Lecture 8: Gamma regression

Claudia Czado

TU München

Overview

- Models with constant coefficient of variation
- Gamma regression: estimation and testing
- Gamma regression with weights

Motivation

Linear models:	$Var(Y_i) = \sigma^2$	constant
Poisson models:	$Var(Y_i) = E(Y_i) = \mu_i$	non constant, linear
Constant coefficient of variation:	$\frac{[Var(Y_i)]^{1/2}}{E(Y_i)} = \sigma$	
	$\Rightarrow Var(Y_i) = \sigma^2 \mu_i^2$	non constant, quadratic

Which transformation stablizes the variance if $Var(Y_i) = \sigma^2\mu_i^2$ holds?

Answer: $Z_i := \log(Y_i)$

Proof: According to a 2nd order Taylor expansion we have

$$Z_i = \log(Y_i) \approx \log(\mu_i) + \frac{1}{\mu_i}(Y_i - \mu_i) - \frac{1}{2\mu_i^2}(Y_i - \mu_i)^2$$

for σ^2 small

$$\Rightarrow \quad E(Z_i) \approx \log(\mu_i) - \frac{1}{2\mu_i^2}\underbrace{Var(Y_i)}_{=\sigma^2\mu_i^2} = \log(\mu_i) - \frac{1}{2}\sigma^2$$

According to a 1st order Taylor expansion we have

$$E(Z_i) \approx \log(\mu_i)$$

$$E(Z_i^2) \approx E\left([\log(\mu_i) + \frac{1}{\mu_i}(Y_i - \mu_i)]^2\right)$$

$$= (\log\mu_i)^2 + 2E\left(\log(\mu_i)\cdot\frac{1}{\mu_i}(Y_i - \mu_i)\right) + E\left(\frac{1}{\mu_i^2}(Y_i - \mu_i)^2\right)$$

$$= (\log\mu_i)^2 + 0 + \frac{1}{\mu_i^2}\sigma^2\mu_i^2 = (\log\mu_i)^2 + \sigma^2$$

$$\Rightarrow \quad Var(Z_i) \approx (\log\mu_i)^2 + \sigma^2 - (\log\mu_i)^2 = \sigma^2$$

For small σ^2 we have $E(Z_i) \approx \log(\mu_i) - \sigma^2/2$ and $Var(Z_i) \approx \sigma^2$

Multiplicative error structure: Log normal model

Assume $Y_i = \underbrace{e^{\beta_0} x_{i1}^{\beta_1} \cdots x_{ip}^{\beta_p}}_{E(Y_i) =: \mu_i} (1 + \epsilon_i)$

where $\epsilon_i := \frac{Y_i - E(Y_i)}{E(Y_i)} \Rightarrow E(\epsilon_i) = 0, \; Var(\epsilon_i) = \frac{Var(Y_i)}{E(Y_i)^2} =: \sigma^2$

$$\Rightarrow \quad \ln(Y_i) = \beta_0 + \beta_1 \ln x_{i1} + \ldots + \beta_p \ln x_{ip} + \ln(1 + \epsilon_i)$$

where $E(\ln(1 + \epsilon_i)) = E\left(\ln(1 + \frac{Y_i - E(Y_i)}{E(Y_i)})\right)$
$= E\left(\ln(\frac{Y_i}{E(Y_i)})\right) = \underbrace{E(\ln(Y_i))}_{\approx \ln \mu_i - \frac{\sigma^2}{2}} - \ln E(Y_i) = -\sigma^2/2$

and $Var(\ln(1 + \epsilon_i)) = Var(\ln(Y_i) - \ln E(Y_i)) = Var(\ln(Y_i)) \approx \sigma^2$

Consider

$$\ln Y_i = \beta_0 + \beta_1 \ln x_{i1} + \ldots + \beta_p \ln x_{ip} - \frac{\sigma^2}{2} + \epsilon_i',$$

where $\epsilon_i' := \ln(1 + \epsilon_i) + \frac{\sigma^2}{2} \quad \Rightarrow E(\epsilon_i') \approx 0,\ Var(\epsilon_i') \approx \sigma^2$

If one assumes $\epsilon_i' \sim N(0, \sigma^2)$ $i.i.d.$ and fits the log normal model

$$\ln Y_i = \alpha_0 + \alpha_1 \ln x_{i1} + \ldots + \alpha_p \ln x_{ip} + \epsilon_i'$$

then $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ are consistent estimates for $\beta_1, \ldots, \beta_p$, but $\hat{\alpha}_0$ is biased for β_0 with approx. bias $-\sigma^2/2$.

Original scale: Gamma regression

If one wants to work on original scale

$$\mu_i = E(Y_i) = e^{\beta_0} x_{i1}^{\beta_1} \cdots x_{ip}^{\beta_p} = e^{\overbrace{(\ln \mathbf{x_i})^t \boldsymbol{\beta}}_{\eta_i}}$$

$$\ln \mathbf{x_i} := (1, \ln(x_{i1}, \ldots, \ln(x_{ip}))^t$$

$$\Rightarrow \quad g(\mu_i) = \ln(\mu_i).$$

We need distribution for $Y_i \geq 0$ with $E(Y_i) = \mu_i$ and $Var(Y_i) = \sigma^2 \mu_i^2$. Such a distribution is the Gamma distribution, i.e.

$$Y \sim Gamma(\nu, \lambda) \quad \text{with} \quad f_Y(y) = \frac{\lambda}{\Gamma(\nu)} (\lambda y)^{\nu - 1} e^{-\lambda y} \qquad y \geq 0$$

$$\Rightarrow E(Y) = \tfrac{\nu}{\lambda} = \mu, \quad Var(Y) = \tfrac{\nu}{\lambda^2} = \left(\tfrac{\nu}{\lambda}\right)^2 \tfrac{1}{\nu} = \mu^2 \underbrace{\frac{1}{\nu}}_{=\sigma^2}.$$

Summary

$$Var(Y_i) = \sigma^2(E(Y_i))^2$$

$$\mu_i = E(Y_i) = e^{\beta_0} x_{i1}^{\beta_1} \cdots x_{ip}^{\beta_p}$$

Log normal model
(normal error on log scale)
Linear model on log scale

GLM with $Y_i \sim \Gamma(\nu, \lambda_i)$
$\mu_i = \frac{\nu}{\lambda_i} \quad \sigma^2 = \frac{1}{\nu}$
Gamma regression

