



## Modelos para Séries Temporais com Base na EGEDD

### Distribuição Dagum Exponencial Generalizada Exponencializada

Apresentador: Vítor Bernardo Silveira Pereira

Orientador: Cleber Bisognin

Coorientadora: Laís Helen Loose

# Motivação

- A identificação de uma **distribuição** apropriada é muito importante para a **análise estatística**;
- A distribuição Dagum, idealizada por Camilo Dagum, tem ganhado muita atenção na sua utilidade para modelagem de dados sobre a distribuição de renda e análise de confiabilidade.
- Distribuições baseadas na Dagum:
  - Distribuição exponenciada **Kumaraswamy-Dagum** (Silva et al., 2015);
  - Distribuição **Dagum Exponencial Generalizada Exponencializada** proposta em (Nasiru et al., 2019);
  - **Gompertz Generalizada**, proposta por El-Gohary et al. (2013), entre outros...

# Motivação

- **Aplicações:** Principalmente na área da modelagem de dados de sobrevivência e em dados econômicos, como análise de confiabilidade, modelagem de renda pessoal e nacionais (micro e macroeconômico).



**Augusto Dagum:** Um dos maiores economistas e estatísticos latinos

# Motivação

- Os modelos de **Regressão Linear Normal** e **Modelos Lineares Generalizados** possuem como pressuposto que a distribuição seja Normal ou pertença a família exponencial regular;
- Na prática, existem dados que não se adequam a distribuição normal ou distribuições pertencentes a família exponencial, então nos últimos anos vêm sendo propostos **modelos alternativos**;

# Motivação

- Os modelos de **Regressão Linear Normal** e **Modelos Lineares Generalizados** possuem como pressuposto que a distribuição seja Normal ou pertença a família exponencial regular;
- Na prática, existem dados que não se adequam a distribuição normal ou distribuições pertencentes a família exponencial, então nos últimos anos vêm sendo propostos **modelos alternativos**;
- Nos novos modelos, tem-se uma reparametrização em termos da média ou quantil, para possibilitar interpretação dos parâmetros em termos de medidas de posição ou precisão;
- Na **EGEDD**, no entanto, **não há forma fechada para média e variância**, desta forma reparametrizamos a distribuição em termos dos quantis;
- Com a reparametrização, uma estrutura de regressão é incluída para **modelar a média ou o quantil**, assim tendo a mesma abordagem dos MLG'S;

# Motivação

- Em Séries Temporais, uma das abordagens mais usuais é a metodologia Box e Jenkins, no entanto também estão surgindo modelos alternativos, denominados de **modelos dinâmicos**;
- Assim a proposta do projeto, é **desenvolver um novo modelo (modelo dinâmico) com suporte nos reais**, baseando-se na EGEDD e na distribuição Dagum;

# Propriedades da EGEDD

## Função de distribuição acumulada

$$F_Y(y) = 1 - \left\{ 1 - \left[ 1 - \left( 1 - \left( 1 + \alpha y^{-\delta} \right)^{-\sigma} \right)^{\gamma} \right]^{\eta} \right\}^{\lambda}, \quad y > 0$$

# Propriedades da EGEDD

## Função de distribuição acumulada

$$F_Y(y) = 1 - \left\{ 1 - \left[ 1 - \left( 1 - \left( 1 + \alpha y^{-\delta} \right)^{-\sigma} \right)^{\gamma} \right]^{\eta} \right\}^{\lambda}, \quad y > 0$$

onde os parâmetros

$$\alpha, \sigma, \lambda, \delta, \eta, \gamma > 0$$

Os parâmetros  $\sigma, \lambda, \delta, \eta, \gamma$  são parâmetros de forma e o parâmetro  $\alpha$  é de escala.



# Função densidade de probabilidade

$$f_Y(y) = \frac{\alpha\sigma\lambda\delta\eta\gamma(1 + \alpha y^{-\delta})^{-\sigma-1} \left[1 - (1 + \alpha y^{-\delta})^{-\sigma}\right]^{\gamma-1}}{y^{\delta+1} \left\{1 - \left[1 - \left(1 - (1 + \alpha y^{-\delta})^{-\sigma}\right)^{\gamma}\right]^{\eta}\right\}^{1-\lambda}} \\ * [1 - (1 - (1 + \alpha y^{-\delta})^{-\sigma})^{\gamma}]^{\eta-1}$$

# Função Quantílica

$$Q_Y(\tau) = \left\{ \frac{1}{\alpha} \left[ \left( 1 - \left( 1 - \left( 1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}^{-\frac{1}{\delta}}, \text{ onde } \tau \in (0, 1).$$

