1 `ologit`: Ordinal Logistic Regression for Ordered Categorical Dependent Variables

Use the ordinal logit regression model if your dependent variable is ordered and categorical, either in the form of integer values or character strings.

1.0.1 Syntax

```
> z.out <- zelig(as.factor(Y) ~ X1 + X2, model = "ologit", 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 automatically order the values. 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()`. See Example 1 for more information on creating ordered factors.

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 = "ologit", data = sanction)
   ```

   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`, estimate the empirical model and returning the coefficients:

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

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* begins with an unobserved continuous variable, Y_i^*, which follows the standard logistic distribution with a parameter μ_i,

$$Y_i^* \sim \text{Logit}(y_i^* \mid \mu_i),$$

  to which we add an observation mechanism

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

  where τ_l (for $l = 0, \ldots, J$) are the threshold parameters with $\tau_l < \tau_m$ for all $l < m$ and $\tau_0 = -\infty$ and $\tau_J = \infty$.

- The *systematic component* has the following form, given the parameters τ_j and β, and the explanatory variables x_i:

$$\Pr(Y \leq j) = \Pr(Y^* \leq \tau_j) = \frac{\exp(\tau_j - x_i\beta)}{1 + \exp(\tau_j - x_i\beta)},$$

  which implies:

$$\pi_j = \frac{\exp(\tau_j - x_i\beta)}{1 + \exp(\tau_j - x_i\beta)} - \frac{\exp(\tau_{j-1} - x_i\beta)}{1 + \exp(\tau_{j-1} - x_i\beta)}.$$

1.0.4 Quantities of Interest

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

$$E(Y = j) \;=\; \pi_j \;=\; \frac{\exp(\tau_j - x_i\beta)}{1+\exp(\tau_j - x_i\beta)} - \frac{\exp(\tau_{j-1} - x_i\beta)}{1+\exp(\tau_{j-1} - x_i\beta)},$$

  given a draw of β from its sampling distribution.

- The predicted value (`qi$pr`) is drawn from the logit distribution described by μ_i, and observed as one of J discrete outcomes.

- 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 (`att.ev`) for the treatment group is

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

  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 (`att.pr`) for the treatment group is

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

  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 = "ologit", 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`: parameter estimates for the explanatory variables.

- `zeta`: a vector containing the estimated class boundaries τ_j.
- `deviance`: the residual deviance.
- `fitted.values`: the $n \times J$ matrix of in-sample fitted values.
- `df.residual`: the residual degrees of freedom.
- `edf`: the effective degrees of freedom.
- `Hessian`: the Hessian matrix.
- `zelig.data`: the input data frame if `save.data = TRUE`.

- From `summary(z.out)`, you may extract:
  - `coefficients`: the parameter estimates with their associated standard errors, and t-statistics.
- 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 logit model is part of the MASS package by William N. Venable and Brian D. Ripley [4]. Advanced users may wish to refer to `help(polr)` as well as [3]. 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] Peter McCullagh and James A. Nelder. *Generalized Linear Models*. Number 37 in Monograph on Statistics and Applied Probability. Chapman & Hall, 2nd edition, 1989.

[4] William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S*. Springer-Verlag, 4th edition, 2002.