Properties of the Gamma distribution

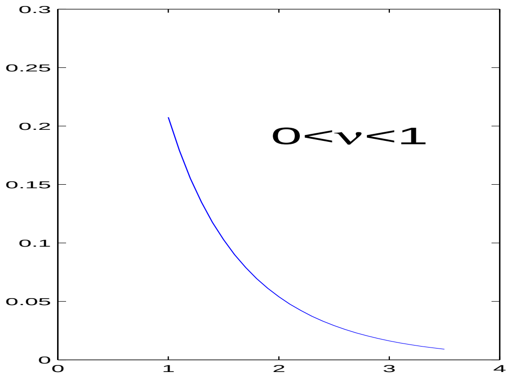
1) Mean parametrization

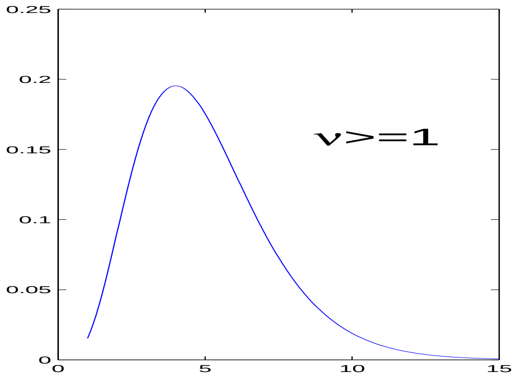
$$
\begin{aligned}
\mu &= \tfrac{\nu}{\lambda} \Rightarrow \lambda = \tfrac{\nu}{\mu} \\
\Rightarrow \qquad f_Y(y) &= \tfrac{\lambda}{\Gamma(\nu)} (\lambda y)^{\nu-1} e^{-\lambda y} \\
&= \tfrac{1}{\Gamma(\nu)} \left(\tfrac{\nu}{\mu}\right) \left(\tfrac{\nu}{\mu} y\right)^{\nu-1} e^{-\frac{\nu}{\mu} y} \\
&= \tfrac{1}{\Gamma(\nu)} \left(\tfrac{\nu y}{\mu}\right)^{\nu} e^{-\frac{\nu}{\mu} y} \, \tfrac{1}{y}
\end{aligned}
$$

$$
\Rightarrow \quad f_Y(y)dy = \underbrace{\tfrac{1}{\Gamma(\nu)} \left(\tfrac{\nu y}{\mu}\right)^{\nu} e^{-\frac{\nu}{\mu} y}}_{Gamma(\mu,\nu)} \, \underbrace{\frac{1}{y}dy}_{d(\ln y)}
$$

Mean reparametrization

2) The form of the density is determined by ν





Skewed to the right

$$\text{Skewness} = \frac{E(Y-\mu)^3}{\sigma^3} = 2\nu^{-1/2} \rightarrow 0 \text{ for } \nu \rightarrow \infty$$

3) Special cases :

$\nu = 1$	exponential distribution
$\nu \rightarrow \infty$	normal distribution

Gamma regression as GLM

For known ν in $Y \sim Gamma(\mu, \nu)$ the log likelihood is given by

$$
\begin{aligned}
l(\mu,\nu,y) &= (\nu-1)\log(y) - \frac{\nu}{\mu}y + \nu\log\nu - \nu\log\mu - \log\Gamma(\nu) \\
&= \nu\left(-\frac{y}{\mu} - \log\mu\right) + c(y,\nu) \\
\Rightarrow \theta &= -1/\mu, \qquad b(\theta) = \log\mu = \log(-1/\theta) = -\log(-\theta) \\
\Rightarrow b'(\theta) &= -\frac{(-1)}{-\theta} = -1/\theta = \mu \quad \text{expectation} \\
b''(\theta) &= \frac{1}{\theta^2} = \mu^2 \quad \text{variance function} \\
a(\phi) &= \phi \ , \quad \phi = 1/\nu \quad \text{dispersion parameter}
\end{aligned}
$$

EDA for Gamma regression

$$\mu_i = e^{\boldsymbol{x}_i^t \boldsymbol{\beta}} \quad \Rightarrow \quad \log(\mu_i) = \boldsymbol{x}_i^t \boldsymbol{\beta}$$

$\Rightarrow$ a plot of x_{ij} versus $\log(Y_i)$ should be linear

$$\mu_i = \frac{1}{\boldsymbol{x}_i^t \boldsymbol{\beta}} \quad \Rightarrow \quad \frac{1}{\mu_i} = \boldsymbol{x}_i^t \boldsymbol{\beta}$$

$\Rightarrow$ a plot of x_{ij} versus $1/Y_i$ should be linear

Example: Canadian Automobile Insurance Claims

Source: Bailey, R.A. and Simon, LeRoy J. (1960). Two studies in automobile insurance. ASTIN Bulletin, 192-217.

Description: The data give the Canadian automobile insurance experience for policy years 1956 and 1957 as of June 30, 1959. The data includes virtually every insurance company operating in Canada and was collated by the Statistical Agency (Canadian Underwriters' Association - Statistical Department) acting under instructions from the Superintendent of Insurance. The data given here is for private passenger automobile liability for non-farmers for all of Canada excluding Saskatchewan.

The variable Merit measures the number of years since the last claim on the policy. The variable Class is a collation of age, sex, use and marital status. The variables Insured and Premium are two measures of the risk exposure of the insurance companies.

Variable Description:

Merit

3	licensed and accident free $\geq$3 years
2	licensed and accident free 2 years
1	licensed and accident free 1 year
0	all others

Class

1	pleasure, no male operator < 25
2	pleasure, non-principal male operator < 25
3	business use
4	unmarried owner or principal operator< 25
5	married owner or principal operator < 25

Variable Description(continued):

Insured	Earned car years
Premium	Earned premium in 1000's (adjusted to what the premium would have been had all cars been written at 01 rates)
Claims	Number of claims
Cost	Total cost of the claim in 1000's of dollars

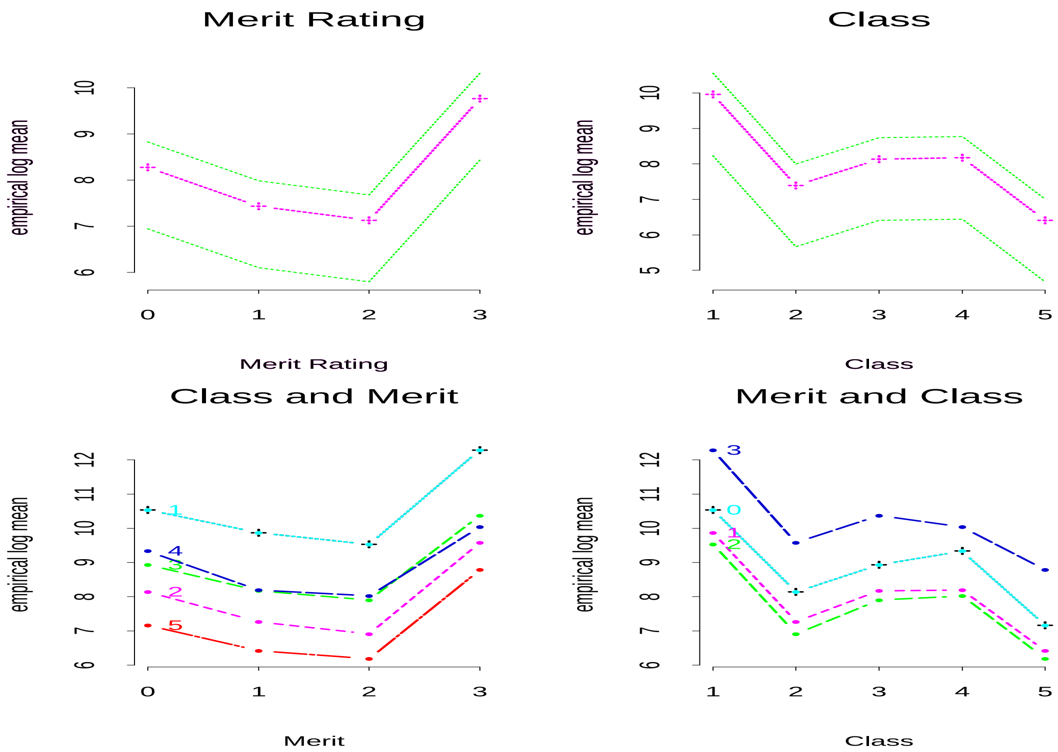
Variable of Interest is Cost.