# Reparametrizações

$$\mu = \left\{ \frac{1}{\alpha} \left[ \left( 1 - \left( 1 - \left( 1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}^{-\frac{1}{\delta}}$$

# Reparametrizações

$$\mu = \left\{ \frac{1}{\alpha} \left[ \left( 1 - \left( 1 - \left( 1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}^{-\frac{1}{\delta}}$$

## Reparametrização por $\delta$

$$\delta = -\frac{1}{\log(\mu)} \log \left\{ \frac{1}{\alpha} \left[ \left( 1 - \left( 1 - \left( 1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}$$

# Submodelos da EGEDD

Distribuição	$\alpha$	$\lambda$	$\sigma$	$\delta$	$\eta$	$\gamma$
EGDD	$\alpha$	1	$\sigma$	$\delta$	$\eta$	$\gamma$
DD	$\alpha$	1	$\sigma$	$\delta$	1	1
EGEBD	1	$\lambda$	$\sigma$	$\delta$	$\eta$	$\gamma$
DB	1	1	$\sigma$	$\delta$	1	1
EGBD	1	1	$\sigma$	$\delta$	$\eta$	$\gamma$
EGEFD	$\alpha$	$\lambda$	1	$\delta$	$\eta$	$\gamma$
EGFD	$\alpha$	1	1	$\delta$	$\eta$	$\gamma$
FD	$\alpha$	1	1	$\delta$	1	1

Notamos que existem parâmetros melhores para fazer as reparametrizações, sem perder a generalidade da EGED como o  $\delta$ .

# Regressão Dagum

Incluindo a estrutura de regressão pelo quantil, por:

$$g(\mu_t) = x_t^\top \beta = \sum_{j=1}^k x_{tj} \beta_j, \quad t = 1, \dots, n,$$

$$\text{com } \beta = (\beta_1, \dots, \beta_k)^\top$$

# Regressão Dagum

Incluindo a estrutura de regressão pelo quantil, por:

$$g(\mu_t) = x_t^\top \beta = \sum_{j=1}^k x_{tj} \beta_j, \quad t = 1, \dots, n,$$

$$\text{com } \beta = (\beta_1, \dots, \beta_k)^\top$$

## Função de ligação

Necessita ser contínua e duas vezes diferenciável, logo a nossa função de ligação é  $g(\mu_t) = \log(\mu_t)$ .

Assim, ficaremos com o vetor de parâmetros (com a reparametrização por  $\delta$ ), sendo:

$$\theta = (\alpha, \sigma, \lambda, \eta, \gamma, \beta^\top)^\top$$

Em que  $\mu_t$  é o quantil da observação  $y_t$  no tempo  $t$ .

# Modelo DARMA(p,q)

Modelo geral para  $\mu_t$ :

$$g(\mu_t) = x_t^\top \beta + u_t$$

Sabendo que pela estrutura ARMA temos uma relação que inclui o componente  $u_t$ :

$$u_t = \frac{\phi(B)}{\theta(B)} \epsilon_t$$

# Modelo DARMA(p,q)

Modelo geral para  $\mu_t$ :

$$g(\mu_t) = x_t^\top \beta + u_t$$

Sabendo que pela estrutura ARMA temos uma relação que inclui o componente  $u_t$ :

$$u_t = \frac{\phi(B)}{\theta(B)} \epsilon_t$$

$$g(\mu_t) = x_t^\top \beta + \frac{\phi(B)}{\theta(B)} \epsilon_t$$

onde:

$$\phi(z) = - \sum_{i=0}^p \phi_i z^i, \quad \theta(z) = - \sum_{j=0}^q \theta_j z^j$$



Desenvolvendo um pouco algebricamente, conseguimos:

$$\phi(B)(g(\mu_t) - x_t^\top \beta) = \phi(B)\epsilon_t$$

$$(1 - \phi_1 B - \phi_2 B^2, \dots, \phi_p B^p) (g(\mu_t) - x_t^\top) = \phi(B) \epsilon_t$$

Desenvolvendo um pouco algebricamente, conseguimos:

$$\phi(B)(g(\mu_t) - x_t^\top \beta) = \phi(B)\epsilon_t$$

$$(1 - \phi_1 B - \phi_2 B^2, \dots, \phi_p B^p) (g(\mu_t) - x_t^\top \beta) = \phi(B) \epsilon_t$$

Até chegar no:

