



UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA

ANDRESSA DE SOUZA FREITAS

**ROBUSTEZ DO TESTE MONTE CARLO PARA TENDÊNCIA ESTOCÁSTICA NA
FAMÍLIA LOCAÇÃO-ESCALA**

Ouro Preto – MG

2020

ANDRESSA DE SOUZA FREITAS

**ROBUSTEZ DO TESTE MONTE CARLO PARA TENDÊNCIA
ESTOCÁSTICA NA FAMÍLIA LOCAÇÃO-ESCALA**

Monografia de Graduação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador: Prof. Dr. Ivair Ramos Silva

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA
OURO PRETO – MG

2020

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

F866r Freitas, Andressa de Souza .
Robustez do teste Monte Carlo para tendência estocástica na família
locação-escala. [manuscrito] / Andressa de Souza Freitas. - 2020.
25 f.: il.: gráf., tab.. + Equação.

Orientador: Prof. Dr. Ivair Ramos Silva.
Monografia (Bacharelado). Universidade Federal de Ouro Preto.
Instituto de Ciências Exatas e Biológicas. Graduação em Estatística .

1. Monte Carlo, Método de. 2. Processo estocástico. 3. Análise de
séries temporais. 4. Estatística - Testes . I. Silva, Ivair Ramos. II.
Universidade Federal de Ouro Preto. III. Título.

CDU 519.2

Bibliotecário(a) Responsável: Celina Brasil Luiz - CRB6 - 1589



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA

**FOLHA DE APROVAÇÃO****Andressa de Souza Freitas****Robustez do Teste Monte Carlo para Tendência Estocástica na Família Localização-Escala**

Membros da banca

Fernando Luiz Pereira de Oliveira - Doutor - UFOP
Ivair Ramos Silva - Doutor - UFOP
Rivert Paulo Braga Oliveira - Doutor - UFOP

Versão final

Aprovado em 21 de outubro de 2020

De acordo

Ivair Ramos Silva



Documento assinado eletronicamente por **Ivair Ramos Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 23/10/2020, às 00:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0095211** e o código CRC **90A517C1**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.008041/2020-93

SEI nº 0095211

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

AGRADECIMENTOS

Agradeço à minha família: mãe Liene Lima de Souza Freitas, pai Dewis Tette de Freitas, irmãos Kristtian Souza e Dawis Tadeu Souza, às minhas sobrinhas queridas Vitória, Ágatha e Samira que, apesar de toda dificuldade, sempre estiveram ao meu lado e são a minha base. Às minhas cunhadas Renata e Verônica, por terem me dado os melhores presentes do mundo que são as minhas sobrinhas. À minha irmã de coração Viviane Aparecida Souza por estar ao meu lado e por sempre me apoiar nos piores momentos e comemorar comigo os melhores. Aos demais familiares que, apesar da distância, guardo com carinho no coração.

Agradeço a todos os professores que fizeram parte da minha formação e ao DEEST como um todo. Em especial ao meu orientador Dr. Ivair Ramos Silva, por toda paciência e cuidado comigo, desde às disciplinas que lecionou para mim até o momento que começou a me orientar na Iniciação Científica e na Monografia. A empatia e profissionalismo de todos os professores foi crucial para que eu chegasse até aqui, tenho muita admiração e gratidão por todos.

Agradeço ao meu namorado Adriano Chamon Tavares por ser luz na minha vida, por ser meu companheiro de todas as horas, por me fortalecer em cada momento de fraqueza e por fazer eu me sentir a pessoa mais amada e importante do mundo.

Aos meus amigos e colegas de curso, em especial a Joziani Mota, Marcos Vinícius, Samuel Sabino e Milele Cássia. São pessoas incríveis que adorei conhecer e foram as pessoas que deixaram a minha graduação mais leve. São companheiros que fiz para vida toda e sempre os guardarei com todo amor do mundo no meu coração. Sem eles, também, eu não teria chegado até aqui.

Agradeço à República Rodoviária por ter me acolhido tão bem e ter sido o meu lar nesses últimos 2 anos. Considero cada um dos moradores e pessoas presentes nessa casinha como parte da minha família, meus irmãos, guardarei para sempre em meu coração.

Agradeço aos meus queridos companheiros de jornada, pessoas que são muito importantes para mim e não poderia jamais deixar de citá-los aqui: Edward Armache, Mônica Chamon, Júlio Tavares, Samuel Chamon, Ivani Souza, e outros mais que me ajudaram nessa caminhada.

E por último, mas com igual importância, à Estatís Jr, empresa que me ajudou a me tornar a profissional que sou hoje.

“É muito melhor lançar-se em busca de conquistas grandiosas, mesmo expondo-se ao fracasso, do que alinhar-se com os pobres de espírito, que nem gozam muito e nem sofrem muito, porque vivem numa penumbra cinzenta, onde não conhecem nem vitória, nem derrota.” – Theodore Roosevelt

FREITAS, Andressa de Souza. **Robustez do Teste Monte Carlo para Tendência Estocástica na Família Locação-Escala**; Ouro Preto, 2020. Trabalho de Conclusão de Curso (Bacharel em Estatística) - Universidade Federal de Ouro Preto, Ouro Preto, 2020.

