

Comparação de desempenho e adequação de três processos para séries temporais INAR

Trabalho de Conclusão de Curso

Discente: Paulo Manoel da Silva Junior

Orientador: Prof^a Dr^a Tatiene Correia de Souza

www.de.ufpb.br



Departamento de
ESTATÍSTICA



Sumário

Introdução

Objetivos

Referencial Teórico

Métodos de Estimação

Previsão do Processo INAR(1)

Medidas de qualidade do ajuste

Metodologia

Resultados

Simulação de Monte Carlo

Aplicação com dados reais

Considerações Finais

Referências



Introdução

As séries temporais de valores inteiros não negativos, conhecidas como *pontos de contagem*, têm ampla aplicação em diferentes áreas, com destaque para o mercado financeiro, epidemiologia, telecomunicações e o setor de seguros. Exemplos típicos incluem o número de chamadas telefônicas por hora, a quantidade de novos clientes que se tornam inadimplentes por dia, o número de clientes que atingem a faixa de *over 90* em determinadas instituições, ou ainda o registro diário de casos de doenças.

Por possuírem natureza discreta, essas séries desafiam a aplicação de modelos clássicos, como AR, MA, ARMA e ARIMA, que pressupõem variáveis contínuas e normalmente distribuídas.



Introdução

Motivação

Diante da inadequação dos modelos tradicionais para lidar com variáveis discretas, surge a necessidade de métodos específicos que respeitem essa natureza.

Nesse contexto, foram propostos os modelos autorregressivos para dados inteiros não negativos. Entre eles, destacam-se os processos **INAR** (*Integer-value autoregressive*), que preservam a estrutura discreta da variável por meio da aplicação de um operador de *thinning*.



Introdução

A formulação original do processo INAR(1) foi apresentada de forma independente por (Al-Osh; Alzaid, 1987) e (McKenzie, 1985). Nesses trabalhos, foi empregado o operador *binomial thinning*, proposto por (Steutel; Harn, 1979), que possibilita a definição de processos autorregressivos compatíveis com a natureza discreta da variável.

Os modelos INAR são amplamente utilizados por incorporarem, em sua formulação, a dependência temporal entre as observações. Essa dependência é representada por meio do operador *binomial thinning*, denotado por $\alpha \circ X_{t-1}$, em que α corresponde à taxa de retenção da série, isto é, a medida da influência da observação anterior sobre a seguinte.



Introdução

O termo de inovação, ϵ_t , representa a contribuição probabilística adicionada ao processo e é modelado por uma distribuição discreta. Usualmente, adota-se a distribuição de Poisson, escolha justificada pela simplicidade das análises e pela facilidade de interpretação dos resultados. Contudo, essa escolha impõe uma limitação, pois na distribuição de Poisson a média é igual à variância — condição que nem sempre é observada nos dados reais. Por essa razão, a literatura já contempla alternativas, como distribuições que permitem que a variância seja diferente da média, a exemplo da binomial negativa e da geométrica.



Objetivo Geral

O objetivo deste trabalho é analisar o desempenho de dois métodos de estimação dos parâmetros do processo INAR(1), considerando três distribuições distintas para o termo de inovação. Os estimadores adotados são: mínimos quadrados condicionais (MQC) e máxima verossimilhança condicional (MVC). Além disso, busca-se aplicar esses métodos a dados reais, possibilitando a comparação entre distribuições quanto à qualidade do ajuste, complementada por análise diagnóstica, e utilizando a esperança condicional para realizar previsões.



Objetivos Específicos

Com a finalidade de alcançar o objetivo geral, estabelecem-se os seguintes objetivos específicos:

1. Realizar estudos de simulação em diferentes cenários;
2. Avaliar o desempenho dos estimadores por meio de medidas de qualidade do ajuste;
3. Aplicar os métodos de estimação a uma série temporal extraída do Banco Central do Brasil;
4. Comparar os resultados dos modelos em função dos métodos de estimação adotados.



Processo INAR

O processo INAR ganha destaque por diversos motivos, entre os quais podem ser citados:

- ▶ **Respeito à natureza discreta dos dados:** garante que tanto os valores ajustados quanto as previsões pertençam ao conjunto dos números inteiros não negativos;
- ▶ **Representação adequada da dependência temporal:** descreve as observações sucessivas por meio do operador *binomial thinning*, fundamentado na convolução binomial;
- ▶ **Flexibilidade na escolha da distribuição:** possibilita empregar diferentes distribuições para o termo de inovação, como a binomial negativa e a geométrica, em comparação à Poisson, de modo a lidar com padrões distintos de dispersão e variabilidade.



Operador *thinning binomial*

Definição

Seja Y uma variável aleatória não negativa que assume valores inteiros e seja $\alpha \in [0, 1]$. O operador *binomial thinning* é definido por

$$\alpha \circ Y = \sum_{i=1}^Y N_i,$$

em que as variáveis aleatórias N_i são independentes e identicamente distribuídas (i.i.d.) segundo uma distribuição Bernoulli de parâmetro α .



Processo INAR(1)

Com base na definição do operador *binomial thinning*, pode-se formalizar o processo autorregressivo de valores inteiros de ordem um.

Definição

Um processo estocástico discreto de valores inteiros não negativos $\{Y_t, t \in \mathbb{Z}\}$ é denominado processo INAR(1) se satisfaz a seguinte equação recursiva:

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t, \quad (1)$$

em que $\alpha \in (0, 1)$ e ϵ_t é uma sequência de variáveis aleatórias i.i.d. assumindo valores inteiros não negativos, tal que $E(\epsilon_t) = \mu_\epsilon$ e $Var(\epsilon_t) = \sigma_\epsilon^2$.



Processo INAR(1)

A interpretação do processo INAR(1) pode ser apresentada da seguinte forma: Y_t representa a população remanescente em uma determinada região no instante t . O termo $\alpha \circ Y_{t-1}$ corresponde à parcela de habitantes que permaneceram na região do período anterior, enquanto ϵ_t representa os novos indivíduos que chegaram no tempo t . Por sua vez, a diferença $Y_{t-1} - \alpha \circ Y_{t-1}$ indica a quantidade de habitantes que deixaram a região.



Propriedades do processo INAR(1)

Seja $\{Y_t, t \in \mathbb{Z}\}$ um processo INAR(1), com $\mu_\epsilon = E(\epsilon_t)$ e $\sigma_\epsilon^2 = \text{Var}(\epsilon_t)$. Então, $\{Y_t, t \in \mathbb{Z}\}$ satisfaz as seguintes propriedades:

1. $E(Y_t) = \frac{\mu_\epsilon}{1 - \alpha}$
2. $\text{Var}(Y_t) = \frac{\sigma_\epsilon^2 + \alpha\mu_\epsilon}{1 - \alpha^2}$
3. $E(Y_t | Y_{t-1}) = \alpha Y_{t-1} + \mu_\epsilon$
4. $\text{Var}(Y_t | Y_{t-1}) = \alpha(1 - \alpha)Y_{t-1} + \sigma_\epsilon^2$
5. $\text{Cov}(Y_t, Y_{t+j}) = \alpha^j \text{Var}(Y_t), \quad j \in \mathbb{N}$
6. $\rho_Y(h) = \text{Corr}(Y_t, Y_{t+h}) = \alpha^h, \quad h \in \mathbb{N}$



Processo Poisson INAR(1)

Seja $\{\epsilon_t, t \in \mathbb{Z}\}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas segundo uma Poisson com parâmetro λ . Então, o processo $\{Y_t, t \in \mathbb{Z}\}$ é denominado **Poisson-INAR(1)**.

Neste caso, a média e a variância das inovações coincidem, ou seja, $\mu_\epsilon = \sigma_\epsilon^2 = \lambda$.

Propriedades da distribuição Poisson:

- ▶ Função de probabilidade: $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$;
- ▶ Esperança: $E(X) = \lambda$;
- ▶ Variância: $\text{Var}(X) = \lambda$.



