

# Full Gamma Regression

Daniel Gstrein  
Universität Innsbruck

---

## Abstract

The **fgamma** package (<https://R-Forge.R-project.org/projects/uibk-rprog-2017/>) fits full gamma regression models estimating the parameters  $\mu$  and  $\sigma$  using maximum likelihood estimation. A brief overview of the package is provided, along with some illustrations and a replication of results from the **gamlss** package.

*Keywords:* full gamma, regression, R.

---

## 1. Introduction

The specific parameterization of the gamma distribution used in **fgamma** is

$$f(y_i|\mu_i, \sigma_i) = \frac{y_i^{(1/\sigma_i^2-1)} \exp[-y_i/(\sigma_i^2\mu_i)]}{(\sigma_i^2\mu_i)^{(1/\sigma_i^2)}\Gamma(1/\sigma_i^2)}$$

for  $y_i > 0$ ,  $\mu_i > 0$  and  $\sigma_i > 0$ .

Here  $E(Y) = \mu$  and  $Var(Y) = \sigma^2\mu^2$ . This reparametrization is obtained by setting  $a = 1/\sigma^2$  and  $s = \sigma^2\mu$  from the specification in **dgamma()**.

The mean  $\mu_i$  and  $\sigma_i$  are linkend to the linear predictors

$$\begin{aligned}\log(\mu_i) &= x_i^\top \beta \\ \log(\sigma_i) &= z_i^\top \gamma\end{aligned}$$

where the regressor vectors  $x_i$  and  $z_i$  can be set up without restrictions, i.e., they can be identical, overlapping or completely different or just including an intercept, etc.

See also [Rigby and Stasinopoulos \(2005\)](#) for a more detailed introduction to this model class as well as a better implementation in the package **gamlss**. The main purpose of **fgamma** is to illustrate how to create such a package *from scratch*.

## 2. Implementation

As usual in many other regression packages for R ([R Core Team 2017](#)), the main model fitting function **fgamma()** uses a formula-based interface and returns an (S3) object of class **fgamma**:

```
fgamma(formula, data, subset, na.action,
  model = TRUE, y = TRUE, x = FALSE,
  control = fgamma_control(...), ...)
```

Actually, the formula can be a two-part Formula (Zeileis and Croissant 2010), specifying separate sets of regressors  $x$  and  $z$  for the mean and sigma submodels, respectively.

The underlying workhorse function is `fgamma_fit()` which has a matrix interface and returns an unclassed list.

A number of standard S3 methods are provided, see Table 1.

Method	Description
<code>print()</code>	Simple printed display with coefficients
<code>summary()</code>	Standard regression summary; returns <code>summary.fgamma</code> object (with <code>print()</code> method)
<code>coef()</code>	Extract coefficients
<code>vcov()</code>	Associated covariance matrix
<code>predict()</code>	(Different types of) predictions for new data
<code>fitted()</code>	Fitted values for observed data
<code>residuals()</code>	Extract (different types of) residuals
<code>terms()</code>	Extract terms
<code>model.matrix()</code>	Extract model matrix (or matrices)
<code>nobs()</code>	Extract number of observations
<code>logLik()</code>	Extract fitted log-likelihood
<code>bread()</code>	Extract bread for <b>sandwich</b> covariance
<code>estfun()</code>	Extract estimating functions (= gradient contributions) for <b>sandwich</b> covariances
<code>getSummary()</code>	Extract summary statistics for <code>mtable()</code>

Table 1: S3 methods provided in **fgamma**.

Due to these methods a number of useful utilities work automatically, e.g., `AIC()`, `BIC()`, `coeftest()` (**lmtest**), `lrtest()` (**lmtest**), `waldtest()` (**lmtest**), `linearHypothesis()` (**car**), `mtable()` (**memisc**), etc.

### 3. Illustration

To illustrate the package's use in practice, a comparison of several gamma regression models is applied to data on average hourly earnings of 526 people (taken from Wooldridge 2010). A gamma regression model using all explanatory variables for both parameters can be set up by:

```
R> data("Wages", package = "fgamma")
R> library("fgamma")
R> f1 <- fgamma(wage ~ . | ., data = Wages)
R> summary(f1)
```

Call:

```
fgamma(formula = wage ~ . | ., data = Wages)
```

Standardized residuals:

Min	1Q	Median	3Q	Max
-2.1442	-0.7395	-0.2713	0.4531	5.5329

Coefficients (mu model with log link):

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	0.444945	0.120222	3.701	0.000215	***
educ	0.080556	0.007493	10.751	< 2e-16	***
exper	0.003669	0.001866	1.966	0.049245	*
tenure	0.021025	0.003584	5.867	4.45e-09	***
marriageNotmarried	-0.129821	0.042835	-3.031	0.002439	**
numdep	-0.014849	0.015008	-0.989	0.322446	
ethnicityNonwhite	0.028303	0.059678	0.474	0.635316	
sexMale	0.282226	0.038529	7.325	2.39e-13	***

Coefficients (sigma model with log link):

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1.174208	0.196651	-5.971	2.36e-09	***
educ	0.013073	0.011753	1.112	0.2660	
exper	0.002385	0.003275	0.728	0.4664	
tenure	0.010696	0.004881	2.191	0.0284	*
marriageNotmarried	0.083048	0.076690	1.083	0.2788	
numdep	-0.044625	0.028507	-1.565	0.1175	
ethnicityNonwhite	0.038160	0.101086	0.378	0.7058	
sexMale	0.028071	0.063852	0.440	0.6602	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -1136 on 16 Df

Number of iterations in BFGS optimization: 29

This model is now modified in the direction that the variables influencing the mean parameter are also employed in the sigma submodel. Clearly, not all explanatory variables are statistically significant.