RESUMO

Um dos modelos mais utilizados em análise de séries temporais é o modelo de equações estruturais. Sob tal abordagem, um processo estocástico é descrito por meio de um funcional linear composto por termos de interpretação prática, tais como nível, tendência, sazonalidade, e qualquer outro componente cíclico de interesse. Em 1983, Nyblom e Makelainen propuseram um teste localmente mais poderoso para testar se o termo de nível de um processo é determinístico ou estocástico, chamado Teste NM. Esse teste foi aperfeiçoado por FRANCO et al, em 2009, através do método de *bootstrap*. Dulcília et al. (2016) estenderam o teste NM para testar se o termo de tendência é determinístico ou estocástico. Para tanto, eles construíram uma versão Monte Carlo para o teste NM. Como demonstrado formalmente por Silva et al. (2020), o teste proposto por Dulcília et al. (2016) garante o controle da probabilidade do Erro Tipo I sob o nível nominal desejado pelo analista. Isto é válido para qualquer processo estocástico com distribuição pertencente à família locação-escala. Este trabalho visa verificar a robustez do teste proposto por Silva et al. (2020) para séries temporais que não pertençam à família locação-escala. Esta verificação se deu através do estudo do Poder do Teste e na verificação do controle da probabilidade do Erro Tipo I, o que foi possível pelo método de Simulação de Monte Carlo. Foram simuladas séries com distribuição Qui-quadrado e Binomial. Com isso, foi possível estudar o desempenho estatístico do teste de Silva et al. (2020) para situações em que os dados não pertencem à família locação-escala.

Palavras-chave: Simulação Monte Carlo; Família locação-escala; Termo de Tendência; Séries Temporais; Teste NM.

FREITAS, Andressa de Souza. **Robustez do Teste Monte Carlo para Tendência Estocástica na Família Localização-Escala**; Ouro Preto, 2020. Trabalho de Conclusão de Curso (Bacharel em Estatística) - Universidade Federal de Ouro Preto, Ouro Preto, 2020.

ABSTRACT

One of the best models utilized in temporal series analysis it's the model of structural equations. Upon this approach, a stochastic process is described through a functional linear composed by practical interpretations terms, which include level, tendency, seasonality and any other cyclical component of interest. In 1983, Nyblom and Makelainen proposed a locally more powerful test for testing if level term of a process is deterministic or stochastic, called NM Test. This test was perfected by FRANCO et al., in 2009, through the bootstrap method. Dulcília et al. (2016) extended the NM test to verify if the tendency term is deterministic or stochastic. To do so, they elaborated a version of Monte Carlo to the NM test. Demonstrated formally by Silva et al. (2020), the test proposed by Dulcília et al. (2016) guarantee the control of the probability of Type I Error at the nominal level required by the analyst. This is valid to any stochastic process with a distribution belonging to the family location-scale. This work plans to verify the robustness of the test proposed by Silva et al. (2020) to temporal series that do not belong to the location-scale family. This inspection was brought through a study of the Power of the Test and in the ratification of the control of probability of the Type I Error, which was possible due to the Simulation of Monte Carlo method. Series with Chi-square and Binomial were simulated. With that, it was possible study the statistic performance of the test of Silva et al. (2020) for situations which the data do not belong to the location-scale family.

Keywords: Simulation of Monte Carlo method; location-scale family; Trend term; time series; NM Test.

LISTA DE TABELAS

Tabela 1: Comparação dos Poderes do teste para cada Cenário, Normal versus Qui-quadrado, com $\alpha = 0,01$	17
Tabela 2: Comparação dos Poderes do teste para cada Cenário, Normal versus Qui-quadrado, com $\alpha = 0,05$	18
Tabela 3: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Qui-quadrado, para alpha igual a 0,01.....	18
Tabela 4: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Qui-quadrado, para alpha igual a 0,05.....	19
Tabela 5: Comparação dos Poderes do teste para cada Cenário, Normal versus Binomial, com $\alpha = 0,01$	19
Tabela 6: Comparação dos Poderes do teste para cada Cenário, Normal versus Binomial, com $\alpha = 0,05$	20
Tabela 7: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Binomial, para alpha igual a 0,01.....	20
Tabela 8: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Binomial, para alpha igual a 0,05.....	20

LISTA DE FIGURAS

Figura 1:Consumo de energia elétrica, em valores monetários (R\$), no estado do Espírito Santo, de janeiro de 1968 a setembro de 1979	22
Figura 2: Histograma da série de consumo de energia elétrica, em valores monetários (R\$), no estado do Espírito Santo, de janeiro de 1968 a setembro de 1979.....	23

SUMÁRIO

1. INTRODUÇÃO	10
2. OBJETIVO	13
3. MATERIAL E MÉTODOS	14
3.1. Poder do Teste Monte Carlo para o Ruído do Termo de Tendência	15
3.2. Simulação de Monte Carlo.....	16
4. RESULTADOS E DISCUSSÃO	17
4.1. Comparação dos poderes: Normal versus Qui-quadrado.....	17
4.2. Comparação dos poderes: Normal versus Binomial.....	19
4.3. Exemplo Prático	21
5. CONCLUSÕES	24
6. REFERÊNCIAS BIBLIOGRÁFICAS	25
SITES ELETRÔNICOS	25

1. INTRODUÇÃO

Uma série temporal é composta de uma média (μ_t) e um ruído (ϵ_t) que, usualmente, é conhecido como ruído branco. Também é comum supor que ϵ_t segue uma distribuição Normal ($0, \sigma_\epsilon^2$). Mas, μ_t depende de seus valores passados ($\mu_{t-1}, \mu_{t-2}, \dots$), mais um termo de tendência (β_t), e um ruído branco (η_t) somado ao termo de tendência. O ruído η_t também é usualmente modelado de acordo com uma Normal ($0, \sigma_\eta^2$). Porém, β_t também depende de seu passado ($\beta_{t-1}, \beta_{t-2}, \dots$) mais um erro (ξ_t) que, como nos outros casos, segue uma Normal ($0, \sigma_\xi^2$). A estrutura descrita pode ser representada da seguinte forma:

$$\mathbf{Y}_t = \mu_t + \epsilon_t, \epsilon_t \sim \text{Normal}(0, \sigma_\epsilon^2) \quad (1.1)$$

$$\mu_t = \mu_{t-1} + \beta_t + \eta_t, \eta_t \sim \text{Normal}(0, \sigma_\eta^2) \quad (2.2)$$

$$\beta_t = \beta_{t-1} + \xi_t, \xi_t \sim \text{Normal}(0, \sigma_\xi^2) \quad (3.3)$$

Testes estatísticos tem sido desenvolvidos com o intuito de estudar o comportamento de séries temporais não estacionárias. Segundo Ernesto et al. (2016), é importante definir se o termo de tendência é determinístico ou estocástico, visto que as previsões feitas por estas diferentes abordagens geram valores e interpretações distintas.