Processo Geométrico INAR(1)

Considere $\{Y_t, t \in \mathbb{Z}\}$ um processo INAR(1), definido conforme a equação 1. Quando $\{\epsilon_t, t \in \mathbb{Z}\}$ é uma sequência de variáveis aleatórias independentes com distribuição geométrica parametrizada pela média μ , então $\{Y_t, t \in \mathbb{Z}\}$ é denominado processo Geométrico-INAR(1). Nesse caso, obtêm-se os seguintes resultados:

- ▶ $p = \frac{1}{1 + \mu}$
- ▶ $P(X = k) = \frac{\mu^k}{(\mu + 1)^{k+1}}$
- ▶ $E(X) = \mu$
- ▶ $Var(X) = \frac{\mu}{(\mu + 1)^2}$



Processo Binomial Negativo INAR(1)

O processo Binomial Negativo-INAR(1) é usualmente parametrizado pela média e pela variância, de modo a facilitar a comparação entre distribuições alternativas. Nesse caso, obtêm-se os seguintes resultados:

$$\blacktriangleright p = \frac{\mu}{\sigma^2}$$

$$\blacktriangleright r = \frac{\mu^2}{\sigma^2 - \mu}$$

$$\blacktriangleright P(X = k) = \binom{k+r-1}{k} \left(1 - \frac{\mu}{\sigma^2}\right)^k \left(\frac{\mu}{\sigma^2}\right)^r$$

$$\blacktriangleright E(X) = \mu$$

$$\blacktriangleright \text{Var}(X) = \sigma^2$$



Estimadores de Mínimos Quadrados Condicionais

Seja $Y_1, Y_2, Y_3, \dots, Y_t$ uma amostra do processo INAR(1) definido na equação 1. O objetivo é estimar o vetor de parâmetros $\theta = (\alpha, \lambda)$.

Sabemos que:

$$E(Y_t | Y_{t-1}) = \alpha Y_{t-1} + \lambda = g(\theta, Y_{t-1}).$$

Considerando a função:

$$Q_n(\theta) = \sum_{t=2}^n [Y_t - g(\theta, Y_{t-1})]^2.$$



Estimadores de Mínimos Quadrados Condicionais

Os estimadores de mínimos quadrados condicionais de α e λ são aqueles que minimizam $Q_n(\theta)$.

$$\hat{\alpha}_{\text{MQC}} = \frac{\sum_{t=2}^n Y_t Y_{t-1} - \frac{1}{n-1} \sum_{t=2}^n Y_t \sum_{t=2}^n Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2 - \frac{1}{n-1} \left(\sum_{t=2}^n Y_{t-1} \right)^2} \quad (2)$$

Após encontrar o estimador de α , utilizamos esse valor para obter o estimador de λ :

$$\hat{\lambda}_{\text{MQC}} = \frac{1}{n-1} \left(\sum_{t=2}^n Y_t - \hat{\alpha}_{\text{MQC}} \sum_{t=2}^n Y_{t-1} \right). \quad (3)$$



Estimadores de Máxima Verossimilhança Condicional

No caso dos estimadores de máxima verossimilhança condicional, temos que $Y_t = \alpha \circ Y_{t-1} + \epsilon_t$. Como ϵ_t depende da distribuição assumida, a função de log-verossimilhança é dada por:

$$l(\alpha, \lambda) = \sum_{t=2}^n \ln P(Y_t | Y_{t-1}),$$

onde a probabilidade de transição depende da distribuição considerada. Em geral,

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}) = P(\alpha \circ Y_{t-1} + \epsilon_t = y_t).$$



Estimadores de Máxima Verossimilhança Condicional

Assim, a função de log-verossimilhança de um processo INAR(1), tomando como exemplo o caso Poisson, é dada por:

$$l(\alpha, \lambda) = \sum_{t=2}^n \ln \left(\sum_{j=0}^{\min(y_t, y_{t-1})} \binom{y_{t-1}}{j} \alpha^j (1 - \alpha)^{y_{t-1}-j} \frac{e^{-\lambda} \lambda^{y_t-j}}{(y_t - j)!} \right).$$

O destacado da equação é a parte que altera dependendo da distribuição de inovação incorporada ao processo.



Previsão do Processo INAR(1)

Seja $y_1, y_2, y_3, \dots, y_t$ uma amostra do processo INAR(1). Supondo que a série seja conhecida até o tempo t , o interesse está na previsão de y_{t+1} .

Para isso, sabemos que o valor de uma constante k , que minimiza o erro quadrático médio, é dado por:

$$E \left[(y_{t+1} - k)^2 \mid y_t \right].$$

O valor ótimo de k é a esperança condicional de y_{t+1} dado y_t , ou seja:

$$k = E[y_{t+1} \mid y_t],$$

o qual será utilizado como previsão de y_{t+1} .



Média da distribuição condicional um passo à frente

Considerando o que foi apresentado, utilizaremos a média da distribuição condicional um passo à frente para a previsão do processo INAR(1).

Em sua tese de doutorado, (Freeland, 1998) demonstrou que:

$$E(Y_{t+k} \mid Y_t) = \alpha^k Y_t + \mu_\epsilon \frac{1 - \alpha^k}{1 - \alpha}.$$

Substituindo $k = 1$, obtemos:

$$E(Y_{t+1} \mid Y_t) = \alpha Y_t + \mu_\epsilon.$$

Portanto, essa é a expressão utilizada para a previsão do processo INAR(1), dado que conhecemos Y_t .



Medidas de qualidade do ajuste

Na etapa de simulação de Monte Carlo, a qualidade do ajuste — ou seja, a capacidade de estimar os parâmetros corretamente — será analisada por meio de dois métodos, com o objetivo de verificar as propriedades dos estimadores e identificar o mais eficiente para diferentes valores de parâmetros da série.

Assim, considerando K simulações de Monte Carlo de amostras de tamanho n , e seja $\hat{\theta}^{(i)}$ a estimativa obtida na i -ésima repetição, utilizaremos o erro quadrático médio (EQM) simulado e o viés simulado, dados respectivamente por:

$$\text{EQM}(\hat{\theta}) = \frac{1}{K} \sum_{i=1}^K \left(\hat{\theta}^{(i)} - \theta \right)^2,$$

$$\text{vis}(\hat{\theta}) = \theta - \hat{\theta}^*,$$



Medidas de qualidade do ajuste

em que $\hat{\theta}^*$ é a estimativa média de θ , isto é:

$$\hat{\theta}^* = \frac{1}{K} \sum_{i=1}^K \hat{\theta}^{(i)}.$$

Para mensurar a qualidade das previsões, utilizaremos o erro médio absoluto (MAE), a raiz do erro quadrático médio (RMSE) e o erro percentual médio (MPE), definidos por:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|,$$



Medidas de qualidade do ajuste

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2},$$

$$\text{MPE} = \frac{100}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t},$$

em que \hat{y}_t é o valor estimado de y_t no tempo t .



Metodologia

A metodologia adotada neste trabalho consistiu em:

- ▶ Definição do processo INAR(1);
- ▶ Estimação dos parâmetros por dois métodos:
 1. Mínimos Quadrados Condicionais (MQC);
 2. Máxima Verossimilhança Condicional (MVC).
- ▶ Consideração de diferentes distribuições de inovação: Poisson, Binomial Negativa e Geométrica (McKenzie, 1986).



Metodologia

Além da etapa de simulação, realizou-se uma aplicação empírica com dados da **população desocupada na Região Metropolitana do Rio de Janeiro**.

- ▶ Tratamento e análise realizados em R;
- ▶ Visualizações gráficas construídas com o pacote ggplot2;
- ▶ Estimação via MVC implementada com o pacote optimx, que fornece funções robustas de otimização.



Metodologia

- ▶ Comparação do desempenho dos estimadores em cenários controlados (simulações) e aplicação prática (dados reais);
- ▶ Etapas da análise: geração das séries, estimação dos parâmetros, avaliação das previsões;
- ▶ Todo o código desenvolvido está disponível em:
github.com/paulomanoel578/TCC.



Simulação de Monte Carlo

O objetivo é comparar o desempenho dos estimadores **MQC** e **MVC** para os parâmetros α e λ do processo INAR(1).

- ▶ Séries geradas pela equação:

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t = 1, \dots, n$$

- ▶ Replicações: $R = 5000$
- ▶ Tamanhos amostrais: $n = \{50, 100, 300, 500\}$
- ▶ Estimativas obtidas: $\hat{\alpha}_{MQC}, \hat{\alpha}_{MVC}, \hat{\lambda}_{MQC}, \hat{\lambda}_{MVC}$
- ▶ Métricas: média, viés e EQM (na Binomial Negativa $\rightarrow \sigma^2$)



Cenários da Simulação

- ▶ **Coeficiente autorregressivo:** $\alpha = \{0.3, 0.5, 0.7\}$ (dependência fraca, moderada e forte).
Extras: $\alpha = 0.1, \lambda = 5$ e $\alpha = 0.9, \lambda = 4$.
- ▶ **Distribuições de inovação:** $\mu = \{1.5, 2.5\}$ (Poisson: $\mu = \lambda$).
- ▶ **Justificativa:** 5000 replicações garantem baixo erro padrão e robustez estatística, por causa da propriedade assintótica do estimador.



Resultados da Simulação

Foram avaliados os estimadores de Mínimos Quadrados Condicionais (MQC) e Máxima Verossimilhança Condicional (MVC) no processo INAR(1) com diferentes distribuições (Poisson, Binomial Negativa e Geométrica).

- ▶ MVC apresentou menor viés e melhor desempenho em quase todos os cenários;
- ▶ MQC foi razoável sob baixa dispersão, mas mais sensível à variabilidade;
- ▶ Em todos os cenários, a medida que o tamanho da amostra aumentava o valor do EQM convergia para zero, evidência da consistência do estimador;
- ▶ Binomial Negativa mostrou maior robustez entre os cenários analisados.



Resultados da Simulação

Tabela: Estimativas dos parâmetros, viés e erro quadrático médio quando $\alpha = 0,9$ e $\lambda = 4$.

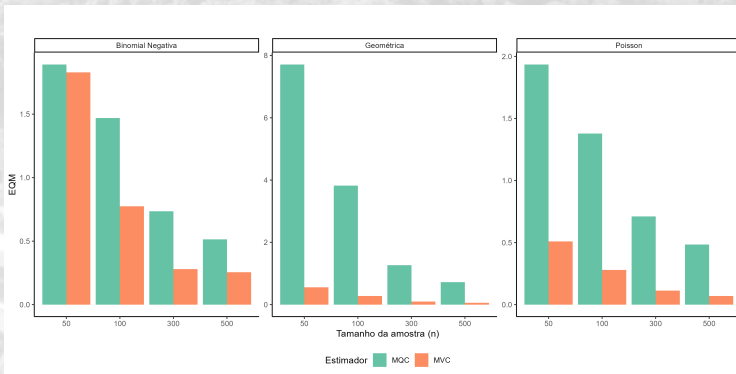
Amostra	Estimativa dos parâmetros											
	Poisson				Binomial Negativa				Geométrica			
	$\hat{\alpha}_{MQC}$	$\hat{\lambda}_{MQC}$	$\hat{\alpha}_{MVC}$	$\hat{\lambda}_{MVC}$	$\hat{\alpha}_{MQC}$	$\hat{\lambda}_{MQC}$	$\hat{\alpha}_{MVC}$	$\hat{\lambda}_{MVC}$	$\hat{\alpha}_{MQC}$	$\hat{\lambda}_{MQC}$	$\hat{\alpha}_{MVC}$	$\hat{\lambda}_{MVC}$
	viés											
$n = 50$	-0,0194	0,6253	-0,0027	0,0770	-0,0191	0,6211	0,0065	0,3938	-0,0454	1,4839	0,0000	-0,0198
$n = 100$	-0,0143	0,5137	-0,0016	0,0517	-0,0150	0,5393	0,0032	0,1635	-0,0271	0,9787	0,0000	-0,0087
$n = 300$	-0,0073	0,2786	-0,0006	0,0178	-0,0079	0,3020	0,0015	0,0217	-0,0108	0,4178	-0,0002	0,0051
$n = 500$	-0,0056	0,2136	-0,0004	0,0105	-0,0057	0,2225	0,0004	0,0733	-0,0070	0,2787	0,0000	0,0029
	erro quadrático médio											
$n = 50$	0,0019	1,9345	0,0005	0,5100	0,0018	1,8906	0,0010	1,8287	0,0066	7,7092	0,0002	0,5542
$n = 100$	0,0011	1,3793	0,0002	0,2789	0,0011	1,4687	0,0006	0,7740	0,0027	3,8210	0,0001	0,2771
$n = 300$	0,0005	0,7112	0,0001	0,1127	0,0005	0,7348	0,0002	0,2791	0,0008	1,2659	0,0000	0,0968
$n = 500$	0,0003	0,4846	0,0000	0,0688	0,0003	0,5139	0,0002	0,2555	0,0004	0,7191	0,0000	0,0573

Fonte: Elaboração Própria



Resultados da Simulação

Figura: Análise do EQM na simulação de monte carlo para $\hat{\lambda}$ quando $\lambda = 4$ e $\alpha = 0,9$.



