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).

1.0.1 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)
```

1.0.2 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 [7]). 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.
    - `"kernHAC"`: HAC standard errors using the weights given in [1].
    - `"weave"`: HAC standard errors using the weights given in [4].
  - `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 [7] for more options.

1.0.3 Example

Attach the sample data:

```
> data(coalition)
```

Estimate the model:

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

```
 How to cite this model in Zelig:
  Kosuke Imai, Gary King, and Olivia Lau. 2011.
  "gamma: Gamma Regression for Continuous, Positive Dependent Variables"
  in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software,"
  http://gking.harvard.edu/zelig
```

View the regression output:

```
> summary(z.out)

Call:
glm(formula = duration ~ fract + numst2, family = Function, data = Data.frame,
    model = FALSE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2510  -0.9112  -0.2278   0.4132   1.5360

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.296e-02  1.329e-02  -0.975  0.33016
fract        1.149e-04  1.723e-05   6.668 1.19e-10 ***
numst2      -1.739e-02  5.881e-03  -2.957  0.00335 **
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

(Dispersion parameter for Gamma family taken to be 0.6291004)

    Null deviance: 300.74  on 313  degrees of freedom
Residual deviance: 272.19  on 311  degrees of freedom
AIC: 2428.1

Number of Fisher Scoring iterations: 6
```

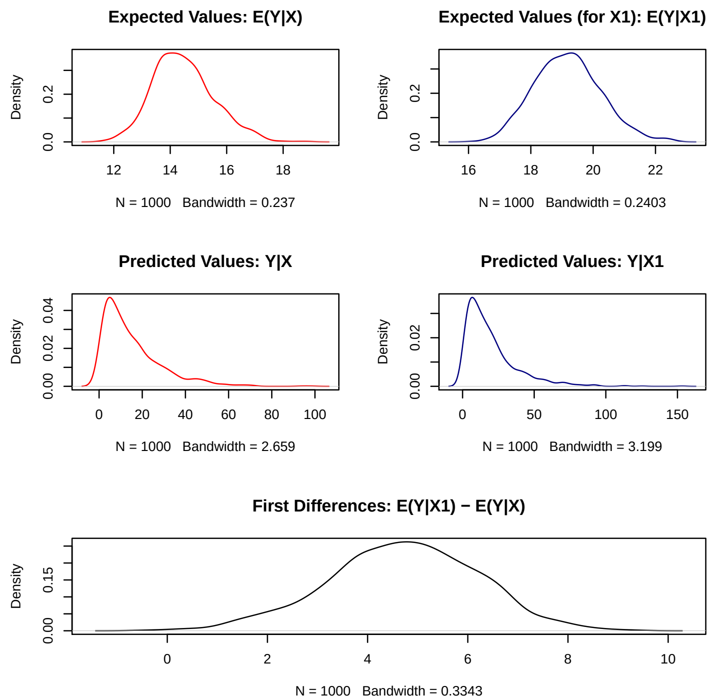
Set the baseline values (with the ruling coalition in the minority) and the alternative values (with the ruling coalition in the majority) for X:

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

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

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

> summary(s.out)

Model:  gamma
Number of simulations:  1000
```

```
Values of X
     fract numst2
1 718.8121      0

Values of X1
     fract numst2
1 718.8121      1

Expected Values: E(Y|X)
   mean    sd    50%   2.5%  97.5%
 14.472 1.076 14.365 12.527 16.821

Expected Values (for X1): E(Y|X1)
   mean    sd    50%   2.5% 97.5%
 19.177 1.063 19.153 17.288 21.43

Predicted Values: Y|X
  mean     sd    50%  2.5%  97.5%
 15.08 13.602 11.049 0.814 50.109

Predicted Values: Y|X1
   mean     sd   50% 2.5%  97.5%
 19.597 18.232 14.31 1.16 69.568

First Differences: E(Y|X1) - E(Y|X)
  mean    sd   50%  2.5% 97.5%
 4.705 1.495 4.754 1.592 7.523

> plot(s.out)
```



1.0.4 Model

- The Gamma distribution with scale parameter α has a *stochastic component*:

$$\begin{aligned} Y &\sim \text{Gamma}(y_i \mid \lambda_i, \alpha) \\ f(y) &= \frac{1}{\alpha^{\lambda_i} \, \Gamma \lambda_i} \, y_i^{\lambda_i - 1} \exp - \left\{ \frac{y_i}{\alpha} \right\} \end{aligned}$$

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

- The *systematic component* is given by

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

1.0.5 Quantities of Interest

- The expected values (`qi$ev`) are simulations of the mean of the stochastic component given draws of α and β from their posteriors:

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

- The predicted values (`qi$pr`) are draws from the gamma distribution for each given set of parameters (α, λ_i).

- If `x1` is specified, `sim()` also returns the differences in the expected values (`qi$fd`),

$$E(Y \mid x_1) - E(Y \mid x)$$

  .

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

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

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

1.0.6 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 = "gamma", 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.

- `residuals`: the working residuals in the final iteration of the IWLS fit.
- `fitted.values`: the vector of fitted values.
- `linear.predictors`: the vector of $x_i\beta$.
- `aic`: Akaike's Information Criterion (minus twice the maximized log-likelihood plus twice the number of coefficients).
- `df.residual`: the residual degrees of freedom.
- `df.null`: the residual degrees of freedom for the null model.
- `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, p-values, and t-statistics.
  - `cov.scaled`: a $k \times k$ matrix of scaled covariances.
  - `cov.unscaled`: a $k \times k$ matrix of unscaled covariances.
- From the `sim()` output object `s.out`, you may extract quantities of interest arranged as matrices indexed by simulation $\times$ `x`-observation (for more than one `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 (α, λ_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 the Gamma Model

Kosuke Imai, Olivia Lau, and Gary King. *logit: Logistic Regression for Dichotomous Dependent*, 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 gamma model is part of the stats package by **(author?)** [6]. Advanced users may wish to refer to `help(glm)` and `help(family)`, as well as [5]. Robust standard errors are implemented via the sandwich package by **(author?)** [7]. Sample data are from [3].

References

[1] Donald W.K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, May 1991.

[2] Kosuke Imai, Olivia Lau, and Gary King. *logit: Logistic Regression for Dichotomous Dependent*, 2011.

[3] Gary King, Michael Tomz, and Jason Wittenberg. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2):341–355, April 2000. http://gking.harvard.edu/files/abs/making-abs.shtml.

[4] Thomas Lumley and Patrick Heagerty. Weighted empirical adaptive variance estimators for correlated data regression. *jrssb*, 61(2):459–477, 1999.

[5] Peter McCullagh and James A. Nelder. *Generalized Linear Models*. Number 37 in Monograph on Statistics and Applied Probability. Chapman & Hall, 2nd edition, 1989.

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

[7] Achim Zeileis. Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17, 2004.