Sendo assim, afirma Ernesto et al. (2016), Nyblom e Makelainen (1983) criaram uma estatística de teste, NM, localmente mais poderosa, com o objetivo de testar se a variância do ruído do termo de nível é determinístico ou estocástico. O teste propunha o uso de aproximação assintótica da distribuição da estatística NM sob a hipótese nula de não aleatoriedade do termo de nível (μ_t). Porém, o teste NM não oferece um real controle da probabilidade do Erro Tipo I, principalmente, em boa parte das aplicações reais, com tamanhos de amostras pequenas e moderadas, Ernesto et al. (2016). Além disso, o teste apresentava limitações, tais como:

- Ser aplicável apenas para o termo de nível (μ_t);
- Necessitar da suposição de normalidade dos erros;
- A aproximação assintótica da estatística NM sob H_0 é de manipulação não amigável;
- Poder elevar a probabilidade do Erro Tipo I (FRANCO et al, 2009).

FRANCO et al (2009) introduziram uma proposta *bootstrap* eficiente para testar o termo de nível. FRANCO et al (2009) compararam o poder do uso do *bootstrap* com o da aproximação assintótica e observaram que o *bootstrap* apresentou um resultado mais satisfatório com relação ao nível de significância, ou seja, a probabilidade do Erro Tipo I tendia a ser próximo do desejado, diferentemente da aproximação assintótica que não tinha esse controle. Mas, segundo Ernesto et al. (2016), assim como o teste de Nyblom e Makelainen, o teste proposto por FRANCO et al (2009), também possuía limitações:

- Foi desenvolvido, exclusivamente, para testar o termo de nível;
- Requeria a estimação de todos os parâmetros do modelo e;
- O NM *bootstrap*, aplicado para testar o termo de tendência, não garante o real controle da probabilidade do Erro Tipo I, ou seja, a probabilidade de erro seria maior do que o α desejado.

Levando-se em consideração as limitações existentes nas propostas atuais, Ernesto et al. (2016) construiu um teste Monte Carlo para a variância do termo de nível (μ_t) e de tendência (β_t) que:

- Garante o nível de significância nominal exato;
- Não depende da suposição de normalidade;
- Não requer a estimação do modelo estrutural como um todo.

Segundo Ernesto et al. (2016), os testes de estocasticidade dos termos de nível e de tendência tinham outros aspectos importantes a serem considerados, como:

- A formulação de estatísticas de teste para os dois termos (de nível e de tendência), que não demandavam a estimação dos parâmetros do modelo;
- O controle analítico da probabilidade do Erro Tipo I para os ruídos do termo de tendência;
- A formulação de testes para casos de ruídos dos termos não observáveis que não assumiam distribuição normal e;
- A possibilidade das variâncias dos ruídos, sob H_1 , serem estocásticas.

Veja que, se β_t é uma tendência determinística, então $\sigma_\xi^2 = 0$. Caso $\sigma_\xi^2 > 0$, β_t será uma tendência estocástica. De acordo com Ernesto et al. (2016), é possível testar se uma série possui

tendência Determinística ou Estocástica. Ela utiliza o Método de Monte Carlo para simular a estatística de teste sob a hipótese nula, onde: $H_0: \sigma_{\xi}^2 = 0$ contra $H_1: \sigma_{\xi}^2 > 0$.

Ernesto et al. (2016) comparou a nova proposta de Monte Carlo com o teste NM original e constatou que, para a variância do termo de Nível (σ_{η}^2) e tamanhos de amostras maiores que 100, os testes NM_{mc} e NM possuem praticamente os mesmos poderes.

Segundo Silva et al. (2020) o uso do Método de Monte Carlo permitiu expandir o teste NM para a variância do termo de tendência e, após obter os resultados, percebeu-se que o poder do teste era satisfatório para séries temporais com distribuição pertencentes à família Localização-Escala, como as distribuições Gaussiana e Logística.

Dizemos que uma variável aleatória tem sua distribuição pertencente à família localização-escala quando, se subtraída de sua média e dividida pelo seu desvio-padrão, a distribuição resultante não mais dependa dos parâmetros de média e variância originais. São exemplos mais comuns desta família de distribuições a Normal, a Poisson, a Gumbel, a Exponencial e a Gama.

No entanto, gostaríamos de estudar o poder do teste proposto por Ernesto et al. (2016) para saber se ele é satisfatório nos casos em que as séries temporais não possuem distribuição pertencente à família localização-escala, tais como: t de Student, Qui-quadrado, Binomial, Hipergeométrica e Uniforme. Casos assim acontecem quando o usuário do teste pressupõe uma distribuição para os dados, como por exemplo Normal, quando estes possuem outra distribuição.

2. OBJETIVO

O objetivo deste trabalho é verificar se o teste proposto por Silva et al. (2020) é válido para séries que não pertençam à família locação-escala. Neste trabalho simularemos séries com distribuição de Qui-quadrado e Binomial. A ideia é compararmos o poder do teste quando aplicado à series temporais com distribuição na família locação-escala (Normal, por exemplo) com o poder do teste quando aplicado às séries temporais que não pertençam a essa família de distribuições. Adicionaremos séries com distribuições fora da família locação-escala, enquanto usaremos o teste pressupondo normalidade dos dados, ou seja, injetaremos um “erro” propositalmente para verificar a robustez do teste. Faremos isso via simulação Monte Carlo.

