

Modelos Lineares Generalizados Duplos

Pedro Henrique Corrêa de Almeida *

Gustavo Almeida Silva[†]

Primavera de 2023

RESUMO

Esse artigo teve como objetivo estudar profundamente os Modelos Lineares Generalizados Duplos. Tais modelos são utilizados para a modelagem dos parâmetros de locação e escala em um contexto de distribuições da família exponencial, sendo uma extensão dos Modelos Lineares Generalizados. O trabalho explicitou os tópicos de estimação, análise de resíduo e diagnósticos. Além disso, um estudo de simulação e uma aplicação em dados reais de tais modelos foi proposta, onde foi possível observar um considerável ganho de precisão ao se modelar os parâmetros de locação e escala (...)

Palavras chave: Modelos Lineares Generalizados, Modelos Lineares Generalizados Duplos, MLG, Generalized Linear Models, Double Generalized Linear Models, DGLM, GAMLSS, Location and Escala Models

*Universidade Federal de Juiz de Fora

[†]Universidade Federal de Juiz de Fora

1 Introdução

A análise de regressão é uma técnica estatística essencial utilizada para investigar e modelar a relação entre variáveis. Seu propósito principal é desenvolver modelos matemáticos que descrevem a dependência de uma variável dependente em relação a uma ou mais variáveis independentes. Amplamente aplicável em áreas como finanças, biologia e engenharia.

Os modelos lineares usuais assumem que sua parte aleatória (erros) possuem distribuição Normal, com parâmetros de locação 0 e escala constante, tal suposição se mostra robusta em diversos contextos, onde assumir tal distribuição facilita interpretações do modelo, testes de significância e certas propriedades exclusivas de uma distribuição Gaussiana. Porém, em contextos como dados de contagem, binários, e limitados na reta real, utilizar tal modelo se mostra inviável e errôneo. Assim, os MLGs surgem como ferramenta para esses casos

Uma das limitações na análise de regressão clássica ocorre quando os dados, devido a natureza destes, apresentam heterogeneidade da variância, fato que vai contra o pressuposto de homogeneidade. Nesse sentido, na busca de lidar com esse tipo de problema, sem necessitar de transformações, modelos de média e dispersão surgem, chamados também de Modelos Lineares Generalizados Duplos, uma extensão dos MLGs. (PAULA, 2013)

2 Especificação do Modelo

Os Modelos Lineares Generalizados (MLG) foram uma teoria de unificação de técnicas estatísticas proposta no artigo (NELDER; WEDDERBURN, 1972) e depois sendo aperfeiçoado em vários livros e artigos.

Considerando uma amostra com n observações independentes, X uma matriz com $p+1$ colunas e \mathbf{y} um vetor de observações amostrado de Y , são definidos os 3 componentes de um MLG

- Y tem distribuição probabilística como membro da Família Exponencial de distribuições, com uma função de probabilidade ou função densidade de probabilidade (para variáveis aleatórias discretas e contínuas, respectivamente)

$$f(y; \theta_i, \Phi) = \exp[\Phi y \theta_i - b(\theta_i) + c(y, \Phi)]$$

,

$$E(y_i) = \mu_i = b$$

$$Var(y_i) = a_i(\phi) V_i$$

- A matriz X de covariáveis relacionadas no chamado preditor linear na forma

$$\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}$$

- Uma Função de Ligação monótona (invertível) e diferenciável $g(\cdot)$, que liga o preditor linear η_i à média de Y onde escrevemos

$$g(\mu_i) = \eta_i$$

(SMYTH, 1989) introduziu os modelos lineares generalizados duplos com modelagem conjunta da média e do parâmetro de precisão (ou dispersão) e desenvolveu um processo de estimação baseado no método de máxima verossimilhança

No caso dos MLGs duplos, são definidos mais 2 componentes

- A matriz \mathbf{Z} de covariáveis relacionadas no chamado preditor linear na forma

$$\lambda_i = \mathbf{Z}_i^T \boldsymbol{\gamma}$$

- Uma Função de Ligação monótona (invertível) e diferenciável $h(\cdot)$, que liga o preditor linear λ ao parâmetro de dispersão ϕ onde escrevemos

$$h(\phi) = \lambda(X_i)$$

Com isso, a função de densidade de probabilidades do modelo será dada por

$$f(y; \theta_i, \phi_i) = \exp[\phi_i y \theta_i b(\theta_i) + c(y, \theta_i)]$$

Onde $c(y, \phi_i) = d(\phi_i) + \phi_i a(y) + u(y)$.

Nesse sentido, temos o logaritmo da função de verossimilhança

$$\begin{aligned} L(\theta) &= \sum_i = 1^n [\phi_i y_i \theta_i b(\theta_i) + d(\phi_i) + \phi_i a(y_i) + u(y_i)] \\ &= \sum_i = 1^n \phi_i t_i + d(\phi_i) + u(y_i), \end{aligned}$$

em que $t_i = y_i \theta_i b(\theta_i) + a(y_i)$. Logo, com $\theta_1 = \theta_2 = \theta_3 \dots = \theta_n$, é equivalente a log-verossimilhança de um modelo da família exponencial com respostas independentes T_1, \dots, T_n . Pelas propriedades da família exponencial

$$\mu_{T_i} = E(T_i) = d'(\phi_i)$$

e

$$Var(T_i) = d''(\phi_i).$$

3 Estimação

Como apresentado em (PAULA, 2013), a função score e a matriz informação de Fisher para $\boldsymbol{\beta}$ podem ser obtidas calculando as derivadas da função de verossimilhança, da seguinte forma

$$\delta L(\boldsymbol{\theta})/\delta\beta_j = \sum_{i=1}^n \left\{ \frac{\omega_i}{V_i} (y_i - \mu_i) x_{ij} \right\}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2$. Dessa forma, temos a função escore dada por

$$U_\beta(\boldsymbol{\beta}) = \frac{\delta L(\boldsymbol{\theta})}{\delta\boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$$

em que \mathbf{X} é uma matriz $n \times p$ de posto completo, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ é a matriz de pesos, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\mathbf{y} = \{y_1, \dots, y_n\}$ e $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_n\}$.

$$\begin{aligned} \partial^2 L(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell &= \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d^2\theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell} \\ &+ \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d\theta_i}{d\mu_i} \frac{d^2\mu_i}{d\eta_i^2} x_{ij}x_{i\ell} - \phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell}, \end{aligned}$$

dessa forma, obtemos o valor esperado

$$\begin{aligned} E \left\{ \partial^2 L(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell \right\} &= -\phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell} \\ &= -\phi \sum_{i=1}^n \frac{(d\mu_i/d\eta_i)^2}{V_i} x_{ij}x_{i\ell} \\ &= -\phi \sum_{i=1}^n \omega_i x_{ij}x_{i\ell}. \end{aligned}$$

escrevendo na forma matricial, concluímos

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = E \left\{ -\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \right\} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X}$$