## Modelo final ARMA:

$$g(\mu_t) = x_t^\top \beta + \sum_{j=1}^n \phi_j (g(\mu_t) - x_t^\top \beta) + \phi(B)\epsilon_t$$

# Estimação dos parâmetros

A estimação dos parâmetros será realizada através de Estimadores de Máxima Verossimilhança (EMV) e a máxima  $L_q$ -verossimilhança ( $ML_qE$ ), com o vetor de parâmetros:

$$\theta = (\alpha, \sigma, \gamma, \eta, \lambda, \delta)$$

# Estimação dos parâmetros

A estimação dos parâmetros será realizada através de Estimadores de Máxima Verossimilhança (EMV) e a máxima  $L_q$ -verossimilhança ( $ML_qE$ ), com o vetor de parâmetros:

$$\boldsymbol{\theta} = (\alpha, \sigma, \gamma, \eta, \lambda, \delta)$$

## Função de log-verossimilhança:

$$\begin{aligned} \ell_t(\boldsymbol{\theta}, y_t) = & \log(\alpha\delta\sigma\gamma\eta\lambda) - (\delta + 1)\log(y_t) - (\sigma + 1)\log(z_t) + (\gamma - 1)\log(1 - z_t^{-\sigma}) \\ & + (\eta - 1)\log\left[1 - (1 - z_t^{-\sigma})^\gamma\right] + (\lambda - 1)\log\left\{1 - \left[1 - (1 - z_t^{-\sigma})^\gamma\right]^\eta\right\} \end{aligned}$$

# Estimador de Máxima Verossimilhança

A estimação do vetor de parâmetros  $\theta$  através do método da máxima verossimilhança, denotado por EMV, é dado por

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \{\ell(\theta)\}$$

# Estimador de Máxima Verossimilhança

A estimação do vetor de parâmetros  $\theta$  através do método da máxima verossimilhança, denotado por EMV, é dado por

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \{\ell(\theta)\}$$

em que  $\Theta \subseteq \{\mathbb{R}_+^5 \times \mathbb{R}^{k+1}\}$  é o espaço de parâmetros e  $\ell(\theta)$  é a função de log-verossimilhança de  $\mathbf{y}$  é dada por

$$\ell(\theta) = \ell(\theta; \mathbf{y}) = \sum_{t=1}^n \ell_t(\theta, y_t),$$

# Estimador de Máxima Verossimilhança

A estimação do vetor de parâmetros  $\boldsymbol{\theta}$  através do método da máxima verossimilhança, denotado por EMV, é dado por

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \{\ell(\boldsymbol{\theta})\}$$

em que  $\Theta \subseteq \{\mathbb{R}_+^5 \times \mathbb{R}^{k+1}\}$  é o espaço de parâmetros e  $\ell(\boldsymbol{\theta})$  é a função de log-verossimilhança de  $\mathbf{y}$  é dada por

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; \mathbf{y}) = \sum_{t=1}^n \ell_t(\boldsymbol{\theta}, y_t),$$

em que

$$\begin{aligned} \ell_t(\boldsymbol{\theta}, y_t) = & \log(\alpha\delta\sigma\gamma\eta\lambda) - (\delta + 1) \log(y_t) - (\sigma + 1) \log(z_t) \\ & + (\gamma - 1) \log(1 - z_t^{-\sigma}) + (\eta - 1) \log\left[1 - (1 - z_t^{-\sigma})^\gamma\right] \\ & + (\lambda - 1) \log\left\{1 - \left[1 - (1 - z_t^{-\sigma})^\gamma\right]^\eta\right\}, \end{aligned}$$

# Estimador de Máxima Verossimilhança

Considerando a reparametrização por  $\delta$  :

$$\delta = -\frac{1}{\log(\mu_t)} \log \left\{ \frac{1}{\alpha} \left[ \left( 1 - \left( 1 - \left( 1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\},$$

