1 `oprobit`: Ordinal Probit Regression for Ordered Categorical Dependent Variables

Use the ordinal probit regression model if your dependent variables are ordered and categorical. They may take on either integer values or character strings.

1.0.1 Syntax

```
> z.out <- zelig(as.factor(Y) ~ X1 + X2, model = "oprobit", data = mydata)
> x.out <- setx(z.out)
> s.out <- sim(z.out, x = x.out)
```

If `Y` takes discrete integer values, the `as.factor()` command will order it automatically. If `Y` takes on values composed of character strings, such as "strongly agree", "agree", and "disagree", `as.factor()` will order the values in the order in which they appear in `Y`. You will need to replace your dependent variable with a factored variable prior to estimating the model through `zelig()`.

1.0.2 Example

1. Creating An Ordered Dependent Variable

   Load the sample data:

   ```
   >  data(sanction)
   ```

   Create an ordered dependent variable:

   ```
   >  sanction$ncost <- factor(sanction$ncost, ordered = TRUE,
   +                            levels = c("net gain", "little effect",
   +                            "modest loss", "major loss"))
   ```

   Estimate the model:

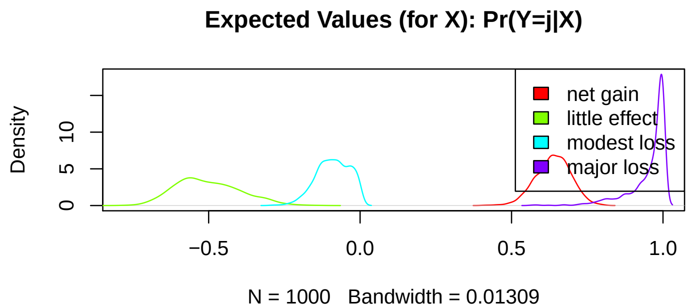
   ```
   >  z.out <- zelig(ncost ~ mil + coop, model = "oprobit", data = sanction)
   >  summary(z.out)
   ```

   Set the explanatory variables to their observed values:

   ```
   >  x.out <- setx(z.out, fn = NULL)
   ```

   Simulate fitted values given `x.out` and view the results:

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



2. First Differences

   Using the sample data `sanction`, let us estimate the empirical model and return the coefficients:

```
>  z.out <- zelig(as.factor(cost) ~ mil + coop, model = "oprobit",
+                 data = sanction)
```

```
> summary(z.out)
```

   Set the explanatory variables to their means, with `mil` set to 0 (no military action in addition to sanctions) in the baseline case and set to 1 (military action in addition to sanctions) in the alternative case:

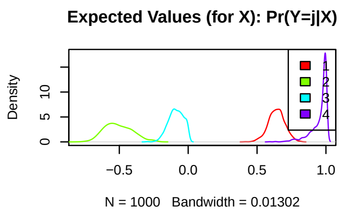
```
>  x.low <- setx(z.out, mil = 0)
>  x.high <- setx(z.out, mil = 1)
```

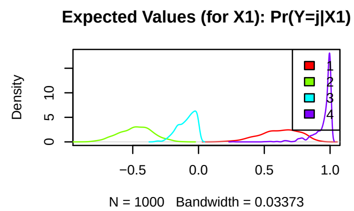
   Generate simulated fitted values and first differences, and view the results:

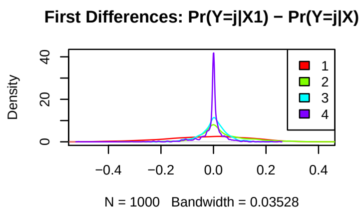
```
>  s.out <- sim(z.out, x = x.low, x1 = x.high)
```

```
> summary(s.out)
```

```
>  plot(s.out)
```







1.0.3 Model

Let Y_i be the ordered categorical dependent variable for observation i that takes one of the integer values from 1 to J where J is the total number of categories.

- The *stochastic component* is described by an unobserved continuous variable, Y_i^*, which follows the normal distribution with mean μ_i and unit variance

  $$Y_i^* \;\sim\; N(\mu_i, 1).$$

  The observation mechanism is

  $$Y_i \;=\; j \quad \text{if} \quad \tau_{j-1} \le Y_i^* \le \tau_j \quad \text{for} \quad j = 1, \ldots, J.$$

  where τ_k for $k = 0, \ldots, J$ is the threshold parameter with the following constraints; $\tau_l < \tau_m$ for all $l < m$ and $\tau_0 = -\infty$ and $\tau_J = \infty$.

  Given this observation mechanism, the probability for each category, is given by

  $$\Pr(Y_i = j) \;=\; \Phi(\tau_j \mid \mu_i) - \Phi(\tau_{j-1} \mid \mu_i) \quad \text{for} \quad j = 1, \ldots, J$$

  where $\Phi(\mu_i)$ is the cumulative distribution function for the Normal distribution with mean μ_i and unit variance.