Resultados da Simulação

- ▶ Ambos os estimadores (MQC e MVC) são consistentes; **MVC apresenta EQM menor**, especialmente na **distribuição geométrica**.
- ▶ **Aumento do tamanho amostral** reduz viés e variabilidade, confirmando propriedades assintóticas.
- ▶ **Comparativo de desempenho:**
 - ▶ Cenário 5, $n = 50$, geométrica: EQM $\hat{\lambda}_{MQC} = 7,7092$, EQM $\hat{\lambda}_{MVC} = 0,5542$ (**redução de 92,81%**).
 - ▶ Cenário 3, $n = 50$, Poisson: EQM $\hat{\lambda}_{MQC} = 1,2236$, EQM $\hat{\lambda}_{MVC} = 0,2511$ (**redução de 79,48%**).
- ▶ **Figura 1:** MVC mantém EQM menores para todas as distribuições e tamanhos amostrais, com destaque para amostras pequenas ($n = 50$).
- ▶ **Conclusão:** MVC é mais preciso e robusto, especialmente quando $\alpha \rightarrow 1$.



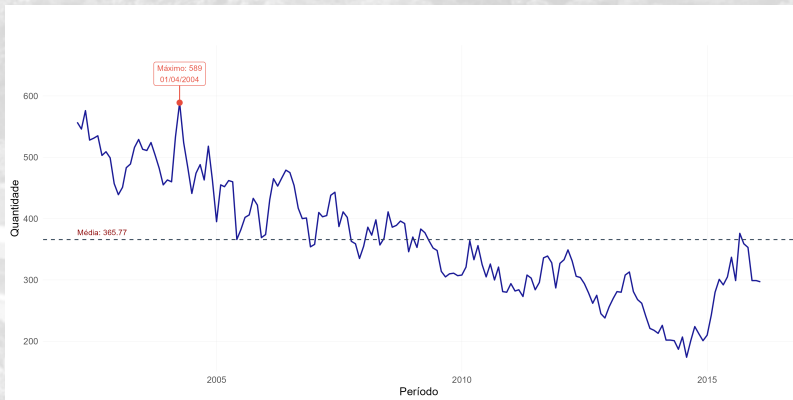
Aplicação com dados reais

- ▶ Banco de dados: série mensal da população desocupada na Região Metropolitana do Rio de Janeiro.
- ▶ Período de análise: março de 2002 a fevereiro de 2016, totalizando 168 observações.
- ▶ Fonte dos dados: **SGS - Banco Central do Brasil**.
- ▶ A variável em estudo é de natureza **discreta e de contagem**, em uma escala de milhares, logo os valores considerados na série devem ser multiplicados por mil.
- ▶ Este contexto motiva a utilização de um modelo INAR(1), capaz de lidar com dependência temporal em séries de contagem.



Aplicação com dados reais

Figura: Série temporal da população desocupada na região metropolitana do Rio de Janeiro, de março de 2002 à fevereiro de 2016, em milhares.



Análise Exploratória

- ▶ Média em torno de **366 mil habitantes**, com mediana de 357,5 mil.
- ▶ **Assimetria positiva leve** (0, 18), indicando maior peso à direita da distribuição.
- ▶ **Dispersão significativa**: desvio padrão de 95,26 mil.
- ▶ **Curtose platicúrtica** (2, 25), sugerindo caudas mais leves que a normal.
- ▶ Apesar da relativa estabilidade, há períodos de afastamento da média, possivelmente relacionados a **crises no mercado de trabalho**.



Estimação dos Parâmetros

- ▶ Foram aplicados dois métodos de estimação:
 - ▶ **MQC (Mínimos Quadrados Condicionais).**
 - ▶ **MVC (Máxima Verossimilhança Condicional).**
- ▶ Distribuições consideradas para a inovação:
 - ▶ Poisson.
 - ▶ Binomial Negativa.
 - ▶ Geométrica.
- ▶ Os resultados mostram que o estimador via **MVC apresentou melhor ajuste** em comparação ao MQC, especialmente nas distribuições mais flexíveis (binomial negativa e geométrica).



Estimação dos Parâmetros

Tabela: Estimativas dos parâmetros da série temporal de desocupados na região metropolitana do Rio de Janeiro, de março de 2002 à fevereiro de 2016, em milhares.

MQC		MVC					
		Poisson		Bin. Negativa		Geométrica	
$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$
0,9471	17,8003	0,7705	82,477	0,8724	85,436	0,8819	41,6999

Fonte: Elaboração Própria



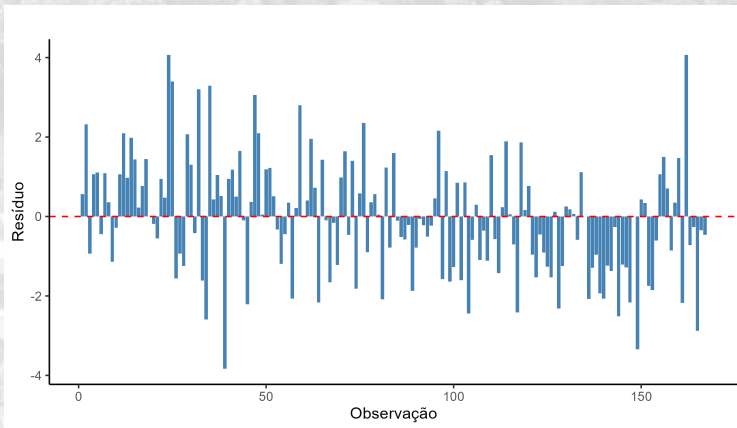
Ajuste do modelo

- ▶ Critérios de informação (**AIC e BIC**) indicaram como melhor ajuste o **modelo INAR(1) com inovação geométrica**, estimado via MVC.
- ▶ Esse modelo apresentou **boa aderência** aos dados e capacidade de capturar a dependência temporal.
- ▶ A análise dos resíduos mostrou ausência de **autocorrelação significativa**, reforçando a adequação do ajuste.
- ▶ Portanto, o modelo selecionado apresenta consistência após a submissão de testes estatísticos, como o de Ljung-Box e através das análises gráficas, pudemos analisar que os erros de pearson, se distribuem ao redor de zero, de forma aleatória.



Ajuste do modelo

Figura: Análise de resíduos da série da série modelada referente população desocupada na região metropolitana do Rio de Janeiro, de abril de 2002 à fevereiro de 2016, em milhares.



Predição com o modelo ajustado

- ▶ A previsão de curto prazo reproduziu de forma **fiel a dinâmica observada** na série original.
- ▶ Os resultados confirmam as evidências obtidas nas simulações, destacando a robustez do MVC em contextos de dependência elevada.
- ▶ Assim, o modelo ajustado fornece **ferramenta consistente** para análises e previsões de séries de contagem.



Predição com o modelo ajustado

Tabela: Medidas de qualidade do ajuste para as estimação da série de população desempregada na região metropolitana do Rio de Janeiro, de acordo com o horizonte de previsão.

h	MAE	RMSE	MPE
1	8,00	8,00	2,69%
3	54,67	54,67	18,33%
6	38,67	42,19	11,37%
12	30,67	34,83	9,40%

Fonte: Elaboração Própria



Previsão da série temporal

- ▶ Modelo ajustado: **INAR(1) com distribuição Geométrica**, selecionado via AIC e BIC.
- ▶ Horizonte de previsão: **12 meses** (mar/2016 a fev/2017).
- ▶ Valores previstos variam de **304 a 341 mil desocupados**.
- ▶ Incremento total de **37 mil pessoas (+12%)**.
- ▶ Crescimento **gradual e suavizado**, típico de processos INAR(1).

Tabela: Valores previstos para a série da população desocupada na região metropolitana do Rio de Janeiro, de março de 2016 à fevereiro de 2017, em milhares.

Período	Previsão
mar/2016	304
jun/2016	319
set/2016	330
dez/2016	337
fev/2017	341



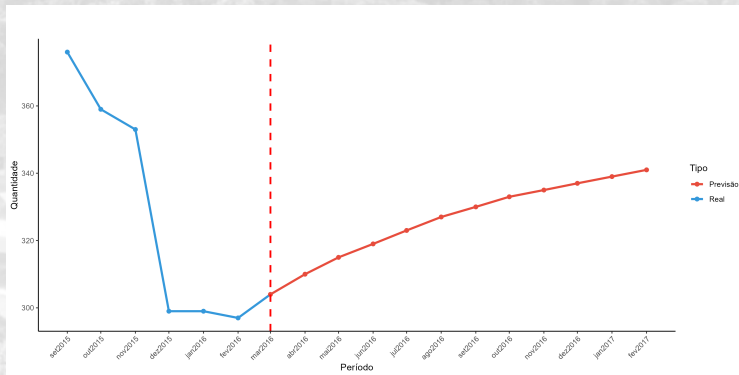
Previsão do série temporal

- ▶ Série tende a se estabilizar em torno de **341 mil unidades da população desocupada**.
- ▶ Intervalo de confiança de 95%: **303 a 350 mil**.
- ▶ Margem de incerteza cresce em horizontes longos (multiperíodo).
- ▶ Resultados sugerem um **cenário de piora progressiva** no mercado de trabalho.
- ▶ Implicações: necessidade de **políticas públicas** voltadas à geração de empregos e estímulo econômico.



Previsão da série temporal

Figura: Detalhado valores previstos para a série da população desocupada na região metropolitana do Rio de Janeiro, de março de 2016 à fevereiro de 2017, em milhares.



Considerações Finais

Objetivo do Trabalho

- ▶ Avaliar a adequação das distribuições de inovação no processo INAR(1).
- ▶ Estimar parâmetros α e λ e realizar previsões em série de contagem.
- ▶ Aplicação em dados reais: população desocupada na Região Metropolitana do Rio de Janeiro (mar/2002 – fev/2016).

Principais Resultados

- ▶ Estimadores via **MQC** e **MVC** mostraram-se eficazes e com propriedades assintóticas validadas.
- ▶ MVC apresentou maior robustez, mesmo em amostras pequenas.
- ▶ Previsões confiáveis obtidas após validação da independência dos resíduos.



Considerações Finais

Flexibilidade e Adequação do Modelo






- ▶ Seleção do modelo via **AIC e BIC**, garantindo melhor ajuste à série.
- ▶ Preservou a natureza **discreta** e **não negativa** da série de contagem.

Perspectivas e Contribuições

- ▶ Resultados fornecem insights estratégicos sobre a dinâmica do mercado de trabalho.
- ▶ Indicam necessidade de monitoramento contínuo e políticas públicas direcionadas.
- ▶ Futuras pesquisas podem explorar extensões do modelo (INAR(p), multivariados) e aplicações em outros contextos de séries de contagem.



Referências

-  FREELAND, R. Keith. **Statistical analysis of discrete time series with application to the analysis of workers compensation claims data**. 1998. Tese (Doutorado) – University of British Columbia.
-  MCKENZIE, Ed. Autoregressive moving-average processes with negative binomial and geometric marginal distributions. **Advances in Applied Probability**, v. 18, n. 3, p. 679–705, 1986.
-  MCKENZIE, Ed. Some simple models for discrete variate time series. **Water Resources Bulletin**, v. 21, n. 4, p. 645–650, 1985.
-  AL-OSH, Mohammed A.; ALZAID, Abul A. First-order integer-valued autoregressive (INAR(1)) process. **Journal of Time Series Analysis**, v. 8, n. 3, p. 261–275, 1987.
-  STEUTEL, Fred W.; HARN, Klaas van. Discrete analogues of self-decomposability and stability. **The Annals of Probability**, v. 7, n. 5, p. 893–899, 1979.



Comparação de desempenho e adequação de três processos para séries temporais INAR

Trabalho de Conclusão de Curso

Discente: Paulo Manoel da Silva Junior

Orientador: Prof^a Dr^a Tatiene Correia de Souza

www.de.ufpb.br