Therefore, the following models are considered:

```
R> f2 <- fgamma(wage ~ educ + exper + tenure + ethnicity + sex + marriage + numdep |
+               educ + exper + tenure, data = Wages)
R> f3 <- fgamma(wage ~ educ + exper + tenure + sex + marriage |
+               tenure, data = Wages)
R> f4 <- fgamma(wage ~ educ + exper + tenure + sex + marriage | 1
+               , data = Wages)
R> BIC(f1, f2, f3, f4)
```

```

      df      BIC
f1 16 2372.241
f2 12 2353.054
f3  8 2329.911
f4  7 2330.707

```

The BIC would choose Model 3 but a likelihood ratio test would prefer the most parsimonious model:

```

R> library("lmtest")
R> lrtest(f1, f2, f3, f4)

```

Likelihood ratio test

```

Model 1: wage ~ . | .
Model 2: wage ~ educ + exper + tenure + ethnicity + sex + marriage + numdep |
      educ + exper + tenure
Model 3: wage ~ educ + exper + tenure + sex + marriage | tenure
Model 4: wage ~ educ + exper + tenure + sex + marriage | 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   16 -1136.0
2   12 -1138.9 -4  5.8738  0.208775
3    8 -1139.9 -4  1.9185  0.750742
4    7 -1143.4 -1  7.0611  0.007877 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## 4. Replication

To assess the reliability of the `fgamma()` implementation, it is benchmarked against the `gamlss()` function of (Rigby and Stasinopoulos 2005), using Model 3.

```

R> library("gamlss")
R> g1 <- gamlss(wage ~ educ + exper + tenure + sex + marriage,
+             sigma.formula = ~ tenure, data = Wages, family = GA)

```

```

GAMLSS-RS iteration 1: Global Deviance = 2280.499
GAMLSS-RS iteration 2: Global Deviance = 2279.795
GAMLSS-RS iteration 3: Global Deviance = 2279.789
GAMLSS-RS iteration 4: Global Deviance = 2279.789

```

```

R> summary(g1)

```

```

*****
Family:  c("GA", "Gamma")

```

```
Call:  gamlss(formula = wage ~ educ + exper + tenure + sex +
  marriage, sigma.formula = ~tenure, family = GA,
  data = Wages)
```

```
Fitting method: RS()
```

```
-----
Mu link function:  log
```

```
Mu Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	0.369770	0.099172	3.729	0.000214	***
educ	0.084983	0.006469	13.138	< 2e-16	***
exper	0.003572	0.001685	2.120	0.034503	*
tenure	0.021293	0.003540	6.014	3.41e-09	***
sexMale	0.276344	0.036916	7.486	3.08e-13	***
marriageNotmarried	-0.104067	0.038741	-2.686	0.007458	**

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
-----
Sigma link function:  log
```

```
Sigma Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-0.958991	0.037420	-25.628	<2e-16	***
tenure	0.011001	0.004317	2.548	0.0111	*

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
-----
No. of observations in the fit:  526
```

```
Degrees of Freedom for the fit:  8
```

```
Residual Deg. of Freedom:  518
```

```
at cycle:  4
```

```
Global Deviance:  2279.789
```

```
AIC:  2295.789
```

```
SBC:  2329.911
```

```
*****
```

It can be easily seen the results can be replicated using both packages:

```
R> summary(f3)
```

```
Call:
```

```
fgamma(formula = wage ~ educ + exper + tenure + sex + marriage |
  tenure, data = Wages)
```

Standardized residuals:

Min	1Q	Median	3Q	Max
-2.2143	-0.7354	-0.2711	0.4599	5.5163

Coefficients (mu model with log link):

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	0.369759	0.099174	3.728	0.000193	***
educ	0.084982	0.006469	13.137	< 2e-16	***
exper	0.003572	0.001685	2.120	0.034018	*
tenure	0.021298	0.003541	6.015	1.80e-09	***
sexMale	0.276316	0.036916	7.485	7.15e-14	***
marriageNotmarried	-0.104075	0.038741	-2.686	0.007221	**

Coefficients (sigma model with log link):

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-0.959003	0.037421	-25.627	<2e-16	***
tenure	0.011003	0.004319	2.548	0.0108	*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -1140 on 8 Df

Number of iterations in BFGS optimization: 28

## References

- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rigby RA, Stasinopoulos DM (2005). “Generalized additive models for location, scale and shape,(with discussion).” *Applied Statistics*, **54**, 507–554.
- Wooldridge J (2010). *Introductory Econometrics: A Modern Approach*. 4th edition. South-Western College Pub.
- Zeileis A, Croissant Y (2010). “Extended Model Formulas in R: Multiple Parts and Multiple Responses.” *Journal of Statistical Software*, **34**(1), 1–13. doi:10.18637/jss.v034.i01.

### Affiliation:

Daniel Gstrein  
 Department of Statistics  
 Faculty of Economics and Statistics  
 Universität Innsbruck  
 Universitätsstr. 15

6020 Innsbruck, Austria

E-mail: [Daniel.Gstrein@uibk.ac.at](mailto:Daniel.Gstrein@uibk.ac.at)