Data:

```
> cancar.data
   Merit Class Insured Premium Claims  Cost
 1     3     1 2757520  159108 217151 63191
 2     3     2  130535    7175  14506  4598
 3     3     3  247424   15663  31964  9589
 4     3     4  156871    7694  22884  7964
 5     3     5   64130    3241   6560  1752
 6     2     1  130706    7910  13792  4055
 7     2     2    7233     431   1001   380
 8     2     3   15868    1080   2695   701
 9     2     4   17707     888   3054   983
10     2     5    4039     209    487   114
11     1     1  163544    9862  19346  5552
12     1     2    9726     572   1430   439
13     1     3   20369    1382   3546  1011
14     1     4   21089    1052   3618  1281
15     1     5    4869     250    613   178
16     0     1  273944   17226  37730 11809
17     0     2   21504    1207   3421  1088
18     0     3   37666    2502   7565  2383
19     0     4   56730    2756  11345  3971
20     0     5    8601     461   1291   382
```

Exploratory Data Analysis (unweighted case):

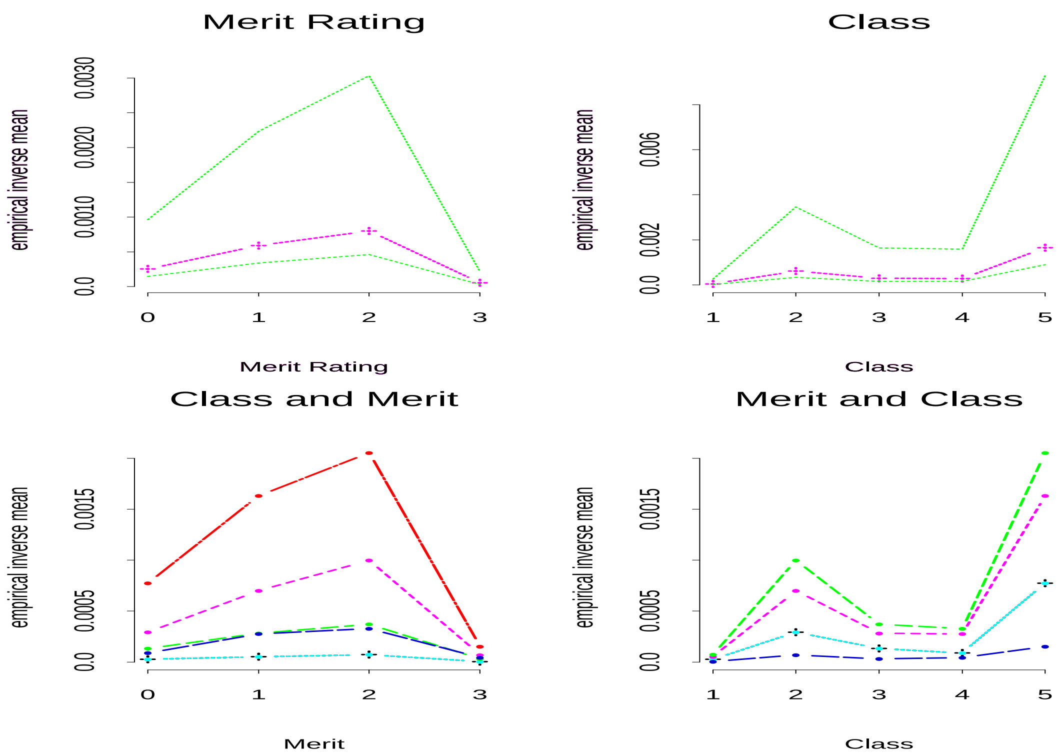
Log Link:



No obvious functional relationships, treatment of covariates as factors might be appropriate. No strong interactions expected.

Exploratory Data Analysis (continued):

Inverse Link:



No obvious functional relationships, treatment of covariates as factors might be appropriate. Some interactions expected.

Residual deviance in Gamma regression

$Y_i \sim Gamma(\mu_i, \nu)$ independent $\qquad \mu_i = e^{\boldsymbol{x}_i^t \boldsymbol{\beta}}$

$$l(\boldsymbol{\mu}, \nu, \mathbf{y}) = \sum_{i=1}^{n} \left\{ \nu \left[-\frac{y_i}{\mu_i} - \ln(\mu_i) \right] - \ln(\Gamma(\nu)) + \nu \ln(\nu y_i) - \ln(y_i) \right\}$$

$$\Rightarrow \quad l(\mathbf{y}, \nu, \mathbf{y}) = \sum_{i=1}^{n} \{ \nu [-1 - \ln(y_i)] - \ln(\Gamma(\nu)) + \nu \ln(\nu y_i) - \ln(y_i) \}$$

– saturated log likelihood

$$\Rightarrow \quad -2 \left(l(\hat{\boldsymbol{\mu}}, \nu, \mathbf{y}) - l(\mathbf{y}, \nu, \mathbf{y}) \right)$$
$$= 2 \left[\sum_{i=1}^{n} \nu \left(-1 - \ln(y_i) + \frac{y_i}{\hat{\mu}_i} + \ln(\hat{\mu}_i) \right) \right]$$
$$= -2 \sum_{i=1}^{n} \nu \left[\ln\left(\frac{y_i}{\hat{\mu}_i} \right) - \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$$

$$\Rightarrow \quad D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = -2 \sum_{i=1}^{n} \left[\ln\left(\frac{y_i}{\hat{\mu}_i} \right) - \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$$ deviance

We have $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) \overset{a}{\sim} \phi \chi^2_{n-p}$ where $\phi = 1/\nu$.

Residual deviance test

$Y_i \sim Gamma(\mu_i, \nu)$ independent $\mu_i = e^{\boldsymbol{x}_i^t \boldsymbol{\beta}}$ (Model (*))

Reject model (*) at level α if $\frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\hat{\phi}} > \chi^2_{n-p,1-\alpha}$

Here $\hat{\phi}$ is an estimate of ϕ.

Partial deviance test

Consider $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t)^t \quad \boldsymbol{\beta}_1 \in \mathbb{R}^{p_1},\ \boldsymbol{\beta}_2 \in \mathbb{R}^{p_2}$

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_1 &= \text{fitted mean in reduced model with design } X_2 \\
\hat{\boldsymbol{\mu}}_2 &= \text{fitted mean in full model with design } X_1 \text{ and } X_2 \\
\hat{\phi}_1 &= \text{estimated dispersion parameter in reduced model} \\
\hat{\phi}_2 &= \text{estimated dispersion parameter in full model}
\end{aligned}
$$

The partial deviance test for $H_0: \quad \boldsymbol{\beta}_1 = \mathbf{0}$ versus $H_1: \quad \boldsymbol{\beta}_1 \neq \mathbf{0}$ is given by:

Reject H_0 at level $\alpha \Leftrightarrow$

$$\frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}}_1)}{\hat{\phi}_1} - \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}}_2)}{\hat{\phi}_2} > \chi^2_{p_1, 1-\alpha}$$

Canonical link

$$\eta_i = \theta_i = -\frac{1}{\mu_i}$$

Since we need $\mu_i > 0$ we need $\eta_i < 0$, which gives restrictions on $\boldsymbol{\beta}$. Therefore the canonical link is not often used. Most often the log link is used.

Estimation of the dispersion parameter

For the estimation of the dispersion parameter $\phi = \sigma^2 = 1/\nu$ method of moments is used

$$E(Y_i) = \mu_i \quad Var(Y_i) = \phi\mu_i^2 \quad \forall i$$

$$\Rightarrow \quad Z_i = \frac{Y_i}{\mu_i}$$ has expectation 1, variance ϕ and is i.i.d.

$$\Rightarrow \quad \hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}\left(\frac{Y_i}{\hat{\mu}_i} - 1\right)^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left(\frac{Y_i-\hat{\mu}_i}{\hat{\mu}_i}\right)^2$$

since p parameters need to be estimated.

Remark: Pearson χ^2 statistic for the gamma regression is given by

$$\chi_P^2 = \sum_{i=1}^{n}\frac{(Y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)} \qquad V(\hat{\mu}_i) = \hat{\mu}_i^2$$

$$= \sum_{i=1}^{n}\left(\frac{Y_i-\hat{\mu}_i}{\hat{\mu}_i}\right)^2$$

$\Rightarrow \hat{\phi} = \chi_P^2/(n-p)$.

Relationship between gamma regression with log link and a linear model on the log scala

For $\sigma^2 = \phi$ small we have $Var(\log(Y_i)) \approx \sigma^2$, therefore we can use

$$\log Y_i = \boldsymbol{x_i^t \alpha} + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2) \; i.i.d. \qquad \text{linear model on log scala}$$

$$\Rightarrow Cov(\hat{\boldsymbol{\alpha}}) = \sigma^2 (X^t X)^{-1}.$$

For the gamma regression with $Y_i \sim \Gamma(\mu_i, 1/\sigma^2)$ and $\mu_i = exp\{\boldsymbol{x_i^t \beta}\}$ we have

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} N(\boldsymbol{\beta}, I(\hat{\boldsymbol{\beta}})^{-1}), \quad \text{where}$$

$$I(\hat{\boldsymbol{\beta}})) = \text{Fisher information} = \left(-E\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^t}\right)\right)\Big|_{\beta=\hat{\beta}}.$$

Here $$l(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(-\frac{y}{\mu_i} - \log\mu_i + const\right)$$

$$\Rightarrow \quad \frac{\partial l(\beta)}{\partial\beta_j} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i x_{ij}}{e^{\mathbf{x}_i^t\beta}} - x_{ij}\right) \qquad j = 1,\dots,p$$

$$\Rightarrow \quad \frac{\partial^2 l(\beta)}{\partial\beta_s\partial\beta_j} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(-\frac{y_i x_{is} x_{ij}}{e^{\mathbf{x}_i^t\beta}}\right)$$

$$\Rightarrow \quad -E\left(\frac{\partial^2 l(\beta)}{\partial\beta_s\partial\beta_j}\right) = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_{is}x_{ij} (\text{independent of } \boldsymbol{\beta})$$

$$\Rightarrow \quad I(\hat{\boldsymbol{\beta}}) = \frac{1}{\sigma^2}X^tX$$

$$\Rightarrow \quad Cov(\hat{\boldsymbol{\beta}}) \approx \sigma^2(X^tX)^{-1} = Cov(\hat{\boldsymbol{\alpha}})$$

Distinction between both models is difficult. If X corresponds to an orthogonal design, i.e. $(X^tX)^{-1} = I_p$, we have that the components of $\hat{\boldsymbol{\alpha}}$ ($\hat{\boldsymbol{\beta}}$) are (asymptotically) independent.

Gamma regression with weights

Example: Y_i = total claim size arising from n_i claims in group i with covariates $\mathbf{x_i}$.

$$\Rightarrow \quad Y_i = \sum_{j=1}^{n_i} Y_{ij}; \qquad Y_{ij} = j^{th} \text{ claim in group } i$$

$$Y_i^s := Y_i/n_i = \text{ average claim size in group } i$$

If $Y_{ij} \sim Gamma(\mu_i, \nu)$, independent, we have

$$E(Y_i^s) = \frac{1}{n_i} \sum_{j=1}^{n_i} \underbrace{E(Y_{ij})}_{=\mu_i} = \mu_i$$

$$Var(Y_i^s) = \frac{1}{n_i} Var(Y_{i1}) = \frac{1}{n_i} \mu_i^2/\nu = \mu_i^2/n_i\nu.$$

Since $n_i Y_i^s = \sum_{j=1}^{n_i} Y_{ij} \sim Gamma(n_i \mu_i, n_i \nu)$

we have $Y_i^s \sim Gamma(\mu_i, n_i \nu)$. This follows since

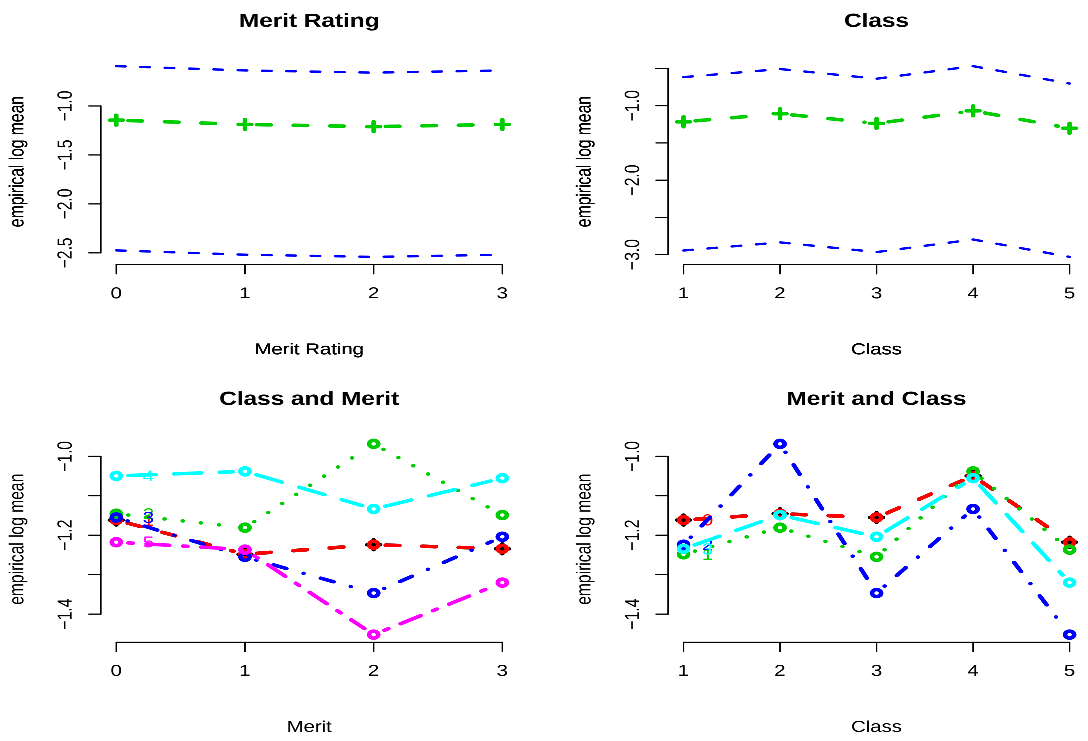
$$m_{Y_i^s}(t) := E\left(e^{tY_i^s}\right) = E(e^{\frac{t n_i Y_i^s}{n_i}}) = \frac{1}{(1 - \frac{n_i \mu_i}{n_i \nu} \frac{t}{n_i})^{n_i \nu}}$$

$$= \frac{1}{(1 - \frac{\mu_i t}{n_i \nu})^{n_i \nu}} = m_X(t) \quad \forall t, \quad \text{where } X \sim \text{Gamma}(\mu_i, n_i \nu).$$

Therefore we need to introduce weights for a gamma regression for Y_i^s. Since $E(Y_i^s) = \mu_i$, the EDA for a weighted gamma regression does not change compared to an unweighted one.

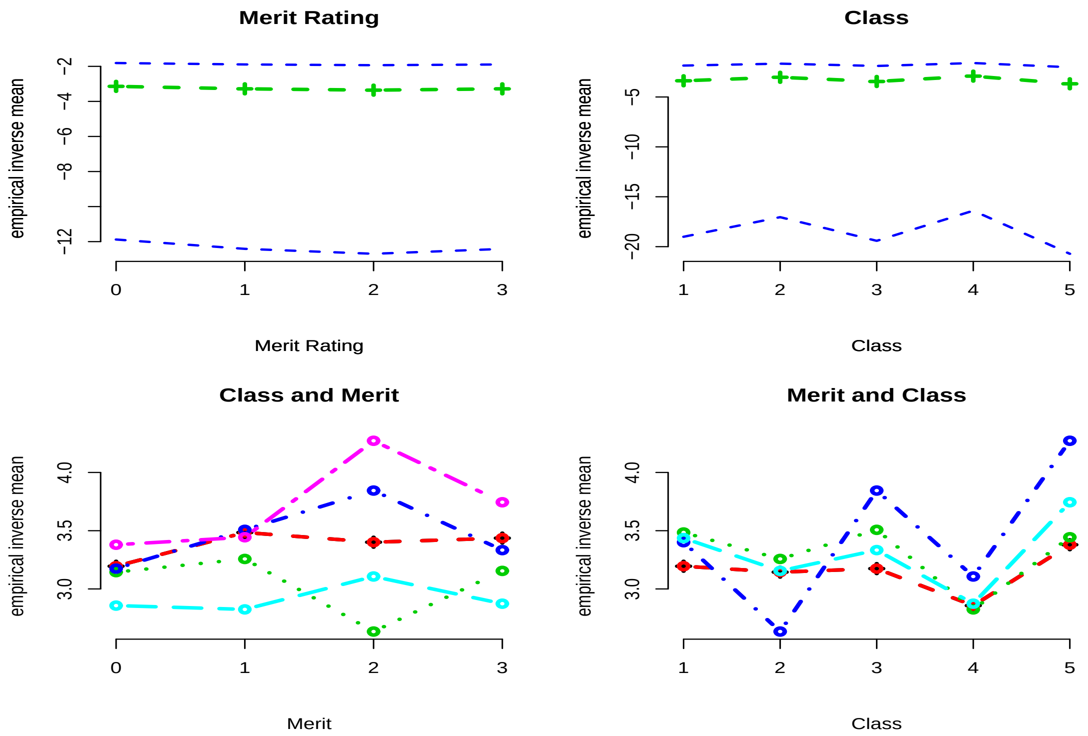
Exploratory Data Analysis (weighted case):

Log Link:



Exploratory Data Analysis (weighted case):

Inverse Link:



For estimation of the dispersion parameter $\sigma^2 = \phi = 1/\nu$ in a weighted model we use

$$\hat{\sigma}^2 = \hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} n_i \left(\frac{Y_i - \hat{\mu}_i}{\hat{\mu}_i} \right)^2$$

Residuals: $E(Y_i^s) = \mu_i \quad Var(Y_i^s) = \frac{\mu_i^2}{n_i \nu} = \mu_i^2 \frac{\sigma^2}{n_i}$

$\Rightarrow \quad \frac{Y_i^s - \mu_i}{\frac{\sigma}{\sqrt{n_i}} \mu_i} = \frac{\sqrt{n_i}}{\sigma} \left(\frac{Y_i^s - \mu_i}{\mu_i} \right)$ has zero expectation and unit variance

i.e. standardized residuals can be defined by

$$r_i^s := \frac{\sqrt{n_i}}{\sigma} \left(\frac{Y_i^s - \hat{\mu}_i}{\hat{\mu}_i} \right).$$

Example: Canadian Automobile Insurance Claims: Gamma Regression:

Main Effects with Log Link

```
> f.gamma.main_Cost/Claims ~ Merit + Class
> r.gamma.log.main_glm(f.gamma.main, family =
Gamma(link = "log"), weights = Claims)
```

Example: Gamma Regression:

```
> summary(r.gamma.log.main,cor=F)
Call: glm(formula = Cost/Claims ~ Merit + Class,
 family= Gamma(link = "log"), weights = Claims)

Coefficients:
             Value Std. Error t value
(Intercept) -1.175      0.016  -75.58
     Merit1 -0.069      0.026   -2.63
     Merit2 -0.070      0.029   -2.41
     Merit3 -0.057      0.016   -3.48
     Class2  0.083      0.026    3.13
     Class3  0.016      0.018    0.86
     Class4  0.160      0.019    8.23
     Class5 -0.081      0.039   -2.08

(Dispersion Parameter for Gamma family
 taken to be 13 )
    Null Deviance:1556 on 19 degrees of freedom
Residual Deviance:157 on 12 degrees of freedom
```

Example: Gamma Regression (continued):

To check the dispersion estimate, use the Pearson Chi Square Statistics.

```
> sum(resid(r.gamma.log.main,type="pearson")^2)
[1] 159
> 159/12
[1] 13
```

For the residual deviance test, we need to scale the deviance

```
> 1-pchisq(156/13,12)
[1] 0.45
```

A p-value of .45 shows no lack of fit.

Main Effects with Inverse Link

```
> f.gamma.main_Cost/Claims ~ Merit + Class
> r.gamma.inverse.main_glm(f.gamma.main,
        family = Gamma, weights = Claims)
> summary(r.gamma.inverse.main,cor=F)
Coefficients:
             Value Std. Error t value
(Intercept)  3.247      0.051   63.72
     Merit1  0.215      0.089    2.42
     Merit2  0.224      0.099    2.25
     Merit3  0.177      0.053    3.32
     Class2 -0.268      0.086   -3.12
     Class3 -0.054      0.063   -0.85
     Class4 -0.498      0.059   -8.39
     Class5  0.287      0.149    1.93
(Dispersion Parameter for Gamma family taken to be 14 )
    Null Deviance:1556 on 19 degrees of freedom
Residual Deviance:167 on 12 degrees of freedom
```

Example: Gamma Regression (continued):

For the residual deviance test, we need to scale the deviance

```
> 1-pchisq(167/14,12)
[1] 0.45
```

A p-value of .45 shows no lack of fit.

Log Link With Interaction

```
> f.gamma.inter_Cost/Claims ~ Merit * Class
> r.gamma.inter_glm(f.gamma.inter, family
         = Gamma, weights = Claims))
```

Example: Gamma Regression (continued):

```
> summary(r.gamma.log.inter,cor=F)

Coefficients:
              Value Std. Error t value
 (Intercept)  3.195         NA      NA
      Merit1  0.289         NA      NA
      Merit2  0.206         NA      NA
      Merit3  0.241         NA      NA
      Class2 -0.051         NA      NA
      Class3 -0.020         NA      NA
      Class4 -0.338         NA      NA
      Class5  0.185         NA      NA
Merit1Class2 -0.176         NA      NA
Merit2Class2 -0.716         NA      NA
Merit3Class2 -0.231         NA      NA
Merit1Class3  0.043         NA      NA
Merit2Class3  0.464         NA      NA
Merit3Class3 -0.083         NA      NA
Merit1Class4 -0.322         NA      NA
Merit2Class4  0.044         NA      NA
Merit3Class4 -0.225         NA      NA
Merit1Class5 -0.225         NA      NA
Merit2Class5  0.686         NA      NA
Merit3Class5  0.123         NA      NA

(Dispersion Parameter for Gamma family taken
 to be NA )
```

Example: Gamma Regression (continued):

```
    Null Deviance: 1556 on 19 degrees of freedom

Residual Deviance:  0 on 0 degrees of freedom

Number of Fisher Scoring Iterations: 1
```

This model is the saturated model since only one observation per cell. Therefore the dispersion estimate can no longer be estimated, since the Pearson Chi Square Statistics is 0. To fit a model with fixed dispersion, use

```
> summary(r.gamma.log.inter,dispersion=1,cor=F)
```

Example: Gamma Regression (continued):

```
Coefficients:
                Value Std. Error t value
 (Intercept)  3.195      0.016  194.24
      Merit1  0.289      0.030    9.66
      Merit2  0.206      0.033    6.19
      Merit3  0.241      0.018   13.39
      Class2 -0.051      0.056   -0.90
      Class3 -0.020      0.040   -0.51
      Class4 -0.338      0.031  -10.74
      Class5  0.185      0.095    1.93
Merit1Class2 -0.176      0.106   -1.67
Merit2Class2 -0.716      0.105   -6.85
Merit3Class2 -0.231      0.062   -3.70
Merit1Class3  0.043      0.075    0.57
Merit2Class3  0.464      0.089    5.21
Merit3Class3 -0.083      0.045   -1.84
Merit1Class4 -0.322      0.062   -5.21
Merit2Class4  0.044      0.071    0.62
Merit3Class4 -0.225      0.037   -6.00
Merit1Class5 -0.225      0.171   -1.32
Merit2Class5  0.686      0.218    3.15
Merit3Class5  0.123      0.106    1.16

(Dispersion Parameter for Gamma family taken to be 1 )
    Null Deviance: 1556 on 19 degrees of freedom
Residual Deviance:  0 on 0 degrees of freedom
```

From this we see that certain interaction effects are strongly significant.

Log Link With Some Interaction Terms

```
> f.gamma.inter3_Cost/Claims ~ Merit + Class +
    I((Merit == 2) * (
    Class == 2)) + I((Merit == 2) * (Class ==
    3)) + I((Merit == 1) * (Class == 4)) + I((
    Merit == 3) * (Class == 4)) + I((Merit ==
    3) * (Class == 2)) + I((Merit == 2) * (
    Class == 5))
> r.gamma.log.inter3_glm(f.gamma.inter3,family
=Gamma(link="log"),weights=Claims)
```

Example: Gamma Regression (continued):

```
> summary(r.gamma.log.inter3,cor=F)
Call:glm(formula=Cost/Claims ~ Merit+Class+I((
     Merit == 2) * (Class == 2)) + I((Merit ==
     2) * (Class == 3)) + I((Merit == 1) * (
     Class == 4)) + I((Merit == 3) * (Class ==
     4)) + I((Merit == 3) * (Class == 2)) + I((
     Merit == 2) * (Class == 5)), family =
     Gamma(link = "log"), weights = Claims)
Deviance Residuals:
  Min    1Q   Median   3Q Max
 -1.8 -0.43 -2.4e-07 0.44 2.2
```

Example: Gamma Regression (continued):

Coefficients:

	Value	Std. Error
(Intercept)	-1.164	0.0076
Merit1	-0.084	0.0127
Merit2	-0.064	0.0145
Merit3	-0.070	0.0081
Class2	0.033	0.0245
Class3	0.023	0.0086
Class4	0.110	0.0153
Class5	-0.074	0.0182
I((Merit==2)*(Class==2))	0.226	0.0590
I((Merit==2)*(Class==3))	-0.142	0.0354
I((Merit==1)*(Class==4))	0.100	0.0331
I((Merit==3)*(Class==4))	0.068	0.0191
I((Merit==3)*(Class==2))	0.052	0.0283
I((Merit==2)*(Class==5))	-0.150	0.0781

Example: Gamma Regression (continued):

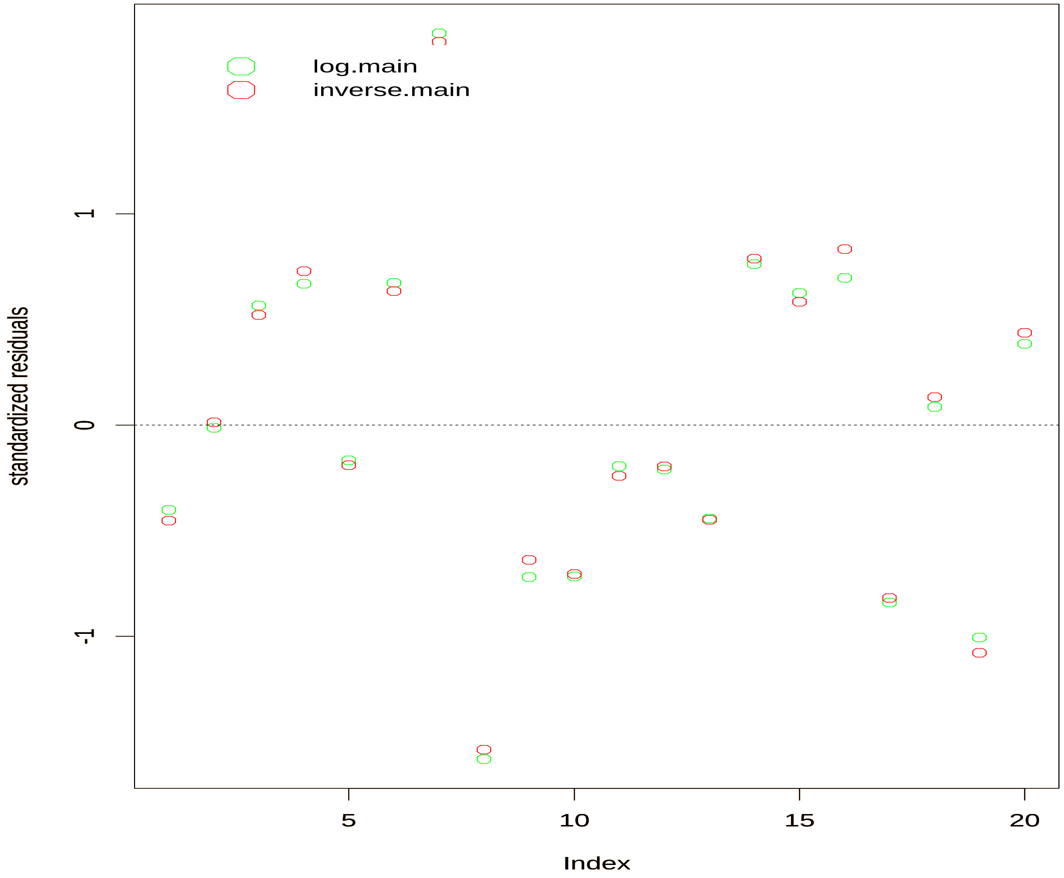
```
                                t value
              (Intercept)      -154.1
                   Merit1        -6.6
                   Merit2        -4.4
                   Merit3        -8.6
                   Class2         1.3
                   Class3         2.7
                   Class4         7.2
                   Class5        -4.1
I((Merit == 2) * (Class == 2))    3.8
I((Merit == 2) * (Class == 3))   -4.0
I((Merit == 1) * (Class == 4))    3.0
I((Merit == 3) * (Class == 4))    3.6
I((Merit == 3) * (Class == 2))    1.8
I((Merit == 2) * (Class == 5))   -1.9

(Dispersion Parameter for Gamma family taken to be 2.7 )
    Null Deviance: 1556 on 19 degrees of freedom
Residual Deviance: 16 on 6 degrees of freedom
```

Example: Gamma Regression (continued):

```
> 1-pchisq(16/2.7,6)
[1] 0.43  # Residual Deviance Test
> 1-pchisq((156/13)-(16/2.7),6)
[1] 0.41   # Partial Deviance Test
```

We see that the joint effect of the interaction terms are not significant, since the partial deviance test yields a p-value of .41. A partial deviance test when only I((Merit == 2) * (Class == 2)) and I((Merit == 2) * (Class == 3)) are included gives a p-value of .39, thus showing that the interaction effects are not present in this data set. The same results are true when an inverse link is used instead.

Standardized Residual Plots::



Difference between log link and inverse link is minimal.

What needs to be done if one wants to use the linear model on the log scale?

$Var(Y_i^s) = \frac{\sigma^2}{n_i}\mu_i^2$

$$
\begin{aligned}
\Rightarrow \qquad E(\log(Y_i^s)) &\approx E\left(\log(\mu_i) + \frac{1}{\mu_i}(Y_i^s - \mu_i) - \frac{1}{\mu_i^2}(Y_i^s - \mu_i)^2\right) \\
&= \log(\mu_i) - \frac{1}{\mu_i^2}\underbrace{Var(Y_i^s)}_{\mu_i^2\frac{\sigma^2}{n_i}} = \log(\mu_i) - \frac{\sigma^2}{2n_i} \\
E\left[(\log(Y_i^s))^2\right] &\approx E\left([\log(\mu_i) + \frac{1}{\mu_i}(Y_i^s - \mu_i)]^2\right) \\
&= [\log(\mu_i)]^2 + \frac{Var(Y_i^s)}{\mu_i^2} \\
&= (\log \mu_i)^2 + \frac{\sigma^2}{n_i}
\end{aligned}
$$

$$\Rightarrow \qquad Var(\log(Y_i^s)) \approx (\log \mu_i)^2 + \frac{\sigma^2}{n_i} - (\log \mu_i)^2 = \frac{\sigma^2}{n_i} \quad \text{for } \sigma^2 \text{ small}$$

$\Rightarrow$ linear model on the log scale needs also to be weighted

Log Normal Models:

```
> f.lognormal.main_log(Cost/Claims)~Merit + Class
> r.lognormal.main_glm(f.lognormal.main,
weights=Claims)
> summary(r.lognormal.main,cor=F)

Call: glm(formula = log(Cost/Claims) ~ Merit +
        Class, weights = Claims)
Deviance Residuals:
 Min   1Q Median  3Q Max
  -6 -1.8  -0.24 2.3 6.2
```

Log Normal Models (continued):

```
Coefficients:
              Value Std. Error t value
(Intercept) -1.175      0.015  -76.33
     Merit1 -0.069      0.026   -2.66
     Merit2 -0.072      0.029   -2.51
     Merit3 -0.057      0.016   -3.50
     Class2  0.082      0.026    3.12
     Class3  0.015      0.018    0.85
     Class4  0.160      0.019    8.29
     Class5 -0.082      0.039   -2.12

(Dispersion Parameter for Gaussian family taken to be 13 )

    Null Deviance:1515 on 19 degrees of freedom

Residual Deviance:156 on 12 degrees of freedom
```

Log Normal Models (continued):

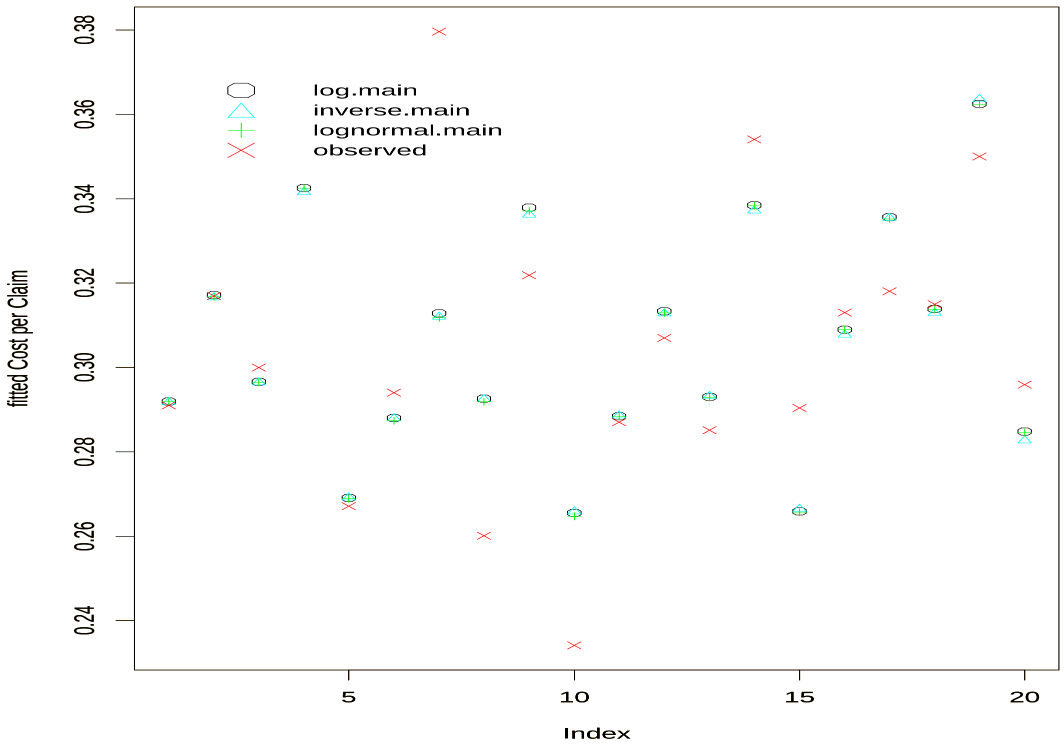
```
> 1-pchisq(156/13,12)
[1] 0.45  # Residual Deviance Test
# Results when Splus function lm() is used:
> summary(lm(f.lognormal.main,weights=Claims),cor=F)
Coefficients:
               Value Std. Error t value
(Intercept)  -1.175   0.015    -76.327
     Merit1  -0.069   0.026     -2.658
     Merit2  -0.072   0.029     -2.508
     Merit3  -0.057   0.016     -3.501
     Class2   0.082   0.026      3.117
     Class3   0.015   0.018      0.849
     Class4   0.160   0.019      8.292
     Class5  -0.082   0.039     -2.124
            Pr(>|t|)
(Intercept)   0.000
     Merit1   0.021
     Merit2   0.028
     Merit3   0.004
     Class2   0.009
     Class3   0.412
     Class4   0.000
     Class5   0.055
```

Log Normal Models (continued):

```
Residual standard error: 3.6 on 12
degrees of freedom
Multiple R-Squared: 0.9
F-statistic: 15 on 7 and 12 degrees of freedom,
the p-value is 4.7e-05
```

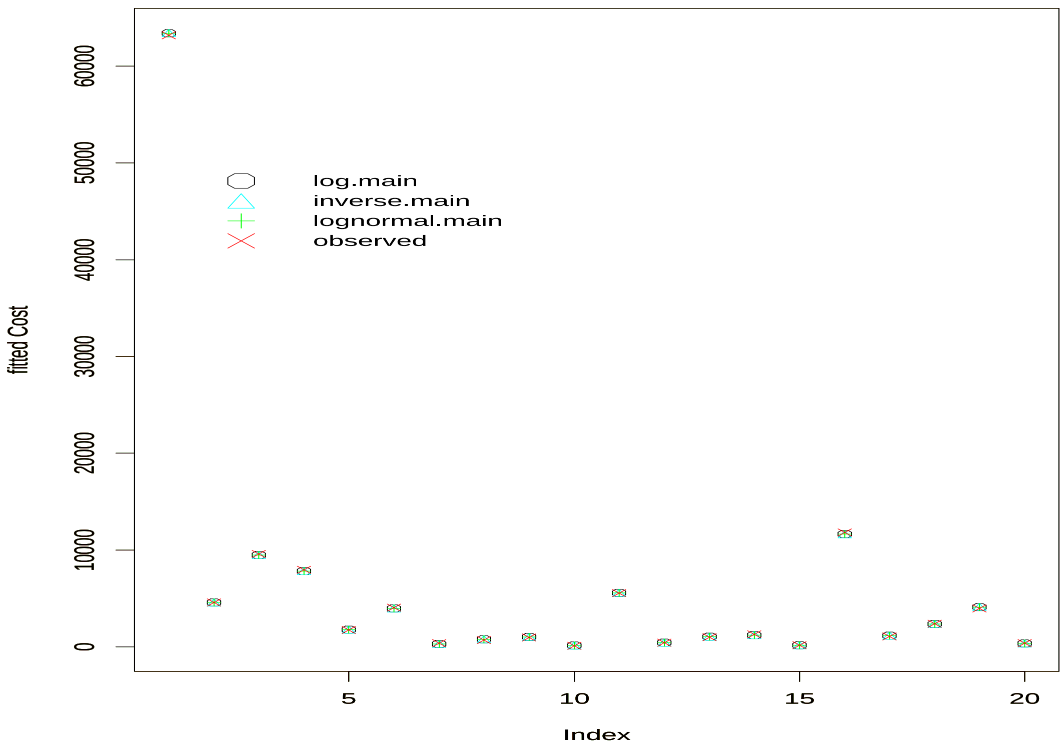
Both approaches (R-squared=.9, residual deviance p-value=.45) show no lack of fit.

Fitted Costs per Claims:



Difference between Models are minimal.

Fitted Cost:



Difference between Models are minimal.