0.1 gamma: Gamma Regression for Continuous, Positive Dependent Variables

Use the gamma regression model if you have a positive-valued dependent variable such as the number of years a parliamentary cabinet endures, or the seconds you can stay airborne while jumping. The gamma distribution assumes that all waiting times are complete by the end of the study (censoring is not allowed).

Syntax

> z.out <- zelig(Y ~ X1 + X2, model = "gamma", data = mydata)
> x.out <- setx(z.out)
> s.out <- sim(z.out, x = x.out, x1 = NULL)

Additional Inputs

In addition to the standard inputs, zelig() takes the following additional options for gamma regression:

- **robust**: defaults to FALSE. If TRUE is selected, zelig() computes robust standard errors via the sandwich package (see Zeileis (2004)). The default type of robust standard error is heteroskedastic and autocorrelation consistent (HAC), and assumes that observations are ordered by time index.

In addition, robust may be a list with the following options:

- **method**: Choose from
  - "vcovHAC": (default if robust = TRUE) HAC standard errors.
  - "weave": HAC standard errors using the weights given in Lumley and Haggerty (1999).

- **order.by**: defaults to NULL (the observations are chronologically ordered as in the original data). Optionally, you may specify a vector of weights (either as order.by = z, where z exists outside the data frame; or as order.by = ~z, where z is a variable in the data frame). The observations are chronologically ordered by the size of z.

- ...: additional options passed to the functions specified in method. See the sandwich library and Zeileis (2004) for more options.

Example

Attach the sample data:

> data(coalition)
Estimate the model:

```r
> z.out <- zelig(duration ~ fract + numst2, model = "gamma", data = coalition)
```

View the regression output:

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

Set the baseline values (with the ruling coalition in the minority) and the alternative values (with the ruling coalition in the majority) for X:

```r
> x.low <- setx(z.out, numst2 = 0)
> x.high <- setx(z.out, numst2 = 1)
```

Simulate expected values (qi$ev) and first differences (qi$fd):

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

- The Gamma distribution with scale parameter $\alpha$ has a *stochastic component*:

$$Y \sim \text{Gamma}(y_i \mid \lambda_i, \alpha)$$

$$f(y) = \frac{1}{\alpha \lambda_i \Gamma \lambda_i} y_i^{\lambda_i-1} \exp \left\{ -\frac{y_i}{\alpha} \right\}$$

for $\alpha, \lambda_i, y_i > 0$.

- The *systematic component* is given by

$$\lambda_i = \frac{1}{x_i \beta}$$

Quantities of Interest

- The expected values ($\text{qi$\text{ev}$}$) are simulations of the mean of the stochastic component given draws of $\alpha$ and $\beta$ from their posteriors:

$$E(Y) = \alpha \lambda_i.$$ 

- The predicted values ($\text{qi$\text{pr}$}$) are draws from the gamma distribution for each given set of parameters ($\alpha, \lambda_i$).

- If $x1$ is specified, $\text{sim()}$ also returns the differences in the expected values ($\text{qi$\text{fd}$}$), $E(Y \mid x1) - E(Y \mid x)$.

- In conditional prediction models, the average expected treatment effect ($\text{att.ev}$) for the treatment group is

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

where $t_i$ is a binary explanatory variable defining the treatment ($t_i = 1$) and control ($t_i = 0$) groups. Variation in the simulations are due to uncertainty in simulating $E[Y_i(t_i = 0)]$, the counterfactual expected value of $Y_i$ for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_i = 0$. 


In conditional prediction models, the average predicted treatment effect (\texttt{att.pr}) for the treatment group is

\[
\frac{1}{\sum_{i=1}^{n} t_i} \sum_{i:t_i=1}^{n} \left\{ Y_i(t_i = 1) - \widehat{Y}_i(t_i = 0) \right\},
\]

where \( t_i \) is a binary explanatory variable defining the treatment (\( t_i = 1 \)) and control (\( t_i = 0 \)) groups. Variation in the simulations are due to uncertainty in simulating \( \widehat{Y}_i(t_i = 0) \), the counterfactual predicted value of \( Y_i \) for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to \( t_i = 0 \).

**Output Values**

The output of each Zelig command contains useful information which you may view. For example, if you run \texttt{z.out <- zelig(y ~ x, model = "gamma", data)}, then you may examine the available information in \texttt{z.out} by using \texttt{names(z.out)}, see the \texttt{coefficients} by using \texttt{z.out$coefficients}, and a default summary of information through \texttt{summary(z.out)}. Other elements available through the \$ operator are listed below.

- From the \texttt{zelig()} output object \texttt{z.out}, you may extract:
  - \texttt{coefficients}: parameter estimates for the explanatory variables.
  - \texttt{residuals}: the working residuals in the final iteration of the IWLS fit.
  - \texttt{fitted.values}: the vector of fitted values.
  - \texttt{linear.predictors}: the vector of \( x_i \hat{\beta} \).
  - \texttt{aic}: Akaike’s Information Criterion (minus twice the maximized log-likelihood plus twice the number of coefficients).
  - \texttt{df.residual}: the residual degrees of freedom.
  - \texttt{df.null}: the residual degrees of freedom for the null model.
  - \texttt{zelig.data}: the input data frame if \texttt{save.data = TRUE}.

- From \texttt{summary(z.out)}, you may extract:
  - \texttt{coefficients}: the parameter estimates with their associated standard errors, \( p \)-values, and \( t \)-statistics.
  - \texttt{cov.scaled}: a \( k \times k \) matrix of scaled covariances.
  - \texttt{cov.unscaled}: a \( k \times k \) matrix of unscaled covariances.

- From the \texttt{sim()} output object \texttt{s.out}, you may extract quantities of interest arranged as matrices indexed by simulation \( \times \) \texttt{x-observation} (for more than one \texttt{x-observation}). Available quantities are:
qi$ev: the simulated expected values for the specified values of x.
qi$pr: the simulated predicted values drawn from a distribution defined by $(\alpha, \lambda_i)$.
qi$fd: the simulated first difference in the expected values for the specified values in x and x1.
qi$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
To cite the gamma Zelig model:


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


See also
The gamma model is part of the base packages by William N. Venable and Brian D. Ripley [Venables and Ripley 2002]. Advanced users may wish to refer to help(glm) and help(family), as well as McCullagh and Nelder [1989]. Robust standard errors are implemented via sandwich package by Achim Zeileis [Zeileis 2004]. Sample data are from King et al. [1990].
Bibliography