Agora, deve-se realizar o mesmo procedimento na obtenção dos estimadores de máxima verossimilhança para $\boldsymbol{\gamma}$, em que

$$\begin{aligned} \partial L(\boldsymbol{\theta})/\partial\gamma_j &= \sum_{i=1}^n \left\{ \frac{d\phi_i}{d\lambda_i} \frac{\partial\lambda_i}{\partial\gamma_j} t_i + d'(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial\lambda_i}{\partial\gamma_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{1}{h'(\phi_i)} z_{ij} t_i + d'(\phi_i) \frac{1}{h'(\phi_i)} z_{ij} \right\} \\ &= \sum_{i=1}^n \frac{z_{ij}}{h'(\phi_i)} \{ t_i + d'(\phi_i) \} \end{aligned}$$

em que $h'(\phi_i) = d\lambda_i/d\phi_i$. Portanto, em forma matricial obtemos

$$\mathbf{U}_\gamma = \mathbf{Z}^T \mathbf{H}_\gamma^{-1} (\mathbf{t} - \boldsymbol{\mu}_T),$$

em que $\mathbf{H}_\gamma = \text{diag} \{h'(\phi_1), \dots, h'(\phi_n)\}$, $\mathbf{t} = (t_1, \dots, t_n)^T$ e $\boldsymbol{\mu}_T = (E(T_1), \dots, E(T_n))^T = (-d'(\phi_1), \dots, -d'(\phi_n))^T$.

Na obtenção da matriz informação de Fisher para $\boldsymbol{\gamma}$, vamos calcular as derivadas segundas da função log verossimilhança

$$\begin{aligned} \partial^2 L(\boldsymbol{\theta}) / \partial \gamma_j \partial \gamma_\ell &= - \sum_{i=1}^n \frac{z_{ij}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) h(\phi_i) \frac{d\phi_i}{d\lambda_i} z_{i\ell} - h''(\phi_i) \{t_i + d'(\phi_i)\} \frac{d\phi_i}{d\lambda_i} z_{i\ell} \right] \\ &= - \sum_{i=1}^n \frac{z_{ij} z_{i\ell}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) - \frac{h''(\phi_i)}{h'(\phi_i)} \{t_i + d'(\phi_i)\} \right] \end{aligned}$$

e cujos valores esperados ficam dados por

$$E \left\{ - \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} \right\} = - \sum_{i=1}^n \frac{d''(\phi_i)}{\{h'(\phi_i)\}^2} z_{ij} z_{i\ell}.$$

Em forma matricial obtemos

$$\mathbf{K}_{\gamma\gamma} = \mathbf{Z}^T \mathbf{P} \mathbf{Z}$$

em que $\mathbf{P} = \mathbf{V}_\gamma \mathbf{H}_\gamma^{-2}$, $\mathbf{V}_\gamma = \text{diag} \{-d''(\phi_1), \dots, -d''(\phi_n)\}$. Devido à ortogonalidade entre os parâmetros θ_i e ϕ_i , segue diretamente a ortogonalidade entre $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$. Assim, a matriz de informação de Fisher para $\boldsymbol{\theta}$ é dado por $\mathbf{K}_{\theta\theta} = \text{diag} \{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\gamma\gamma}\}$.

Assim como na teoria dos modelos lineares generalizados, podemos estabelecer um método iterativo para encontrar $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$. De acordo com (PAULA, 2013), é possível atualizar essas estimativas a partir das expressões

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \left(\mathbf{X}^T \boldsymbol{\Phi}^{(m)} \mathbf{W}^{(m)} \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\Phi}^{(m)} \mathbf{W}^{(m)} \mathbf{y}^{*(m)} \\ \boldsymbol{\gamma}^{(m+1)} &= \left(\mathbf{Z}^T \mathbf{P}^{(m)} \mathbf{Z} \right)^{-1} \mathbf{Z}^T \mathbf{P}^{(m)} \mathbf{z}^{*(m)} \end{aligned}$$

em que $\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$, $\mathbf{z}^* = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{V}_\gamma^{-1} \mathbf{H}_\gamma (\mathbf{t} - \boldsymbol{\mu}_T)$ e $m = 0, 1, 2, \dots$

Conforme (SMYTH, 1989) esse processo iterativo pode ser resolvido alternando-se as duas equações até a convergência. No chute inicial para ϕ_i pode-se considerar $\phi_i = \hat{\phi}$, $\forall i = 1, \dots, n$, onde $\hat{\phi}$ é a estimativa de ϕ obtida de um MLG.

Sob as condições de regularidade, temos que para n grande $\hat{\boldsymbol{\beta}} \sim N_p \left(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1} \right)$ e $\hat{\boldsymbol{\gamma}} \sim N_q \left(\boldsymbol{\gamma}, \mathbf{K}_{\gamma\gamma}^{-1} \right)$, respectivamente. Além disso, devido à ortogonalidade entre $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ segue a independência assintótica entre $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$.

3.1 Estimação via Função de Verossimilhança Restrita

Como proposto em (SMYTH; VERBYLA, 1999), a estimação via função de verossimilhança restrita considera uma estimação conjunta de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, tal método é utilizado nos pacotes do *software R*, *DGLM* e *GAMLSS*.

Considere o seguinte modelo normal linear ponderado

$$y \sim N(X\beta, \sigma^2 V)$$

Onde X é o posto k e V é a matriz de pesos conhecida. $\hat{\beta}$ é o estimador de máxima verossimilhança de β . Para $\hat{\sigma}^2$, o estimador de verossimilhança restrito é dado por

$$\hat{\sigma}^2 = \frac{1}{n-k} (\mathbf{y} - X\hat{\beta})^T V^{-1} (\mathbf{y} - X\hat{\beta})$$

O modelo normal linear ponderado geral é dado por

$$y \sim N(X\beta, \Sigma(\gamma))$$

Com Σ dependente do vetor γ de parâmetros desconhecidos

Como dito anteriormente o estimador de MQO de γ dado $pmb\beta$ é dado por

$$Z^T W_d Z \gamma = Z^T W_d \mathbf{z}_d$$

Onde

$$W_d = \text{diag}\left(\frac{1}{h(\phi_i)^2 V_d(\phi_i)}\right)$$

(MCCULLAGH; TIBSHIRANI, 1990) considera a expressão obtida via expansão de Taylor para a variância $d(y_i, \hat{\mu}_i)$ onde a partir dessa expressão a variância de $\hat{\gamma}$ pode obtida

$$\text{var}(\hat{\gamma}) \approx I_{\gamma\gamma}^{-1}$$

Com

$$I_{\gamma\gamma} = \frac{1}{2} Z^T W_d^* Z$$

E

$$W_d^* = W_d - 2\text{diag}\left(\frac{h_i}{\phi^2 h(\phi_i)_i^2}\right) + H^2$$

Tal expressão nos dá um método direto para a conversão da estimação via máxima verossimilhança de γ para uma aproximação via máxima verossimilhança restrita.

(SMYTH; VERBYLA, 1999) cita que o método de mínimos quadrados ponderados iterativo para γ é dado apenas pela substituição de \mathbf{d} para $\mathbf{d}^* = (z_{d1}^*, z_{d1}^*, \dots, z_{dn}^*)^T$ e a matriz de pesos W_d por W_d^* . Tal substituição garante não apenas que $\hat{\gamma}$ seja aproximadamente não viciado, mas também que o submodelo de dispersão nos dê os erros padrão ajustados.

Em problemas reais, a matrix H^2 é custosa de se estimar, assim (MCCULLAGH; TIBSHIRANI, 1990) utiliza uma aproximação via $\text{diag}(h_i^2)$

Assim a matriz de pesos W_d^* é dada por

$$W_d^* \approx W_d - 2diag\left(\frac{h_i}{\phi_i^2 h(\phi_i)^2} + h_i^2\right)$$

Na prática, a a função de ligação h para o parâmetro de dispersão será Log, onde nesse caso a expressão para W_d^* se torna mais simples.

Para $h(\phi) = 1/\phi$ e $V_d = \phi_i^2$, temos que $W_d^* \approx diag(1 - h_i^2)$

4 Análise de Resíduos e Diagnóstico

Nesta seção vamos apresentar métodos de análise de resíduos e diagnóstico no contexto dos MLGs duplos. Diferentemente dos MLGs, nos MLGs duplos podemos definir desvios para a média e para precisão.

4.1 Análise de Resíduos

No caso dos desvios para média, utiliza-se a mesma expressão da classe dos MLGs, substituindo ϕ por ϕ_i , ou seja, $D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \boldsymbol{\phi}) = \sum_{i=1}^n d_1^{*2}(y_i; \hat{\mu}_i, \phi_i)$ em que $d_1^{*2}(y_i; \hat{\mu}_i, \phi_i) = 2\phi_i \left[y_i (\tilde{\theta}_i - \hat{\theta}_i) + \left\{ b(\tilde{\theta}_i) - b(\hat{\theta}_i) \right\} \right]$.

Para ϕ_i grande $\forall i$ o desvio $D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \boldsymbol{\phi})$ pode ser comparado com os percentis da distribuição \mathbf{X}^2 com $(n - p)$ graus de liberdade.

O residuo Studentizado, no caso dos MLGs duplos será dado por

$$t_{D_{1i}} = \frac{d_1^*(y_i; \hat{\mu}_i, \hat{\phi}_i)}{\sqrt{1 - \hat{h}_{ii}}},$$

em que $d_1^*(y_i; \hat{\mu}_i, \hat{\phi}_i) = \pm \sqrt{d_1^{*2}(y_i; \hat{\mu}_i, \hat{\phi}_i)}$, o sinal continua sendo o mesmo de $(y_i - \hat{\mu}_i)$ e \hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz.

No modelo normal heteroscedástico, assume a forma

$$t_i^* = \frac{y_i - \hat{y}_i}{\hat{\sigma}_i \sqrt{1 - \hat{h}_{ii}}},$$

em que $\hat{h}_{ii} = \hat{\sigma}_i^2 \mathbf{x}_i^T (\mathbf{X}^T \hat{\boldsymbol{\Phi}} \mathbf{X})^{-1} \mathbf{x}_i$ com $\hat{\boldsymbol{\Phi}} = diag \{ \sigma_1^{-2}, \dots, \sigma_n^{-2} \}$.

$$\hat{\mathbf{H}} = \hat{\boldsymbol{\Phi}}^{1/2} \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\boldsymbol{\Phi}} \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Phi}}^{1/2} \hat{\mathbf{W}}^{1/2},$$

ou seja,

$$\hat{h}_{ii} = \hat{\phi}_i \hat{\omega}_i \mathbf{x}_i^T (\mathbf{X}^T \hat{\boldsymbol{\Phi}} \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i.$$

Além disso, definimos o desvio para a precisão por $D_2^*(\mathbf{y}; \hat{\boldsymbol{\phi}}, \boldsymbol{\mu}) = \sum_{i=1}^n d_2^{*2}(y_i; \hat{\phi}_i, \mu_i)$, em que $d_2^{*2}(y_i; \hat{\phi}_i, \mu_i) = 2 \left[t_i (\tilde{\phi}_i - \hat{\phi}_i) + \left\{ d(\tilde{\phi}_i) - d(\hat{\phi}_i) \right\} \right]$, $\tilde{\phi}_i$ é solução para ϕ_i sob o

modelo saturado sendo dada por $d'(\tilde{\phi}_i) = -t_i$. Considerando ϕ_i grande $\forall i$ o desvio $D_2^*(\mathbf{y}; \hat{\phi}, \hat{\mu})$ pode ser comparado com os percentis da distribuição qui-quadrado com $(n-q)$ graus de liberdade.

O residuo componente do desvio para a precisão fica dado por

$$t_{D_{2i}} = \frac{d_2^*(y_i; \hat{\phi}_i, \hat{\mu}_i)}{\sqrt{1 - \hat{r}_{ii}}},$$

em que $d_2^*(y_i; \hat{\phi}_i, \hat{\mu}_i) = \pm \sqrt{d_2^{*2}(y_i; \hat{\phi}_i, \hat{\mu}_i)}$, o sinal sendo o mesmo de $\{\hat{t}_i + d'(\hat{\phi}_i)\}$ e \hat{r}_{ii} é o i -ésimo elemento da diagonal principal da matriz

$$\hat{\mathbf{R}} = \hat{\mathbf{P}}^{1/2} \mathbf{Z} (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{P}}^{1/2},$$

ou seja,

$$\hat{r}_{ii} = \hat{p}_i \mathbf{z}_i^T (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{z}_i.$$

A fim de analisar os resíduos, (PAULA, 2013) sugere o gráfico normal de probabilidades e contra os valores ajustados para t_{D_1} e t_{D_2} .

4.2 Influência

Com o objetivo de investigar se existem pontos influentes em um modelo MLG duplo, calcula-se LD_i , como nos modelos MLG, porém com $\hat{\phi}_i$ no lugar de $\hat{\phi}$, ou seja, esta será definida por

$$LD_i^\beta = \left\{ \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right\} t_{S_i}^2,$$

em que

$$t_{S_i} = \frac{\sqrt{\hat{\phi}_i} (y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i (1 - \hat{h}_{ii})}}$$

Na avaliação da sensibilidade da estimativa $\hat{\gamma}$ quando a i -ésima observação é deletada usaremos uma aproximação de um passo

$$\hat{\gamma}_{(i)} = \hat{\gamma} - \frac{(\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{z}_i \{t_i + d'(\hat{\phi}_i)\}}{h'(\hat{\phi}_i) (1 - \hat{r}_{ii})},$$

em que \hat{r}_{ii} é o i -ésimo elemento da diagonal principal da matriz \mathbf{R} . Uma medida para avaliar a influência nas estimativas dos parâmetros da precisão fica dada por

$$\begin{aligned} LD_i^\gamma &= (\hat{\gamma}_{(i)} - \hat{\gamma})^T (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z}) (\hat{\gamma}_{(i)} - \hat{\gamma}) \\ &= \left\{ \frac{\hat{r}_{ii}}{1 - \hat{r}_{ii}} \right\} t_{T_i}^2, \end{aligned}$$

em que

$$\begin{aligned} t_{T_i} &= \frac{t_i + d'(\hat{\phi}_i)}{h'(\hat{\phi}_i) \sqrt{\hat{p}_i(1 - \hat{r}_{ii})}} \\ &= \frac{t_i + d'(\hat{\phi}_i)}{\sqrt{-d''(\hat{\phi}_i)(1 - \hat{r}_{ii})}}. \end{aligned}$$

A fim de analisar possíveis pontos de influência, (PAULA, 2013) sugere o gráfico normal de probabilidades e contra os valores ajustados para LD_i^β e LD_i^γ .

5 Testes de Hipótese

Uma vez que, os MLGs duplos são um extensão dos MLGs podemos utilizar a teoria básica dos MLGs, realizando as devidas alterações, considerando ϕ_i não constante. Dessa forma, podemos relizar os testes de hipótese para β e γ .

Com isso definimos o teste da razão de verossimilhanças para β dado por

$$\xi_{RV} = 2 \left\{ L(\hat{\beta}, \hat{\gamma}) - L(\beta^0, \hat{\gamma}) \right\}.$$

da mesma forma, podemos testar em relação a γ

$$\xi_{RV}^\gamma = 2 \left\{ L(\hat{\beta}, \hat{\gamma}) - L(\hat{\beta}, \gamma^0) \right\}.$$

nota-se, quando $\gamma_j^0 = 0, \forall j \neq 0$ estamos no caso específico da classe dos MLGs.

O teste de Wald para β é definido, nesse caso, por

$$\xi_W = [\hat{\beta} - \beta^0]^T \hat{\text{Var}}^{-1}(\hat{\beta}) [\hat{\beta} - \beta^0],$$

da mesma forma para γ

$$\xi_W^\gamma = [\hat{\gamma} - \gamma^0]^T \hat{\text{Var}}^{-1}(\hat{\gamma}) [\hat{\gamma} - \gamma^0],$$

onde $\text{Var}^{-1}(\hat{\beta})$ e $\text{Var}^{-1}(\hat{\gamma})$ foram definidos na seção 3 deste trabalho.

6 Estudo de Simulação

O escopo desse trabalho se limita apenas em Modelos Lineares Generalizados Duplos, mas é importante destacar o avanço das classes mais gerais desses modelos nas últimas décadas, os chamados GAMLSS, classe de modelo que dá o nome do pacote que será utilizado nessa seção

Os chamados Modelos Aditivos Generalizados para Localização, Escala e Forma (Generalized Additive Models for Location, Escala and Shape - GAMLSS) surgiram como uma

extensão aos os MLGs Duplos, onde além de oferecer a possibilidade de modelagem dos parâmetros de locação e escala (também chamado de dispersão), oferece suporte para a modelagem de assimetria e curtose.

O estudo de simulação utilizou 2 exemplos com distribuições distintas, Normal e Gama, para o estudo da estimação dos parâmetros de locação e escala

Nesse sentido, β e de γ foram fixados

$$\beta^T = (5, 0, 2, 3) \quad \gamma^T = (2, 3, 0, 0)$$

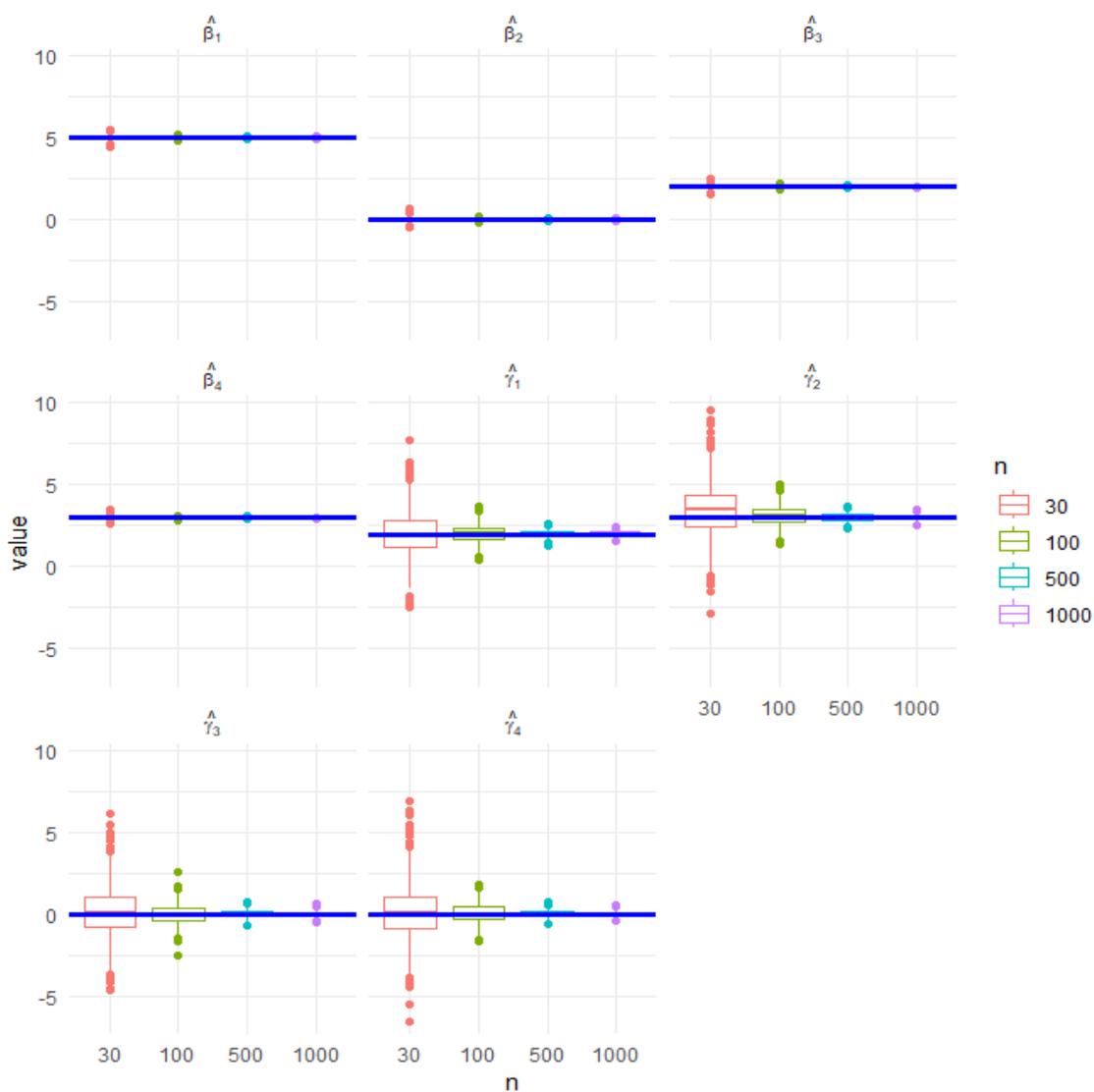
Os tamanhos amostrais 30, 100, 500 e 1000 foram simulados 1000 vezes cada. As funções de ligação para μ foram a função identidade e logarítmica para, respectivamente, distribuições normal e gamma. A função de ligação utilizada para ϕ foi a função logarítmica nos dois casos.

Dado as condições iniciais, estudo de simulação se baseou no seguinte algoritmo:

1. Criação de um vetor de tamanhos amostrais e fixação dos valores de β e γ verdadeiros
2. Cálculo dos componentes lineares $\eta = \mathbf{X}^T \beta$ e $\lambda = \mathbf{X}^T \gamma$
3. Geração de valores de um distribuição Normal ($\mu_i = \eta_i, \sigma_i^2 = \frac{1}{e^{\lambda_i}}$) e Gama ($\alpha_i = e^{\lambda_i}, \theta_i = \frac{e^{\lambda_i}}{e^{\eta_i}}$)
4. Estimação de β e γ via modelos duplos
5. Construção de gráficos e cálculo de métricas de $\hat{\beta}$ e $\hat{\gamma}$

A partir de tal sequência, o seguinte gráfico da estimativas por tamanho amostral foi contruido.

O gráfico tras na linha azul central o verdadeiro valor do parâmetro, enquanto cada boxplot representa o tamanho amostral utilizado e simulado 1000 vezes para a estimação do parâmetro desejado.



Além dos gráficos, a métrica EQM(Erro Quadrático Médio) para cada estimador foram calculados e dispostos na seguinte tabela

Tabela 1 – Valores de EQM - Modelo Normal

Parâmetro	n = 30	n = 100	n = 500	n = 1000
$\hat{\beta}_1$	0.02	0.00	0.00	0.00
$\hat{\beta}_2$	0.02	0.00	0.00	0.00
$\hat{\beta}_3$	0.01	0.00	0.00	0.00
$\hat{\beta}_4$	0.02	0.00	0.00	0.00
$\hat{\gamma}_1$	2.02	0.27	0.04	0.02
$\hat{\gamma}_2$	2.68	0.32	0.05	0.03
$\hat{\gamma}_3$	2.71	0.35	0.05	0.03
$\hat{\gamma}_4$	2.74	0.33	0.05	0.03

A partir do gráfico, foi possível ver que para o menor tamanho amostral simulado, as estimativas para β já apresentavam uma boa precisão,

Para um n igual a 100, o EQM de $\hat{\beta}$ se mostraram aproximadamente 0

Já para $\hat{\gamma}$, vemos uma precisão menor dos estimadores, dado sua complexidade de estimação. Seu EQM foi em média 2,5 para um tamanho amostral igual a 30, onde esse se aproximou de 0 para um n de 500.

No caso da distribuição Gama, tivemos o seguinte resultado.

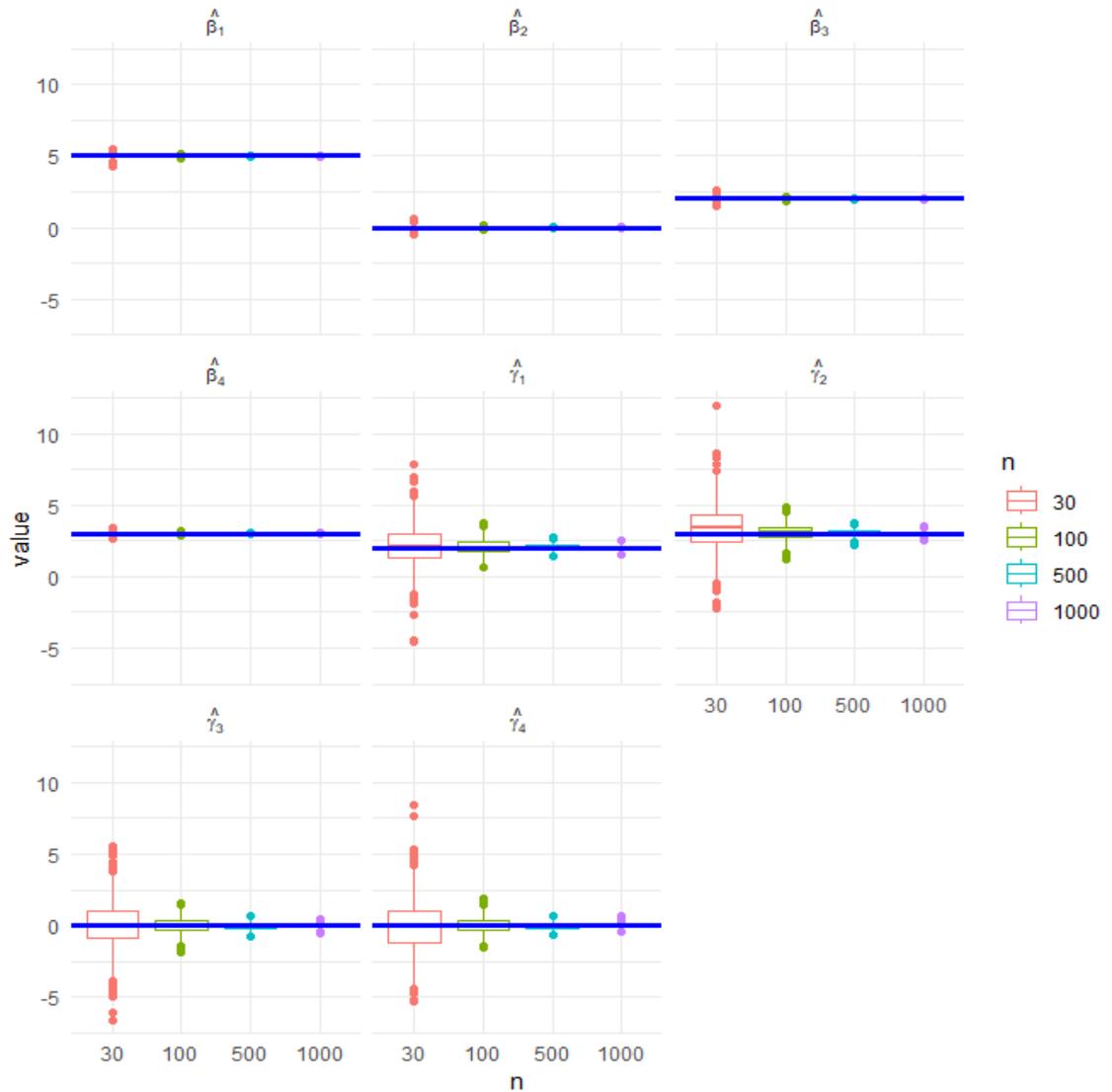
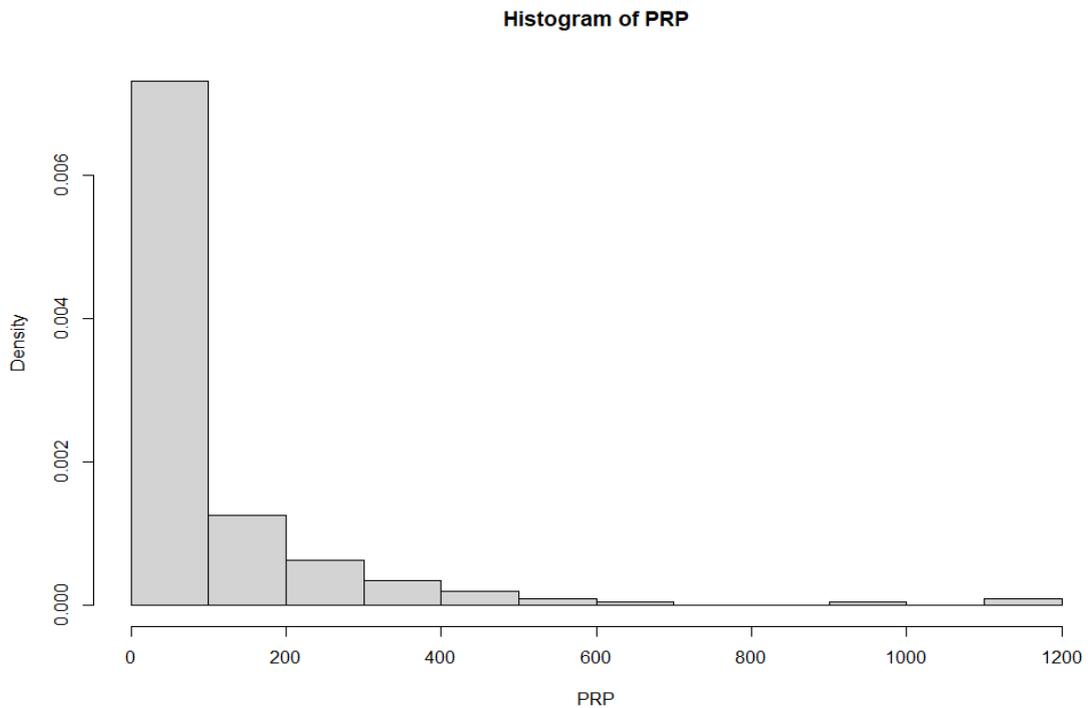


Tabela 2 – Valores de EQM - Modelo Gama

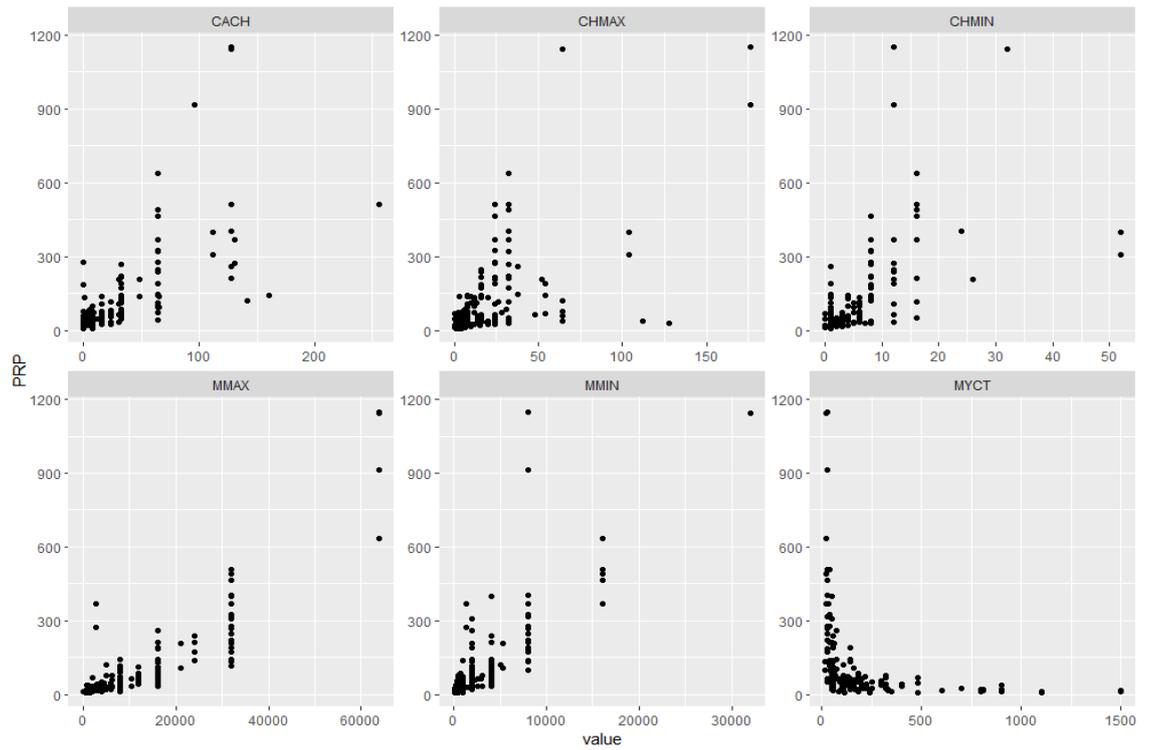
Parâmetro	n = 30	n = 100	n = 500	n = 1000
$\hat{\beta}_1$	0.02	0.00	0.00	0.00
$\hat{\beta}_2$	0.02	0.00	0.00	0.00
$\hat{\beta}_3$	0.01	0.00	0.00	0.00
$\hat{\beta}_4$	0.02	0.00	0.00	0.00
$\hat{\gamma}_1$	1.85	0.25	0.04	0.02
$\hat{\gamma}_2$	2.36	0.32	0.05	0.02
$\hat{\gamma}_3$	2.49	0.29	0.05	0.03
$\hat{\gamma}_4$	2.47	0.29	0.05	0.02

7 Aplicação

Na aplicação de dados reais, consideramos os dados apresentados em (EIN-DOR; FELDMESSER, 1987) sobre a performance de CPUs (Unidade Central de Processamento). Considera-se a variável resposta *PRP* (performance relativa publicada), e, as variáveis dependentes *CACHE* (tamanho de memória cache), *CHMIN* (número mínimo de canais), *CHMAX* (número máximo de canais), *MCYT* (tempo de ciclo da máquina), *MMIN* (mínimo de memória principal), *MMAX* (máximo de memória principal). Na figura abaixo segue o histograma da variável resposta.