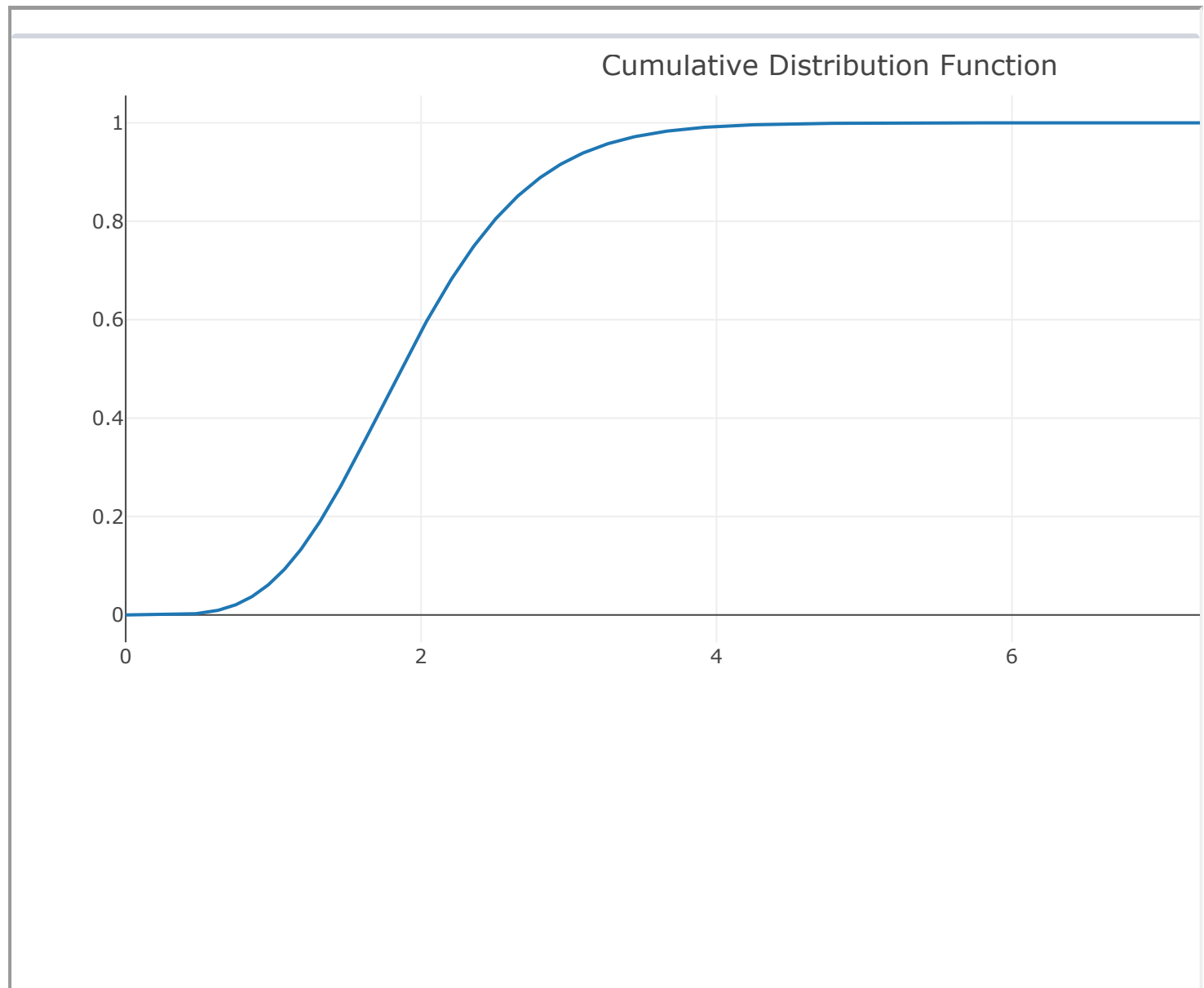
o vetor de parâmetros do modelo proposto é dado por

$$\boldsymbol{\theta} = (\alpha, \sigma, \gamma, \eta, \lambda, \boldsymbol{\beta}^\top)^\top$$



# Outros Procedimentos Estatísticos:

- Construção de testes de hipótese para verificar distribuição dos dados, verificando se é para a EGEDD ou um submodelo;
- Utilização de métodos numéricos para execução de simulações (Newton-Rapson, Gosolnp com os métodos CG, BFGS e L-BFGS-B);
- Utilização de Simulações de Monte Carlo para estimação de parâmetros e análise de diagnósticos dos modelos;
- Desenvolvimento de Intervalos de Confiança e testes de hipóteses baseados na distribuição assintótica dos EMV's;
- Utilização dos estimadores de Entropia Extensiva;
- Aplicação a dados reais.



# Referências Bibliográficas

- Dagum, C. (1977). El modelo log-logístico y la distribución del ingreso en la argentina. El Trimestre Económico, 44(176 (4):837–864
- Nasiru, S., Mwita, P. N., e Ngesa, O. (2019). Exponentiated generalized exponential dagum distribution. Journal of King Saud University-Science, 31(3):362–371.
- R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria