

Correções do AIC para pequenas e grandes amostras em séries temporais: um estudo de simulação com AICc e CAIC

Mariana L. Garcia, Viviane L. D. de Mattos,
FURG - Programa de Pós Graduação em Modelagem Computacional,
95500-000, Rio Grande, RS
E-mail: marianalimagarcia@gmail.com, vivianeldm.furg@gmail.com

Resumo: A seleção de modelos em séries temporais costuma ser conduzida por critérios de informação, em especial pelo Critério de Informação de Akaike (AIC). No entanto, sua formulação apresenta limitações em amostras pequenas e não garante consistência assintótica em amostras grandes, o que motivou o desenvolvimento de correções como o critério de Akaike corrigido (AICc) e o critério de Akaike consistente (CAIC). Este trabalho discute o papel desses três critérios na seleção de modelos de séries temporais e investiga empiricamente seu desempenho sob diferentes tamanhos amostrais. Foi conduzido um estudo de simulação de Monte Carlo em cenário de mundo especificado, no qual séries temporais foram geradas por um modelo ARMA(1,1) estacionário, sem tendência, e ajustadas por uma família de modelos AR, MA e ARMA com ordens entre 1 e 4. Para cada tamanho amostral, comparou-se a proporção de vezes em que cada critério selecionou o modelo verdadeiro. Em amostras pequenas e moderadas, o AICc apresentou maior taxa de acerto do que o AIC e o CAIC; em amostras grandes, o CAIC passou a dominar, aproximando-se de probabilidade 1 de recuperação do modelo verdadeiro, enquanto AIC e AICc se estabilizaram em patamares inferiores. Os achados reforçam que a escolha do critério de informação em séries temporais deve considerar explicitamente o tamanho amostral, evitando o uso automático do AIC como padrão.

Palavras-chave: *Critérios de informação, AIC, AICc, CAIC, Séries temporais, Seleção de modelos.*

1 Introdução

A seleção de modelos é uma etapa central em análises estatísticas aplicadas, especialmente quando se trabalha com modelos paramétricos ajustados a dados observacionais. Em vez de buscar um “modelo verdadeiro” em sentido estrito, costuma-se escolher, dentro de uma classe de modelos candidatos, aquele que fornece uma representação adequada da estrutura dos dados e boa capacidade preditiva. Nesse contexto, os critérios de informação se consolidaram como ferramentas de uso corrente, por incorporarem explicitamente o compromisso entre ajuste e complexidade do modelo com base em fundamentos da teoria da informação e da divergência de Kullback–Leibler.

Entre esses critérios, o Critério de Informação de Akaike (AIC) ocupa posição destacada desde a década de 1970, sendo amplamente utilizado em áreas como econometria, epidemiologia, ecologia e ciência de dados. Sua formulação assintótica, entretanto, apresenta limitações bem documentadas: em amostras pequenas, o AIC tende a favorecer modelos excessivamente complexos; em amostras grandes, não garante consistência assintótica, podendo manter preferência por modelos mais complexos mesmo quando o modelo verdadeiro está entre os candidatos. Essas limitações motivaram o desenvolvimento de desdobramentos orientados a regimes amostrais distintos, entre os quais se destacam o AICc, voltado a amostras pequenas, e o CAIC, que intensifica a penalização em função do tamanho amostral.

No contexto de séries temporais, essas questões assumem relevância particular. Modelos como AR, MA, ARMA e ARIMA são amplamente empregados para descrever a dinâmica temporal de dados em diversas áreas, e a escolha de suas ordens costuma ser feita por meio de critérios de informação. Em revisão anterior da literatura, baseada em buscas recentes na base Scopus, observou-se que o AIC segue sendo o critério predominante na seleção de modelos de séries temporais, especialmente em aplicações com modelos ARIMA e SARIMA. Estudos como [10], na previsão de emissões de CO_2 , [9], na modelagem da atividade sazonal de gripe, adotam o AIC como critério principal de escolha do modelo, em geral sem discutir de forma explícita por que esse critério é mais adequado do que alternativas, nem como sua escolha se relaciona ao tamanho da série ou aos objetivos específicos da análise. Mesmo quando outros critérios são considerados em conjunto, como o Bayesiano (BIC) ou o Hannan–Quinn (HBC), essa multiplicidade costuma ser apresentada como procedimento padrão, sem discussão das diferenças conceituais entre eles ou das implicações práticas dessa escolha.

De forma semelhante, quando o critério de informação de Akaike corrigido (AICc) e o critério de Akaike consistente (CAIC) aparecem em aplicações recentes, seu uso tende a ser pontual e pouco sistematizado. No estudo [5], por exemplo, o AICc é considerado em conjunto com o AIC e o BIC na seleção de modelos ARIMA, sem que se explicita claramente o papel específico de cada critério ou as razões para combiná-los. Já em [6], o CAIC é utilizado na escolha de um modelo SARIMA de referência, novamente ao lado de outros critérios, mas sem justificativa metodológica mais detalhada para sua adoção. Esse padrão reforça a ideia de que, embora as correções do AIC sejam bem estabelecidas na literatura teórica, ainda há um descompasso entre essas recomendações e a forma como os critérios de informação são empregados na prática aplicada em séries temporais.

Diante desse cenário, o objetivo deste trabalho é discutir o papel do AIC e de suas correções AICc e CAIC na seleção de modelos de séries temporais em diferentes regimes amostrais, combinando uma síntese conceitual com evidências empíricas obtidas por simulação. A Seção 2 revisa brevemente os critérios de informação e a formulação do AIC; a Seção 3 apresenta o AICc e o CAIC como correções do AIC para pequenas e grandes amostras em séries temporais; a Seção 4 descreve um estudo de simulação de Monte Carlo em cenário de mundo especificado; e a Seção 5 discute os resultados obtidos e suas implicações práticas.

2 Critérios de Informação e o AIC

Critérios de seleção de modelos são utilizados para escolher, entre diferentes modelos candidatos, aqueles que apresentam uma combinação adequada entre qualidade de ajuste e complexidade estrutural. A ideia central é favorecer modelos que generalizem bem para novos dados, evitando tanto estruturas excessivamente simples quanto modelos demasiadamente complexos, mais propensos ao sobreajuste [3].

A fundamentação teórica desses critérios está ligada à divergência de Kullback–Leibler, que mede a separação entre a distribuição verdadeira geradora dos dados, $g(y)$, e a distribuição estimada por um modelo $f(y|\theta)$ [4], [3]. Como $g(y)$ é desconhecida, a divergência é aproximada a partir da amostra, combinando um termo de ajuste, usualmente baseado na log-verossimilhança maximizada, com uma penalização da complexidade do modelo. Em termos gerais, um critério de informação para um modelo j (M_j) pode ser expresso como uma soma de dois componentes [4]:

$$IC(M_j) = \text{Termo de ajuste}(M_j) + \text{Penalidade}(M_j), \quad (1)$$

em que diferentes critérios se distinguem sobretudo pela forma da penalização.

O Critério de Informação de Akaike (AIC) foi o primeiro a ganhar ampla difusão e permanece como referência central nesse campo. Proposto originalmente nos anos 1970, o AIC reinterpretou o princípio da máxima verossimilhança sob a ótica da teoria da informação, mostrando que é

possível construir um estimador assintoticamente eficiente da divergência de Kullback–Leibler esperada entre o modelo ajustado e a distribuição verdadeira dos dados (AKAIKE, 1998). Em termos práticos, para um modelo estimado por máxima verossimilhança, o AIC assume a forma