3. MATERIAL E MÉTODOS

Na proposta de Ernesto et al. (2016), foi demonstrado que estatística NM, quando pertencente à família locação-escala, pode ser representada por uma nova estatística, U, que por sua vez pode ser simulada sob a hipótese nula, ou seja:

$$NM = \frac{\sum_{t=1}^T [\sum_{s=t}^T (Y_s - \bar{Y})]^2}{(T-1) \sum_{t=1}^T (Y_t - \bar{Y})^2} \quad (3.1)$$

corresponde a

$$U = \frac{\sum_{t=1}^T [\sum_{s=t}^T (Z_s - \bar{Z})]^2}{(T-1) \sum_{t=1}^T (Z_t - \bar{Z})^2} \quad (3.2)$$

onde Z é uma variável aleatória que pode ser simulada diretamente da forma padronizada da distribuição estabelecida pelo usuário, ou seja, sem que este precise conhecer os valores dos parâmetros do modelo estrutural.

DEMONSTRAÇÃO PARA O CASO GAUSSIANO:

$$\begin{aligned} NM &= \frac{\sum_{t=1}^T [\sum_{s=t}^T (Y_s - \bar{Y})]^2}{(T-1) \sum_{t=1}^T (Y_t - \bar{Y})^2} \\ &= \frac{\sigma_\epsilon^{-2} \sum_{t=1}^T [\sum_{s=t}^T (Y_s - \mu_0 + \mu_0 - \bar{Y})]^2}{(T-1) \sigma_\epsilon^{-2} \sum_{t=1}^T (Y_t - \mu_0 + \mu_0 - \bar{Y})^2} \\ &= \frac{\sum_{t=1}^T [\sum_{s=t}^T (\frac{Y_s - \mu_0}{\sigma_\epsilon} - \frac{\bar{Y} - \mu_0}{\sigma_\epsilon})]^2}{(T-1) \sum_{t=1}^T (\frac{Y_t - \mu_0}{\sigma_\epsilon} - \frac{\bar{Y} - \mu_0}{\sigma_\epsilon})^2} \\ \text{Sob } H_0 &= \frac{\sum_{t=1}^T [\sum_{s=t}^T (Z_s - \bar{Z})]^2}{(T-1) \sum_{t=1}^T (Z_t - \bar{Z})^2} \end{aligned}$$

3.1. Poder do Teste Monte Carlo para o Ruído do Termo de Tendência

Define-se como poder estatístico de um teste a probabilidade de se rejeitar a hipótese nula. Pode-se escrever da seguinte forma: $P(\text{Rejeitar } H_0 | \sigma_\epsilon^2 \neq 0)$. Espera-se que, quanto mais longe um parâmetro esteja do valor estipulado sob a hipótese nula, maior será o poder do teste em detectar que a hipótese nula deve ser rejeitada. Por outro lado, deve-se controlar a probabilidade do Erro Tipo I sob o nível α desejado, ou seja, $P(\text{Rejeitar } H_0 | \sigma_\epsilon^2 = 0) \leq \alpha$.

Para testar a variância do termo de tendência, considere o modelo abaixo:

$$Y_t = \mu_t + \epsilon_t, \epsilon_t \sim \text{Normal}(0, \sigma_\epsilon^2) \quad (3.3)$$

$$\mu_t = \mu_{t-1} + \beta_t + \eta_t, \eta_t \sim \text{Normal}(0, \sigma_\eta^2) \quad (3.4)$$

$$\beta_t = \beta_{t-1} + \xi_t, \xi_t \sim \text{Normal}(0, \sigma_\xi^2) \quad (3.5)$$

Queremos testar:

$$\begin{cases} H_0: \sigma_\xi^2 = 0, (\text{Tendência Determinística}) \\ H_1: \sigma_\xi^2 > 0 (\text{Tendência Estocástica}). \end{cases}$$

Segundo SILVA et al (2020), devemos definir uma transformação $X_t = Y_t - Y_{t-1}$, com esperança μ_t e variância σ_t^2 . O método requer assumir que, para $\sigma_\xi^2 = 0$, a distribuição de probabilidade $F_Z(z)$ da transformação

$$Z_t = \frac{Y_t - g_1(\mu_t)}{g_2(\sigma_t^2)}, t = 2, \dots, n, \quad (3.6)$$

seja conhecida, ou seja, que Y_t pertença à família locação-escala.

A estatística de teste proposta é:

$$U(\vec{W}) = \frac{\sum_{l=1}^N [\sum_{s=l}^N (W_s - \bar{W})]^2}{(N-1) \sum_{l=1}^N (W_l - \bar{W})^2} \quad (3.7)$$

onde $\vec{W} = (W_1, \dots, W_N)$, e,

$$W_1 = \begin{cases} Y_{2l} - Y_{2l-1}, & \text{se } n \text{ for par,} \\ Y_{2l+1} - Y_{2l}, & \text{caso contrário.} \end{cases} \quad (3.8)$$

O poder do teste será verificado nos casos em que o usuário suponha normalidade dos dados em uma série que, na verdade, não possua distribuição Normal e nem pertença à família locação-escala. Logo, não mudaremos nada no método proposto por Ernesto et al. (2016), mas sim nos dados de entrada, ou seja, na amostra que iremos simular. Vamos introduzir uma simulação do teste sob dois cenários distintos, um em que os dados seguem uma distribuição Qui-quadrado, e outro em que seguem uma Binomial, porém “ignorando” este fato na hora de aplicar o teste.

É sabido que a esperança e a variância de uma distribuição Qui-quadrado são, respectivamente, v e $2v$. E para a distribuição Binomial são $E(Y_i) = np$ e $\text{Var}(Y_i) = np(1-p)$. Aplicaremos esses fatos para gerar os diferentes cenários de variâncias dos termos de tendência sob a hipótese alternativa usados na simulação Monte Carlo.

Utilizaremos vários cenários para fazer a comparação dos poderes em cada caso. Os cenários incluem tamanhos de amostra $n = 30$, $n = 50$ e $n = 100$ e variâncias (σ_ξ^2) tomadas da sequência $0, 0,2, \dots, 2$. Adotaremos níveis de significância iguais a $0,01$ e $0,05$ para cada cenário.

