zelig(durable ~ age + quant, model = "tobit", data = tobin)
   ```

   Setting values for the explanatory variables to their sample averages:

   ```
   > x.out <- setx(z.out)
   ```

   Simulating quantities of interest from the posterior distribution given `x.out`.

   ```
   > s.out1 <- sim(z.out, x = x.out)
   ```

   ```
   > summary(s.out1)
   ```

2. Simulating First Differences
   Set explanatory variables to their default(mean/mode) values, with high (80th percentile) and low (20th percentile) liquidity ratio (`quant`):

   ```
   > x.high <- setx(z.out, quant = quantile(tobin$quant, prob = 0.8))
   > x.low <- setx(z.out, quant = quantile(tobin$quant, prob = 0.2))
   ```

   Estimating the first difference for the effect of high versus low liquidity ratio on duration(`durable`):

   ```
   > s.out2 <- sim(z.out, x = x.high, x1 = x.low)
   ```

   ```
   > summary(s.out2)
   ```

Model

- Let Y_i^* be a latent dependent variable which is distributed with *stochastic* component

  $$Y_i^* \sim \text{Normal}(\mu_i, \sigma^2)$$

  where μ_i is a vector means and σ^2 is a scalar variance parameter. Y_i^* is not directly observed, however. Rather we observed Y_i which is defined as:

  $$Y_i = \begin{cases} Y_i^* & \text{if} \quad c < Y_i^* \\ c & \text{if} \quad c \geq Y_i^* \end{cases}$$

  where c is the lower bound below which Y_i^* is censored.

- The *systematic component* is given by

$$\mu_i = x_i \beta,$$

where x_i is the vector of k explanatory variables for observation i and β is the vector of coefficients.

Quantities of Interest

- The expected values (`qi$ev`) for the tobit regression model are the same as the expected value of $Y*$:

$$E(Y^*|X) = \mu_i = x_i \beta$$

- The first difference (`qi$fd`) for the tobit regression model is defined as

$$\text{FD} = E(Y^* \mid x_1) - E(Y^* \mid x).$$

- In conditional prediction models, the average expected treatment effect (`qi$att.ev`) for the treatment group is

$$\frac{1}{\sum t_i} \sum_{i:t_i=1} [E[Y_i^*(t_i = 1)] - E[Y_i^*(t_i = 0)]],$$

where t_i is a binary explanatory variable defining the treatment ($t_i = 1$) and control ($t_i = 0$) groups.

Output Values

The output of each Zelig command contains useful information which you may view. For example, if you run:

```
z.out <- zelig(y ~ x, model = "tobit.bayes", data)
```

then you may examine the available information in `z.out` by using `names(z.out)`, see the draws from the posterior distribution of the `coefficients` by using `z.out$coefficients`, and view a default summary of information through `summary(z.out)`. Other elements available through the `$` operator are listed below.

- From the `zelig()` output object `z.out`, you may extract:
  - `coefficients`: draws from the posterior distributions of the estimated parameters. The first k columns contain the posterior draws of the coefficients β, and the last column contains the posterior draws of the variance σ^2.
  - `zelig.data`: the input data frame if `save.data = TRUE`.
  - `seed`: the random seed used in the model.

- From the `sim()` output object `s.out`:
  - `qi$ev`: the simulated expected value for the specified values of `x`.
  - `qi$fd`: the simulated first difference in the expected values given the values specified in `x` and `x1`.
  - `qi$att.ev`: the simulated average expected treatment effect for the treated from conditional prediction models.

How to Cite

To cite the *oprobit.bayes* Zelig model use:

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "oprobit.bayes: Bayesian Ordered Probit Regression," in Kosuke Imai,Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," `http://gking.harvard.edu/zelig`.

To cite Zelig as a whole, please reference these two sources:

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," `http://GKing.harvard.edu/zelig`.

> Kosuke Imai, Gary King, and Olivia Lau. 2008. "Toward A Common Framework for Statistical Analysis and Development," *Journal of Computational and Graphical Statistics*, forthcoming, `http://gking.harvard.edu/files/abs/z-abs.shtml`.

See also

The tobit function is part of the survival library by Terry Therneau, ported to R by Thomas Lumley. Advanced users may wish to refer to `help(survfit)` in the survival library and Venables and Ripley (2002).Sample data are from King et al. (1990).

Bibliography

Huber, P. J. (1981), *Robust Statistics*, Wiley.

King, G., Alt, J., Burns, N., and Laver, M. (1990), “A Unified Model of Cabinet Dissolution in Parliamentary Democracies,” *American Journal of Political Science*, 34, 846–871, http://gking.harvard.edu/files/abs/coal-abs.shtml.

Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, Springer-Verlag, 4th ed.

White, H. (1980), “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica*, 48, 817–838.