- The *systematic component* is given by

$$\mu_i \ = \ x_i \beta$$

where x_i is the vector of explanatory variables and β is the vector of coefficients.

1.0.4 Quantities of Interest

- The expected values (`qi$ev`) for the ordinal probit model are simulations of the predicted probabilities for each category:

$$E(Y_i = j) \ = \ \Pr(Y_i = j) \ = \ \Phi(\tau_j \mid \mu_i) - \Phi(\tau_{j-1} \mid \mu_i) \quad \text{for} \quad j = 1, \ldots, J,$$

given draws of β from its posterior.

- The predicted value (`qi$pr`) is the observed value of Y_i given the underlying standard normal distribution described by μ_i.

- The difference in each of the predicted probabilities (`qi$fd`) is given by

$$\Pr(Y = j \mid x_1) \ - \ \Pr(Y = j \mid x) \quad \text{for} \quad j = 1, \ldots, J.$$

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

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

where t_i is a binary explanatory variable defining the treatment ($t_i = 1$) and control ($t_i = 0$) groups, and n_j is the number of treated observations in category j.

- In conditional prediction models, the average predicted treatment effect (`qi$att.pr`) for the treatment group in category j is

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

where t_i is a binary explanatory variable defining the treatment ($t_i = 1$) and control ($t_i = 0$) groups, and n_j is the number of treated observations in category j.

1.0.5 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 = "oprobit", data)`, then you may examine the available information in `z.out` by using `names(z.out)`, see the `coefficients` by using `z.out$coefficients`, and a default summary of information through `summary(z.out)`. Other elements available through the `$` operator are listed below.

- From the `zelig()` output object `z.out`, you may extract:
  - `coefficients`: the named vector of coefficients.
  - `fitted.values`: an $n \times J$ matrix of the in-sample fitted values.
  - `predictors`: an $n \times (J-1)$ matrix of the linear predictors $x_i \beta_j$.
  - `residuals`: an $n \times (J-1)$ matrix of the residuals.
  - `zeta`: a vector containing the estimated class boundaries.
  - `df.residual`: the residual degrees of freedom.
  - `df.total`: the total degrees of freedom.
  - `rss`: the residual sum of squares.
  - `y`: an $n \times J$ matrix of the dependent variables.
  - `zelig.data`: the input data frame if `save.data = TRUE`.
- From `summary(z.out)`, you may extract:
  - `coef3`: a table of the coefficients with their associated standard errors and t-statistics.
  - `cov.unscaled`: the variance-covariance matrix.
  - `pearson.resid`: an $n \times (m-1)$ matrix of the Pearson residuals.
- From the `sim()` output object `s.out`, you may extract quantities of interest arranged as arrays. Available quantities are:
  - `qi$ev`: the simulated expected probabilities for the specified values of `x`, indexed by simulation $\times$ quantity $\times$ `x`-observation (for more than one `x`-observation).
  - `qi$pr`: the simulated predicted values drawn from the distribution defined by the expected probabilities, indexed by simulation $\times$ `x`-observation.
  - `qi$fd`: the simulated first difference in the predicted probabilities for the values specified in `x` and `x1`, indexed by simulation $\times$ quantity $\times$ `x`-observation (for more than one `x`-observation).
  - `qi$att.ev`: the simulated average expected treatment effect for the treated from conditional prediction models.
  - `qi$att.pr`: the simulated average predicted treatment effect for the treated from conditional prediction models.

How to Cite the Ordinal Logit Model

Kosuke Imai, Olivia Lau, and Gary King. *ologit: Ordinal Logistic Regression for Ordered Categorical Dependent Variables*, 2011

How to Cite the Zelig Software Package

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`.

Imai, Kosuke, Gary King, and Olivia Lau. (2008). "Toward A Common Framework for Statistical Analysis and Development." Journal of Computational and Graphical Statistics, Vol. 17, No. 4 (December), pp. 892-913.

See also

The ordinal probit function is part of the VGAM package by Thomas Yee [3]. In addition, advanced users may wish to refer to `help(vglm)` in the VGAM library. Additional documentation is available at `http://www.stat.auckland.ac.nz/ ~\protect\kern+.1667em\relaxyee`http://www.stat.auckland.ac.nz/ yee.Sample data are from [2]

References

[1] Kosuke Imai, Olivia Lau, and Gary King. *ologit: Ordinal Logistic Regression for Ordered Categorical Dependent Variables*, 2011.

[2] Lisa Martin. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton University Press, 1992. Please inquire with Lisa Martin before publishing results from these data, as this dataset includes errors that have since been corrected.

[3] T. W. Yee and T. J. Hastie. Reduced-rank vector generalized linear models. *Statistical Modelling*, 3:15–41, 2003.