3.2. Simulação de Monte Carlo

Seja u_0 uma observação de $U(\vec{W})$. Assim, seja $\vec{Z}_1, \dots, \vec{Z}_{m-1}$ uma sequência de $m-1$ vetores N -dimensionais de Monte Carlo, sendo $Z_{i,j}$ a i -ésima observação da j -ésima amostra, com $i = 1, \dots, N$, $j = 1, \dots, (m-1)$, com $Z_{i,j} \sim F_Z(z)$. Seja, ainda, U_1, \dots, U_{m-1} a sequência de estatísticas de teste obtidas por se aplicar a estatística $U(\cdot)$ para cada uma das amostras \vec{Z}_j , isto é, $U_j = U(\vec{Z}_j)$.

Com isso, define-se a medida de evidência de Monte Carlo:

$$G = \sum_{j=1}^{m-1} I(U_j \geq u_0). \quad (3.9)$$

Para um nível de significância qualquer de interesse, $\alpha \in (0, 1)$, seja $\psi(G, \alpha, m)$ a função de decisão binária baseada na estatística G , onde:

$$\psi (G, \alpha, m) = \begin{cases} 1, & \text{rejeitamos } H_0 \text{ (Tendência Estocástica),} \\ 0, & \text{não rejeitamos } H_0 \text{ (Tendência Determinística).} \end{cases} \quad (3.10)$$

onde $\psi (G, \alpha, m) = 1$ se $G \leq (\alpha m - 1)$ e $\psi (G, \alpha, m) = 0$, caso contrário. A aplicação desse critério de decisão equivale ao uso do p-valor de Monte Carlo dado por:

$$P_{mc} = \frac{G+1}{m}, \quad (3.11)$$

onde rejeita-se H_0 para valores de $P_{mc} \leq \alpha$. O chamamos aqui de MC_t , dado que estamos testando o termo de tendência com o teste proposto.

4. RESULTADOS E DISCUSSÃO

Após fazermos as devidas adaptações para cada cenário, disponibilizamos os resultados das simulações em tabelas com o intuito de fazer as comparações para a verificação da sua robustez do MC_t .

4.1. Comparação dos poderes: Normal versus Qui-quadrado

As Tabelas 1 e 2 apresentam os poderes estimados do MC_t para séries Normais e Qui-quadrado (χ_v^2) para $\alpha = 0,01$ e $\alpha = 0,05$, respectivamente.

Sabe-se que, quanto maior o poder melhor é o método proposto. Portanto, o teste parece robusto, visto que apresentou valores altos de poder quando a suposição de normalidade foi violada na tentativa de detectar se a variância do termo de tendência era maior do que zero.

Tabela 1: Comparação dos Poderes do teste para cada Cenário, Normal versus Qui-quadrado, com $\alpha = 0,01$

σ_ξ^2	$\epsilon_t \sim \text{Normal} (1, \sigma_\epsilon^2)$			$\epsilon_t \sim \chi_v^2 (v, 2v)$			
	$n =$	30	50	100	30	50	100
0,0		0,01	0,01	0,01	0,01	0,01	0,09
0,2		0,14	0,34	0,64	0,70	1,00	1,00
0,4		0,23	0,46	0,74	0,98	1,00	1,00

0,6	0,29	0,52	0,79	1,00	1,00	1,00
0,8	0,33	0,56	0,81	1,00	1,00	1,00
1,0	0,36	0,58	0,82	1,00	1,00	1,00
1,2	0,39	0,61	0,84	1,00	1,00	1,00
1,4	0,4	0,61	0,84	1,00	1,00	1,00
1,6	0,42	0,62	0,86	1,00	1,00	1,00
1,8	0,43	0,63	0,86	1,00	1,00	1,00
2,0	0,44	0,65	0,86	1,00	1,00	1,00

Tabela 2: Comparação dos Poderes do teste para cada Cenário, Normal versus Qui-quadrado, com $\alpha = 0,05$.

σ_{ξ}^2	$\epsilon_t \sim \text{Normal}(1, \sigma_{\epsilon}^2)$			$\epsilon_t \sim \chi_v^2(v, 2v)$			
	n =	30	50	100	30	50	100
0,0		0,05	0,05	0,05	0,05	0,05	0,05
0,2		0,28	0,49	0,77	0,85	1,00	1,00
0,4		0,39	0,62	0,85	0,99	1,00	1,00
0,6		0,45	0,66	0,88	1,00	1,00	1,00
0,8		0,45	0,7	0,9	1,00	1,00	1,00
1,0		0,53	0,72	0,91	1,00	1,00	1,00
1,2		0,54	0,73	0,92	1,00	1,00	1,00
1,4		0,56	0,75	0,93	1,00	1,00	1,00
1,6		0,58	0,76	0,93	1,00	1,00	1,00
1,8		0,58	0,76	0,94	1,00	1,00	1,00
2,0		0,6	0,77	0,94	1,00	1,00	1,00

Para verificar se, de fato, houve controle do Erro Tipo I, ou seja, se o nível de significância permanecia menor ou igual a $\alpha_1 = 0,01$ e $\alpha_2 = 0,05$, fizemos a mesma simulação acima, porém sob a hipótese nula ($\sigma_{\xi}^2 = 0$), fixando diferentes valores da variância da série Y_t . Nota-se pelas Tabelas 3 e 4 que, a não ser pela variação intrínseca à simulação Monte Carlo, o MC_t preserva controle sobre a probabilidade do Erro Tipo I. Portanto, podemos afirmar que o teste é robusto para séries com distribuição Qui-quadrado.

Tabela 3: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Qui-quadrado, para alpha igual a 0,01.

σ_{ϵ}^2	Probabilidade do Erro Tipo I			
	n =	30	50	100
0,2		0,0120	0,0060	0,0075

0,5	0,0090	0,0105	0,0085
1,0	0,0075	0,0100	0,0105
1,5	0,0075	0,0105	0,0105
2,0	0,0105	0,0065	0,0125

Tabela 4: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Qui-quadrado, para α igual a 0,05.