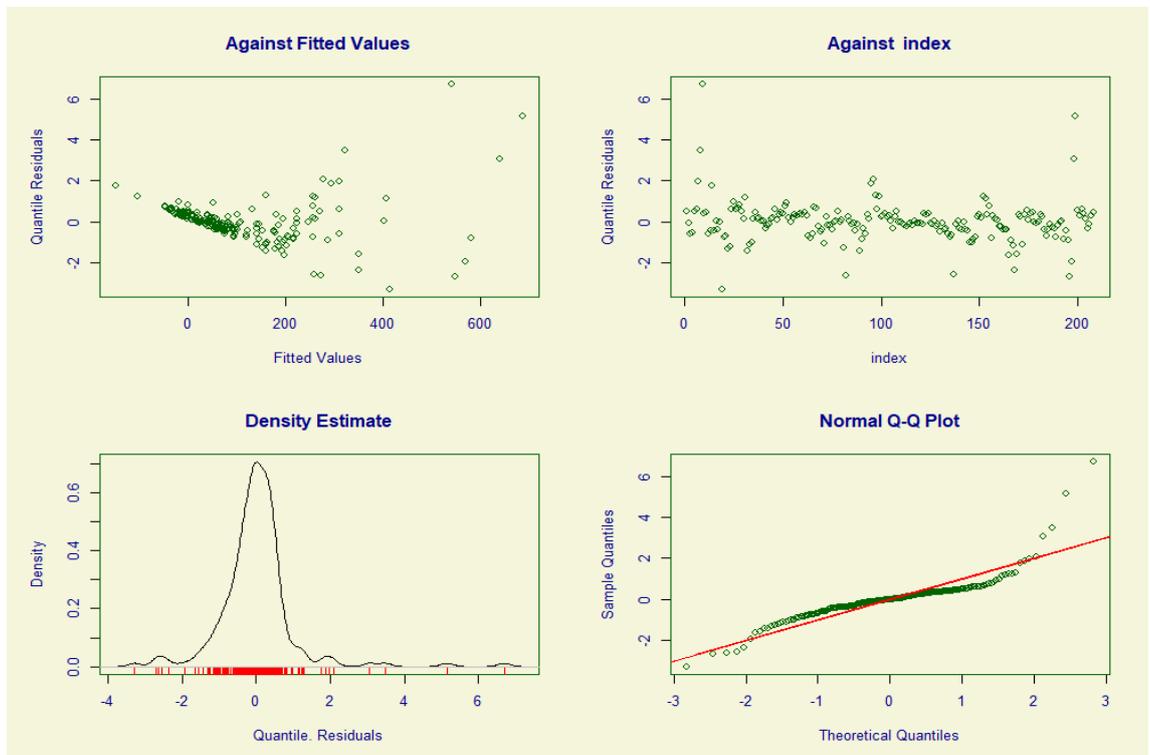


Primeiramente, foram analisadas a relação das variáveis dependentes com a performance relativa, como é possível notar na figura abaixo

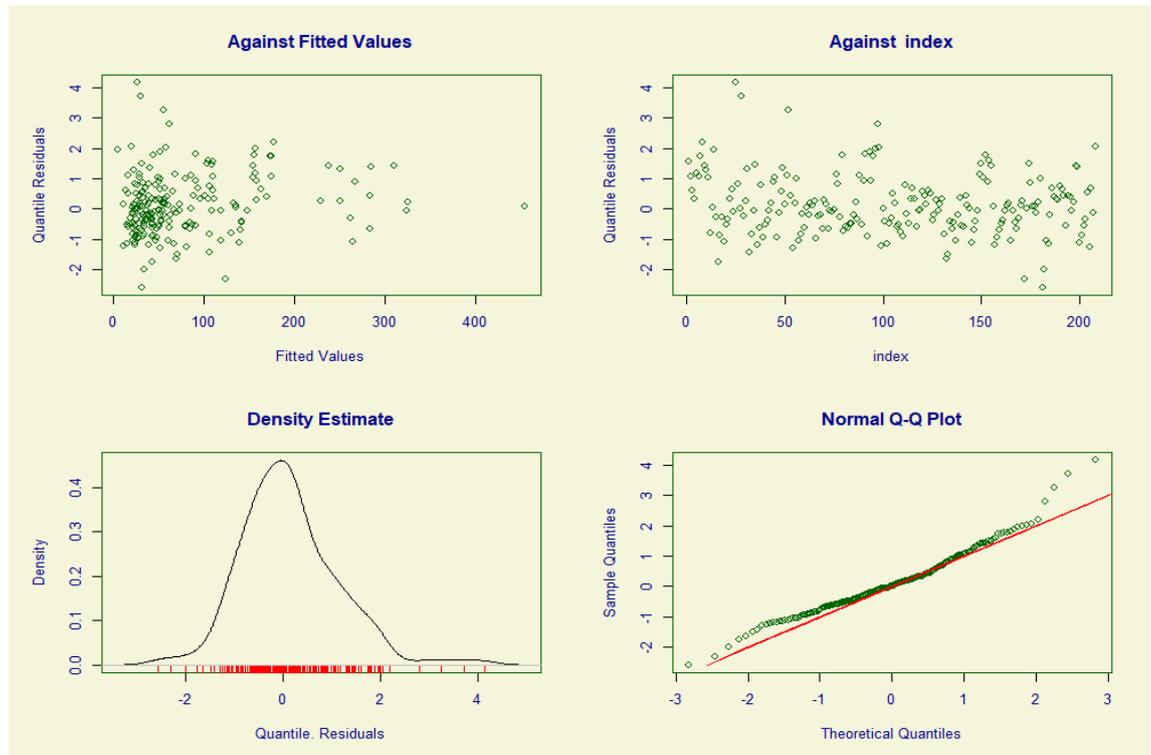


Nota-se que existe uma relação positiva com as todas as variáveis, a não ser o tempo do ciclo. Além disso, já existem indícios de haver heterogeneidade, onde a variabilidade parece aumentar conforme as variáveis independentes aumentam.

Em um primeiro momento, ajustamos um modelo normal com todas as variáveis, e verificamos seu desempenho em relação a variabilidade dos resíduos.



Nota-se um desempenho muito ruim, onde claramente a variabilidade dos resíduos aumenta em relação aos valores ajustados. Nesse sentido, vamos considerar um modelo normal heterocedástico, modelando a variância utilizando todas as variáveis independentes.



Utilizando o modelo duplo normal, nota-se uma melhora muito grande principalmente em relação a heterogeneidade dos resíduos padronizados, onde agora não existe um padrão contra os valores ajustados.

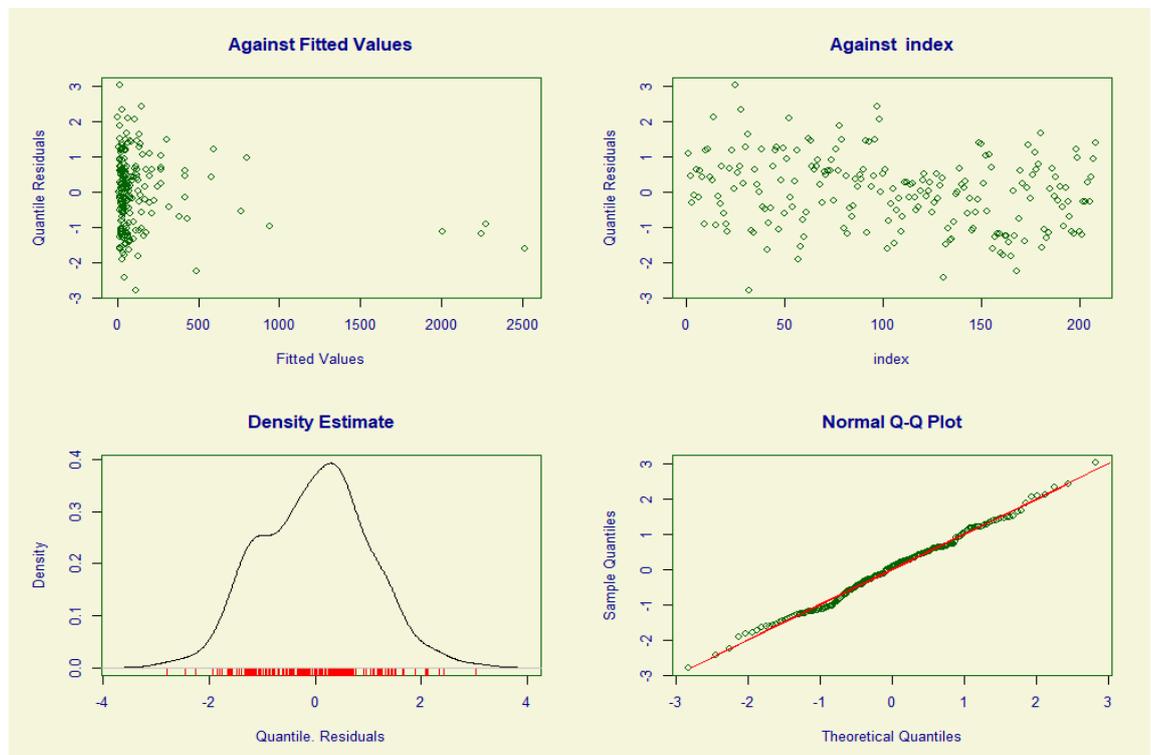
Após isso, foram consideradas as distribuições Gamma e Normal inversa, utilizando as ligações, log, identidade e inversa para μ e para $\sigma(\frac{1}{\sqrt{\phi}})$. O método *forward stepwise* foi utilizado considerando o critério de informação bayesiano (*BIC*) para seleção de variáveis. Os resultados são apresentados na tabela abaixo