$$AIC = -2 \log L(\hat{\theta}) + 2k, \quad (2)$$

em que $L(\hat{\theta})$ é o valor maximizado da função de verossimilhança e k representa o número de parâmetros livres do modelo. O termo $-2 \log L(\hat{\theta})$ atua como medida de falta de ajuste, enquanto a penalização $2k$ corrige o viés decorrente do uso da log-verossimilhança maximizada como estimador da discrepância entre f e g . Sob condições regulares e para tamanhos amostrais grandes, o AIC fornece um estimador assintoticamente eficiente da divergência de Kullback–Leibler esperada. Entretanto, em amostras pequenas o critério tende a favorecer modelos excessivamente complexos, pois a penalização constante $2k$ não compensa adequadamente a variabilidade do termo de ajuste quando n é reduzido. Além disso, o AIC não é consistente em amostras grandes, podendo manter preferência por modelos mais complexos mesmo quando o modelo verdadeiro está entre os candidatos. Essas limitações motivaram o desenvolvimento de correções específicas, como o AICc e o CAIC, que ajustam o comportamento do AIC em regimes de amostras pequenas e grandes.

3 Correções do AIC para pequenas e grandes amostras em séries temporais

Em muitas aplicações, os critérios de informação são utilizados para selecionar modelos de séries temporais, como AR, MA e ARMA, ajustados a observações ordenadas no tempo. Nesses casos, a dependência temporal entre as observações impõe desafios adicionais à modelagem e torna particularmente relevante a escolha de modelos com bom desempenho preditivo [8]. A seleção de ordens em modelos ARMA e ARIMA costuma ser feita por critérios como o AIC, mas o comportamento desse critério depende fortemente do tamanho amostral e da relação entre o número de parâmetros e o número de observações.

O AICc surgiu como uma correção explícita do AIC para amostras pequenas, inicialmente proposta no contexto de modelos de regressão e posteriormente estendida a modelos de séries temporais [3]. Sua derivação baseia-se em uma correção de ordem finita para o viés da log-verossimilhança maximizada, de modo a melhorar a aproximação da divergência de Kullback–Leibler esperada quando o tamanho da amostra é limitado. Para um modelo com k parâmetros livres ajustado a uma amostra de tamanho n , o AICc pode ser escrito como

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}, \quad (3)$$

em que o termo adicional cresce à medida que a razão k/n se aproxima de 1, aumentando a penalização em situações em que o número de parâmetros é grande em relação ao tamanho da amostra. Quando n é grande, o AICc converge para o AIC, mantendo assim a eficiência assintótica do critério original [4]. Em séries temporais curtas, nas quais é comum trabalhar com modelos relativamente parametrizados (como AR(p), MA(q) ou ARMA(p,q) com poucos pontos disponíveis), o AICc tem sido recomendado por reduzir o sobreajuste observado no AIC [7].

O CAIC, por sua vez, foi proposto por [2] como uma modificação do AIC orientada a amostras grandes, substituindo a penalização constante $2k$ por um termo que cresce com o logaritmo do tamanho amostral. Para um modelo ajustado por máxima verossimilhança, o critério assume a forma

$$CAIC = -2 \log L(\hat{\theta}) + k[(\log n) + 1], \quad (4)$$

em que $L(\hat{\theta})$ é o valor da função de verossimilhança maximizada, k é o número de parâmetros livres no modelo e n é o tamanho da amostra. A penalização extra visa garantir consistência assintótica sob condições regulares, ou seja, conforme n cresce, a probabilidade de selecionar o modelo verdadeiro converge para 1. Em modelos de séries temporais, essa característica é particularmente relevante quando se dispõe de séries longas e de um conjunto amplo de modelos candidatos, pois penalizações mais severas reduzem o risco de escolher modelos mais complexos que capturam ruído em vez da estrutura essencial da série.

Na prática aplicada, o uso do AICc e do CAIC em séries temporais é bem menos frequente do que o do AIC, e muitas vezes a escolha do critério não vem acompanhada de justificativas explícitas em termos de tamanho amostral ou de objetivos da análise. Neste trabalho, esses dois critérios são considerados justamente como correções do AIC voltadas a regimes amostrais distintos: o AICc, ajustando o comportamento do critério em amostras pequenas, e o CAIC, intensificando a penalização em amostras grandes. A simulação apresentada na Seção 4 explora como essas diferenças de penalização se traduzem, na prática, em diferentes probabilidades de selecionar o modelo verdadeiro em séries ARMA geradas sob um cenário controlado.

4 Estudo de Simulação

Com o objetivo de comparar o desempenho do critério de informação AIC e de suas correções AICc e CAIC na seleção de modelos de séries temporais sob diferentes tamanhos amostrais, foi conduzido um estudo de simulação de Monte Carlo em um cenário de mundo especificado, isto é, em que o modelo verdadeiro pertence à classe de modelos candidatos. O processo gerador dos dados foi um modelo ARMA(1,1) estacionário, sem termo constante, portanto de média zero, e sem tendência determinística, com coeficientes 0,6 (componente autorregressiva) e 0,4 (componente de médias móveis) e erros distribuídos como ruído branco normal padrão, conforme

$$X_t = 0,6X_{t-1} + \varepsilon_t + 0,4\varepsilon_{t-1} \quad , \quad \varepsilon_t \sim N(0,1) \quad (5)$$

Em cada repetição da simulação, gerou-se inicialmente uma sequência de inovações de comprimento suficiente para o maior tamanho amostral considerado, acrescido de um *burn-in* de 200 observações, que foi descartado para reduzir o efeito das condições iniciais. Para isolar o efeito do tamanho da amostra, utilizou-se a técnica de números aleatórios comuns: para cada repetição foi fixada uma sequência de inovações, reutilizada em todos os valores de n .

Foram considerados tamanhos amostrais variando de 25 a 1000 observações, abrangendo séries curtas, médias e longas. Para cada um desses valores, geraram-se 1000 repetições independentes do processo ARMA(1,1). Em cada série simulada ajustaram-se, por máxima verossimilhança, diferentes modelos ARMA sem constante. O conjunto de modelos candidatos incluiu estruturas autorregressivas (AR), de médias móveis (MA) e ARMA, com ordens variando entre 1 e 4 em cada componente, de forma que o modelo verdadeiro ARMA(1,1) estivesse entre os candidatos, juntamente com modelos mais simples e mais complexos.

Para cada modelo ajustado, calcularam-se os valores de AIC, AICc e CAIC a partir da log-verossimilhança maximizada e do número de parâmetros. Em cada repetição, e para cada critério, o modelo selecionado foi aquele que apresentou o menor valor do respectivo critério. A medida principal de desempenho foi a taxa de acerto na seleção do modelo verdadeiro: para cada critério e tamanho amostral, registrou-se, em cada repetição, se o modelo escolhido correspondia ao ARMA(1,1), estimando-se a proporção de acertos como a média dessas indicações nas 1000 repetições. Intervalos de confiança de 95% para essas proporções, baseados na distribuição binomial, foram utilizados na construção das curvas apresentadas na Figura 1, que mostram a evolução da taxa de acerto em função de n : o painel (a) apresenta todo o intervalo de tamanhos amostrais e o painel (b) exibe um zoom para $n \leq 100$.

A Figura 1 evidencia três regimes distintos. Em tamanhos amostrais pequenos, o AICc apresenta desempenho superior aos demais critérios, sugerindo que a correção para amostras

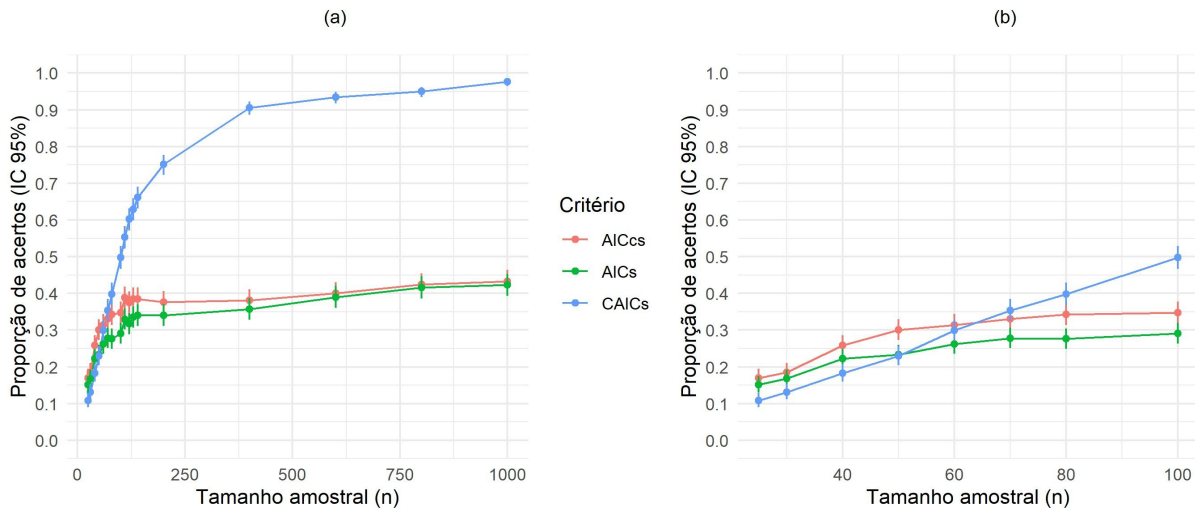


Figura 1: Proporção de acertos por critério e tamanho amostral: (a) intervalo completo de n ; (b) zoom para $n \leq 100$.

pequenas efetivamente reduz o sobreajuste e aumenta a probabilidade de seleção do modelo verdadeiro. Em tamanhos intermediários, as taxas de acerto de AICc e CAIC tornam-se próximas, e o CAIC passa a superar gradualmente o AICc à medida que n cresce. Nos tamanhos amostrais grandes, o comportamento do CAIC torna-se marcadamente distinto, superando o AIC e o AICc e atingindo proporções de acerto próximas de 1, padrão coerente com suas propriedades de consistência assintótica.

Além da análise marginal das taxas de acerto, foram realizadas comparações pareadas entre os critérios por meio do teste de McNemar. Para cada tamanho amostral e para cada par (AIC vs AICc, AIC vs CAIC e AICc vs CAIC), construiu-se uma tabela 2x2 com as contagens de repetições em que ambos acertaram, apenas o primeiro acertou, apenas o segundo acertou ou ambos erraram. De forma geral, as diferenças entre as taxas de acerto mostraram-se estatisticamente significativas ao nível de 5% na maior parte dos tamanhos amostrais: para AICc versus AIC, apenas $n = 30$ não apresentou evidência de diferença; para AIC versus CAIC, a não significância ocorreu apenas em $n = 50$; e, para AICc versus CAIC, as diferenças deixaram de ser significativas apenas em $n = 60$ e $n = 70$. Esses resultados reforçam que as curvas observadas na Figura 1 não são meras flutuações aleatórias.

5 Discussão e Considerações Finais

Os resultados da simulação ilustram como as diferenças na estrutura de penalização do AIC, AICc e CAIC se traduzem em comportamentos distintos na seleção de modelos de séries temporais. Em amostras pequenas, o AICc apresentou maior probabilidade de recuperar o modelo ARMA(1,1) verdadeiro do que o AIC e o CAIC, sugerindo que a correção adicional em função da razão k/n reduz a tendência ao sobreajuste observada no AIC. Em amostras grandes, o CAIC aproximou-se rapidamente de taxas de acerto próximas de 100%, enquanto AIC e AICc estabilizaram em níveis bem mais modestos, coerentemente com a ideia de consistência assintótica associada à penalização mais severa do CAIC.

Do ponto de vista prático, esses resultados indicam que a escolha do critério de informação não é neutra no contexto de séries temporais. Em séries curtas ou moderadas, o uso do AICc parece mais alinhado ao objetivo de controlar o sobreajuste e aumentar a chance de recuperar estruturas dinâmicas adequadas, enquanto em séries mais longas, com muitos modelos candidatos, o CAIC se mostra mais apropriado quando a parcimônia e a identificação do modelo verdadeiro são prioridades. Em contraste, o uso quase automático do AIC, frequentemente observado na

literatura aplicada, tende a ignorar o papel do tamanho amostral e pode levar, em ambos os regimes, a decisões menos favoráveis do que aquelas obtidas com seus refinamentos. Nesse sentido, os achados deste estudo dialogam criticamente com o padrão identificado em trabalhos recentes, nos quais o AIC é adotado como critério padrão e o AICc e o CAIC aparecem apenas de forma pontual e pouco justificada.

Essas conclusões são obtidas em um cenário controlado e relativamente simples: séries geradas por um único modelo ARMA(1,1) estacionário, sem tendência e com erros gaussianos, e um conjunto limitado de modelos candidatos. Outros cenários, que envolvam má especificação do modelo, estruturas sazonais, componentes determinísticos mais complexos ou distribuições de erro não normais, podem produzir padrões diferentes e merecem investigação específica. Ainda assim, o estudo evidencia que as propriedades teóricas de AICc e CAIC têm consequências práticas relevantes na seleção de modelos de séries temporais.

Em síntese, o trabalho reforça a necessidade de escolhas mais conscientes de critérios de informação. Em vez de adotar o AIC como padrão por inércia, recomenda-se considerar explicitamente o tamanho amostral e os objetivos da análise: AICc como opção preferencial em amostras pequenas e moderadas, CAIC em amostras grandes nas quais a parcimônia seja central. Estudos futuros podem estender esse tipo de comparação para outros cenários de modelagem e para critérios adicionais, contribuindo para aproximar a prática aplicada das recomendações teóricas disponíveis na literatura e reduzir o descompasso entre o que se conhece sobre esses critérios e a forma como são efetivamente utilizados.

Referências

- [1] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: Selected papers of Hirotugu Akaike. New York: Springer, 1998. p. 199–213.
- [2] BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, v. 52, n. 3, p. 345–370, 1987.
- [3] CAVANAUGH, J. E.; NEATH, A. A. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 11, n. 3, e1460, 2019.
- [4] KONISHI, S.; KITAGAWA, G. Information criteria and statistical modeling. New York: Springer Science & Business Media, 2008.
- [5] LI, Y. et al. Impact of COVID-19 on epidemic trend of hepatitis C in Henan Province assessed by interrupted time series analysis. *BMC Infectious Diseases*, v. 23, n. 1, p. 691, 2023.
- [6] LV, C.-X. et al. Time series analysis of hemorrhagic fever with renal syndrome in mainland China by using an XGBoost forecasting model. *BMC Infectious Diseases*, v. 21, n. 1, p. 839, 2021.
- [7] MCQUARRIE, A. D. R.; TSAI, C.-L. Regression and time series model selection. Singapore: World Scientific, 1998.
- [8] MORETTIN, P. A.; TOLOI, C. M. C. Análise de séries temporais: modelos lineares univariados. São Paulo: Editora Blucher, 2018.
- [9] ORANG, A. et al. Forecasting seasonal influenza activity in Canada—Comparing seasonal Auto-Regressive integrated moving average and artificial neural network approaches for public health preparedness. *Zoonoses and Public Health*, v. 71, n. 3, p. 304–313, 2024.
- [10] REHMAN, M. Z.; DAR, A. A.; WANGMO, T. Forecasting CO₂ emissions in India: A time series analysis using ARIMA. *Processes*, v. 12, n. 12, p. 2699, 2024.