σ_{ϵ}^2	Probabilidade do Erro Tipo I			
	n =	30	50	100
0,2		0,0475	0,0455	0,0605
0,5		0,0445	0,0520	0,0505
1,0		0,0525	0,0515	0,0545
1,5		0,0485	0,0470	0,0460
2,0		0,0495	0,0425	0,0485

4.2. Comparação dos poderes: Normal versus Binomial

Abaixo, podemos ver os resultados do Poder do Teste, onde comparamos uma simulação do Método de Monte Carlo para séries Normais e Binomial. Na Tabela 5 adotou-se um nível de significância de $\alpha = 0,01$ e na Tabela 6 adotou-se $\alpha = 0,05$.

Assim como no tópico anterior, queremos testar a robustez do teste para suposições errôneas de normalidade. E, nesse caso, o teste parece robusto também, visto que apresentou valores altos de poder quando a suposição de normalidade da série foi violada na tentativa de detectar se a variância do termo de tendência era maior do que zero.

Tabela 5: Comparação dos Poderes do teste para cada Cenário, Normal versus Binomial, com $\alpha = 0,01$.

σ_{ξ}^2	$\epsilon_t \sim \text{Normal}(1, \sigma_{\epsilon}^2)$			$\epsilon_t \sim \text{Bin}(n, p)$			
	n =	30	50	100	30	50	100
0,0		0,01	0,01	0,01	0,01	0,01	0,01
0,2		0,14	0,34	0,64	0,63	0,97	1,00
0,4		0,23	0,46	0,74	0,95	1,00	1,00
0,6		0,29	0,52	0,79	0,99	1,00	1,00
0,8		0,33	0,56	0,81	0,99	1,00	1,00
1,0		0,36	0,58	0,82	1,00	1,00	1,00
1,2		0,39	0,61	0,84	1,00	1,00	1,00
1,4		0,4	0,61	0,84	1,00	1,00	1,00
1,6		0,42	0,62	0,86	1,00	1,00	1,00

1,8	0,43	0,63	0,86	1,00	1,00	1,00
2,0	0,44	0,65	0,86	1,00	1,00	1,00

Tabela 6: Comparação dos Poderes do teste para cada Cenário, Normal versus Binomial, com $\alpha = 0,05$

σ_{ξ}^2	$\epsilon_t \sim \text{Normal}(1, \sigma_{\epsilon}^2)$			$\epsilon_t \sim \text{Bin}(n, p)$			
	n =	30	50	100	30	50	100
0,0		0,05	0,05	0,05	0,05	0,05	0,05
0,2		0,28	0,49	0,77	0,78	0,99	1,00
0,4		0,39	0,62	0,85	0,99	1,00	1,00
0,6		0,45	0,66	0,88	1,00	1,00	1,00
0,8		0,45	0,7	0,9	1,00	1,00	1,00
1,0		0,53	0,72	0,91	1,00	1,00	1,00
1,2		0,54	0,73	0,92	1,00	1,00	1,00
1,4		0,56	0,75	0,93	1,00	1,00	1,00
1,6		0,58	0,76	0,93	1,00	1,00	1,00
1,8		0,58	0,76	0,94	1,00	1,00	1,00
2,0		0,6	0,77	0,94	1,00	1,00	1,00

E assim como no tópico anterior, fizemos os testes para verificar o controle do Erro Tipo I quando os dados da série não são Normais, mas sim Binomiais. Nota-se pelas Tabelas 7 e 8 que, a não ser por variações intrínsecas à simulação Monte Carlo, a probabilidade do Erro Tipo I parece estar sob controle. Portanto, temos evidências de que o teste é robusto para séries com distribuição Binomial.

Tabela 7: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Binomial, para alpha igual a 0,01.

σ_{ϵ}^2	Probabilidade do Erro Tipo I			
	n =	30	50	100
0,2		0,0105	0,0110	0,0120
0,5		0,0070	0,0060	0,0100
1,0		0,0090	0,0085	0,0110
1,5		0,0090	0,0125	0,0100
2,0		0,0055	0,0090	0,0140

Tabela 8: Probabilidade do Erro Tipo I sob H_0 , quando ϵ é Binomial, para alpha igual a 0,05.

σ_{ϵ}^2	Probabilidade do Erro Tipo I			
	n =	30	50	100
0,2		0,0479	0,0450	0,0490
0,5		0,0515	0,0450	0,0485

1,0	0,0485	0,0540	0,0445
1,5	0,0615	0,0475	0,0575
2,0	0,0440	0,0530	0,0515

4.3.Exemplo Prático

Considere a série seguinte: Energia em relação aos valores mensais do consumo de energia elétrica no estado do Espírito Santo, de janeiro de 1968 a setembro de 1979. Queremos saber se essa série possui uma tendência determinística ou estocástica. As hipóteses são $H_0: \sigma_{\xi}^2 = 0$ contra $H_1: \sigma_{\xi}^2 > 0$. Ao fazermos o teste proposto por Ernesto et al. (2016), iremos supor que a série segue uma distribuição Normal, no entanto ela não será Normal.

Primeiramente, importamos a série para o software R Studio. Em seguida, transformamos os dados para série temporal através da função $ts()$ e plotamos a série. Como pode-se observar pela Figura 1, a série parece possuir uma certa tendência positiva. Então, vamos verificar se ela é estacionária ou não estacionária através do teste de Dickey-Fuller, onde as hipóteses são:

$$\begin{cases} H_0: A \text{ série é não estacionária;} \\ H_1: A \text{ série é estacionária.} \end{cases}$$

Os resultados do teste apontaram que a série não é estacionária, ou seja, não houve rejeição da hipótese nula ao nível de significância de 5%. Portanto, foi comprovado que a série possui uma certa tendência.