Tabela 3 – BIC dos modelos ajustados

Distribuição	Ligação μ	Ligação $\frac{1}{\sqrt{\phi}}$	BIC
Gama	Log	Log	1961.18
Gama	Identidade	Log	-
Gama	Inversa	Log	-
Gama	Log	Identidade	-
Gama	Log	Inversa	1958.77
Gama	Identidade	Identidade	-
Gama	Identidade	Inversa	-
Gama	Inversa	Identidade	-
Gama	Inversa	Inversa	-
Normal Inversa	Log	Log	1966.42
Normal Inversa	Identidade	Log	1968.7
Normal Inversa	Inversa	Log	-
Normal Inversa	Log	Identidade	-
Normal Inversa	Log	Inversa	1929.6
Normal Inversa	Identidade	Identidade	-
Normal Inversa	Identidade	Inversa	1951.54
Normal Inversa	Inversa	Identidade	-
Normal Inversa	Inversa	Inversa	-

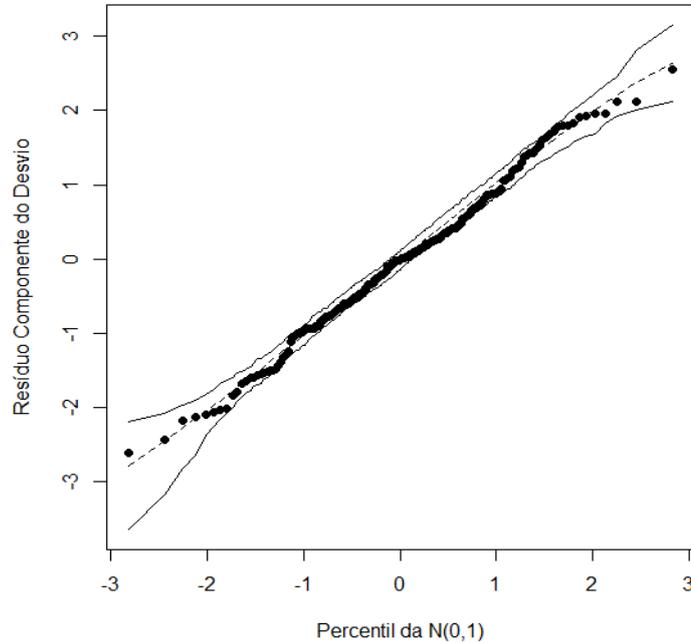
- Valores negativos na transformação resultando em erro

A partir da tabela, vemos que o modelo que obteve o menor BIC foi utilizando a distribuição normal inversa com função ligação log para μ e inversa para σ . Com isso, os resíduos foram analisados.



Nota-se um bom desempenho dos resíduos quando comparamos com o *qqplot* da

normalidade. Em relação aos gráficos contra os valores ajustados e index, o modelo está válido.



A partir do envelope, nota-se um bom ajuste do modelo, com o envelope incluindo todos os resíduos.

Tabela 4 – Coeficientes para $\log(\mu)$

Parâmetro	Estimativa	p-valor
Intercepto	3.543	2e-16 ***
MYCT	-8.888e-04	1.43e-11 **
MMIN	6.651e-05	1.54e-07 ***
MMAX	3.436e-05	2e-16 ***
CACH	8.997e-03	2.38e-12 ***

Tabela 5 – Coeficientes para $\frac{1}{\sigma}(\sqrt{\phi})$

Parâmetro	Estimativa	p-valor
Intercepto	5.384880219	1.46e-07 ***
MMAX	0.001175228	9.35e-13 **
CHMAX	0.201854466	0.00549 ***

Ao analisar os coeficientes, nota-se que todos os coeficientes foram significativos ao nível de 5%, onde temos uma melhora da performance conforme os tamanhos de memória mínimo, máximo e cache aumentam, enquanto existe um decaimento em relação ao tempo do ciclo da unidade de processamento.

Em relação aos coeficientes para $\frac{1}{\sigma}(\sqrt{\phi})$, temos que as variáveis de máximo de memória e número máximo de canais com estimativas positivas, ou seja, a variabilidade

diminui quanto maior os valores destas variáveis.

8 Conclusão

Ao longo do trabalho, os Modelos Lineares Generalizados Duplos foram profundamente estudados, onde os tópicos de definição e estimação de modelos, análise de resíduos e diagnósticos foram explorados.

Os modelos tiveram 2 rodadas de aplicação. Na primeira etapa, tais modelos foram aplicados em dados simulados, onde vimos o comportamento desses em diferentes tamanhos amostrais. Em uma segunda etapa, os modelos duplos foram aplicados em dados reais, onde foi possível observar um grande ganho de desempenho em relação aos modelos Gaussianos e MLGs.

Portanto, os Modelos Lineares Generalizados Duplos representam uma ferramenta estatística robusta para modelagem, particularmente quando os conjuntos de dados trabalhados complexos, heterogêneos e multidimensionais.

REFERÊNCIAS

EIN-DOR, P.; FELDMESSER, J. Attributes of the performance of central processing units: A relative performance prediction model. *Communications of the ACM*, ACM New York, NY, USA, v. 30, n. 4, p. 308–317, 1987. Citado na página 13.

MCCULLAGH, P.; TIBSHIRANI, R. A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 52, n. 2, p. 325–344, 1990. Citado na página 6.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, Oxford University Press, v. 135, n. 3, p. 370–384, 1972. Citado na página 2.

PAULA, G. A. Modelos de regressão com apoio computacional. 2013. Citado na pág, v. 1, n. 9, p. 10, 2013. Citado 5 vezes nas páginas 2, 3, 5, 8 e 9.

SMYTH, G. K. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 51, n. 1, p. 47–60, 1989. Citado 2 vezes nas páginas 3 e 5.

SMYTH, G. K.; VERBYLA, A. P. Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics: The official journal of the International Environmetrics Society*, Wiley Online Library, v. 10, n. 6, p. 695–709, 1999. Citado 2 vezes nas páginas 5 e 6.

Agradecimentos

Os autores agradecem o apoio da Universidade Federal de Juiz de Fora, ao departamento de estatística da UFJF, e ao Prof. Dr. Clécio Ferreira, ministrando da disciplina Modelos Lineares Generalizados.