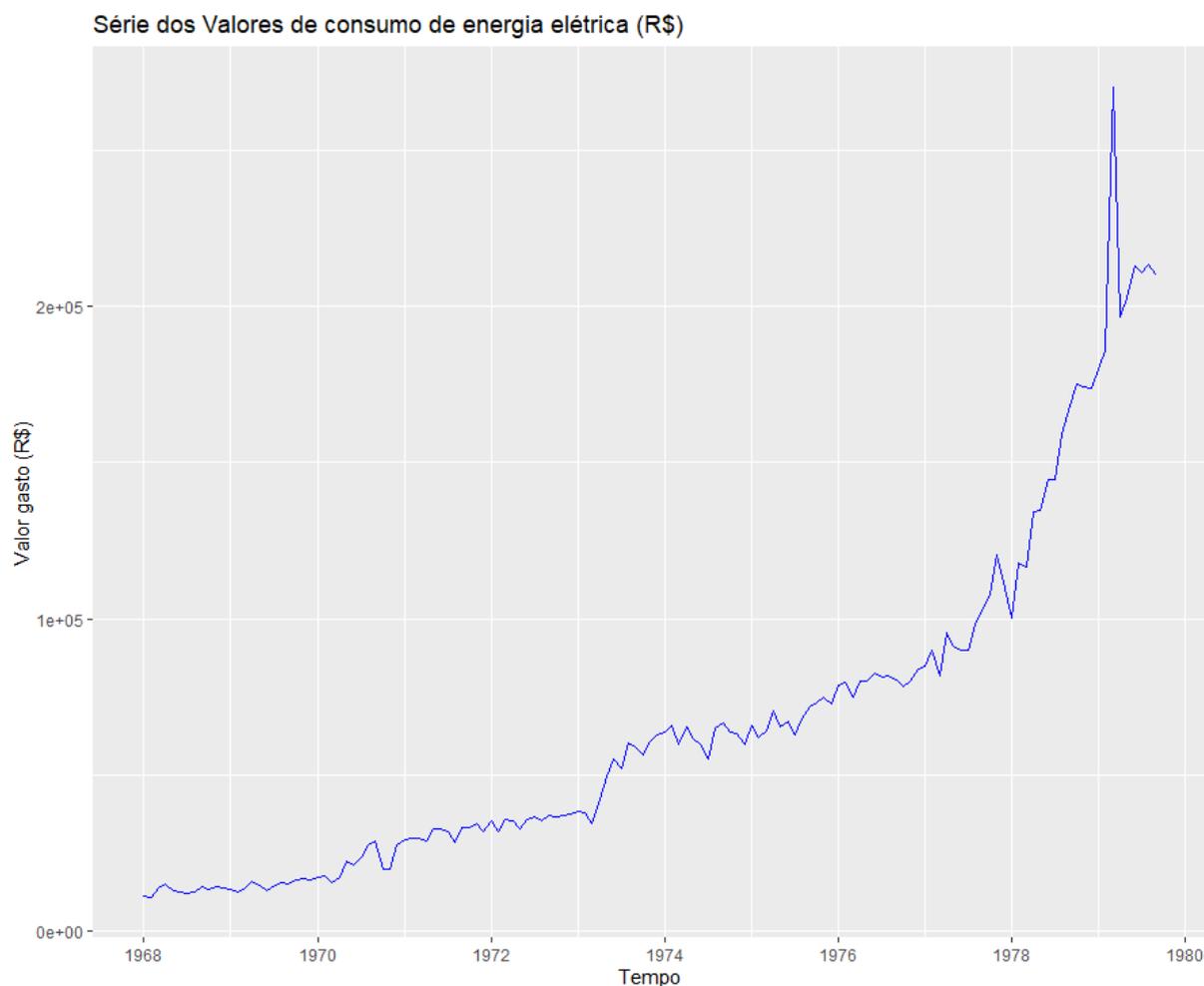


Figura 1: Consumo de energia elétrica, em valores monetários (R\$), no estado do Espírito Santo, de janeiro de 1968 a setembro de 1979

Observe o histograma da série na Figura 2. Como podemos notar, a distribuição da série não parece ser Normal. Então, fizemos o Teste de Shapiro-Wilk para obter evidência sobre esta suspeita. As hipóteses eram:

$$\begin{cases} H_0: \text{Os dados são Normais;} \\ H_1: \text{Os dados não são Normais.} \end{cases}$$

Como o p-valor foi de $5,476e-11$, rejeitamos a hipótese nula ao nível de significância de 5%, ou seja, existem fortes evidências de que os dados não são Normais.

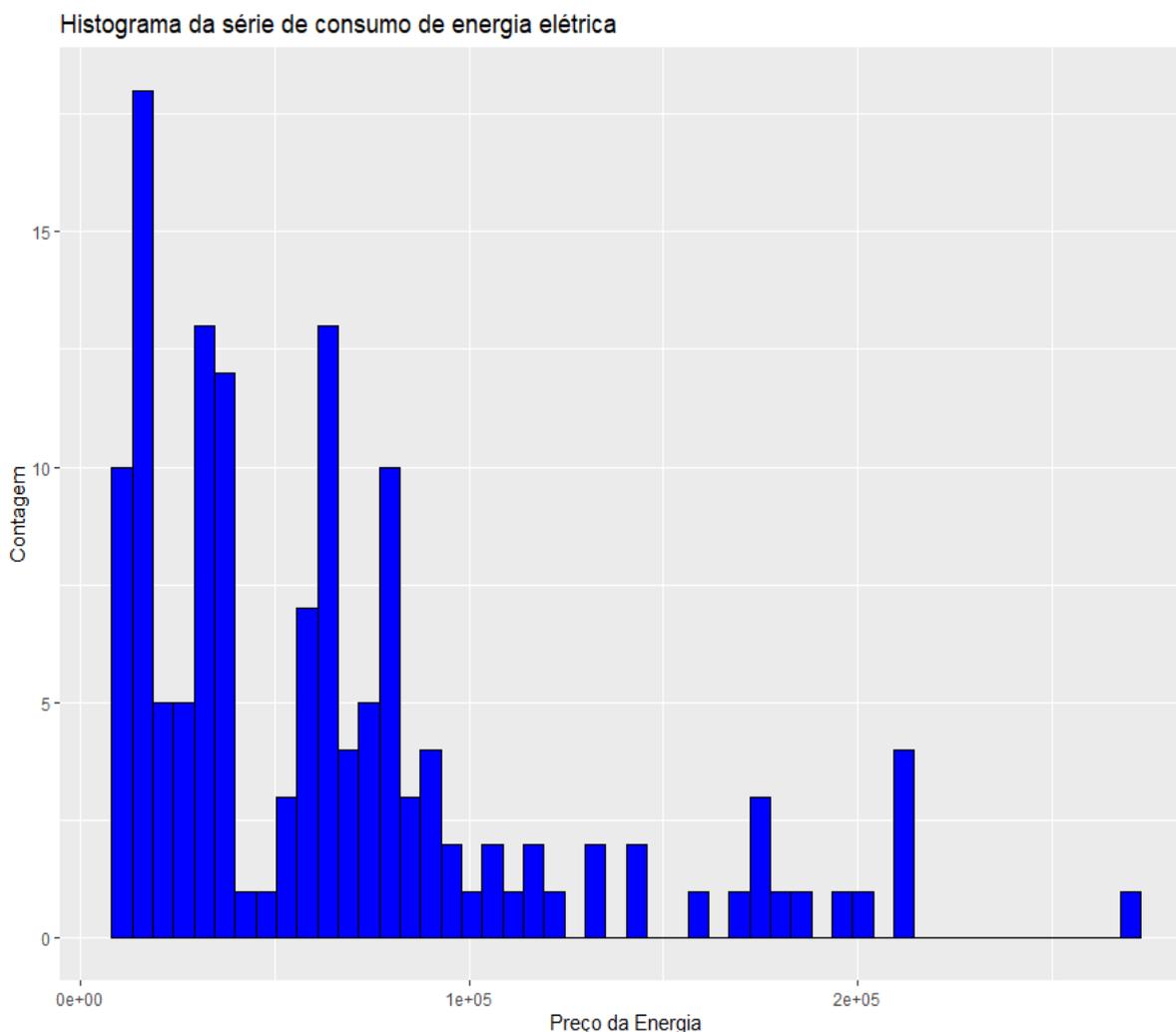


Figura 2: Histograma da série de consumo de energia elétrica, em valores monetários (R\$), no estado do Espírito Santo, de janeiro de 1968 a setembro de 1979

Por fim, após concluirmos que os dados não eram Normais e ainda possuíam uma certa tendência, utilizamos a Estatística MC_t para saber se esta tendência era determinística ou estocástica. As hipóteses eram:

$$\begin{cases} H_0: \text{Tendência Determinística;} \\ H_1: \text{Tendência Estocástica.} \end{cases}$$

Feito o teste com $m = 5000$, chegamos ao valor-p de 0,1882. Portanto, não rejeitamos a hipótese nula sob um nível de significância de 5%, ou seja, a série possui uma tendência determinística.

5. CONCLUSÕES

Quando os dados não são normais, mas sim binomiais ou Qui-quadrados, o Teste MC_t consegue detectar quando a variância do termo de tendência é maior que zero, ou seja, a desvios da suposição de normalidade, ele tende a possuir um poder alto. E ainda possui controle da probabilidade do Erro Tipo I para quaisquer tamanhos de amostra.

Portanto, o teste MC_t parece robusto a erros de especificação da verdadeira distribuição dos dados. É possível realizar testes sobre as características da tendência de séries temporais, com um poder satisfatório, mesmo quando elas não pertençam à família locação-escala.

6. REFERÊNCIAS BIBLIOGRÁFICAS

SILVA, Ivair Ramos; ERNESTO, Dulcília Carlos Guezimane; OLIVEIRA, Fernando Luiz Pereira de; MARQUES, R.; OLIVEIRA, A. Monte Carlo Test for Stochastic Trend in Linear Structural Models for the Location-Scale Family. *Brazilian Review of Econometrics* (In press).

ERNESTO, Dulcília Carlos Guezimane; OLIVEIRA, Fernando Luiz Pereira de; SILVA, Ivair Ramos. Teste Monte Carlo para a Tendência Estocástica em Modelos de Espaços de Estados. Universidade Federal de Viçosa. Viçosa. 2016. (Mestrado)

SILVA, Ivair Ramos; ASSUNÇÃO, Renato; AZEVEDO, Marcelo. Testes Monte Carlo Convencionais e Sequenciais Comparação dos Poderes e dos Tempos de Execução. Universidade Federal de Minas Gerais. Belo Horizonte. 2007.

REIS, Marcelo Menezes. Teste de hipóteses: Poder do Teste. Disponível em: <<https://www.inf.ufsc.br/~marcelo.menezes.reis/Aula09CPGCC.pdf>> Universidade Federal de Santa Catarina. Santa Catarina. Acesso em: 19 jun. 2020.

SITES ELETRÔNICOS

Polo Soluções. O que é o Poder do Teste Estatístico. Disponível em: <<http://polosolucoes.com.br/o-que-e-poder-do-teste-estatistico/#:~:text=Para%20muitos%20belts%2C%20o%20poder,um%20conceito%20frequentemente%20n%C3%A3o%20utilizado.&text=Portanto%2C%20definimos%20o%20poder%20de,%C3%A9%20igual%20a%201%20%E2%80%93%20CE%B2>> Acesso em: 19 jun. 2020.

LABTROP. Disponível em: <<http://labtrop.ib.usp.br/lib/exe/fetch.php?media=cursos:planeco:planeco:material:PowerAnalysisSampleSize.pdf>> Acesso em: 19 jun. 2020.

MORETTIN, Pedro A.; TOLOI, Clelia M. C. Análise de Séries Temporais. ABE- Projeto Fisher e Editora Edgard Blucher, 2004. Disponível em: <<https://www.ime.usp.br/~pam/ST.html>> Acesso em: 17 jul. 2020.