

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**ESTRATÉGIA PARA A DEFINIÇÃO DO NÚMERO DE
DEFASAGENS NA APLICAÇÃO DO TESTE ADF**

por

Anderson Garcia Silveira

Tese para obtenção do Título de
Doutor em Modelagem Computacional

Rio Grande, maio de 2023

Anderson Garcia Silveira

“ESTRATÉGIA PARA A DEFINIÇÃO DO NÚMERO DE DEFASAGENS NA APLICAÇÃO DO TESTE ADF”

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Doutor. Área de concentração: Modelagem Computacional.

Aprovado em: 23/05/2023

BANCA EXAMINADORA

Documento assinado digitalmente
 **VIVIANE LEITE DIAS DE MATTOS**
Data: 02/06/2023 13:46:54-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Viviane Leite Dias de Mattos (Orientadora - FURG)

Documento assinado digitalmente
 **Luiz Ricardo Nakamura**
Data: 05/06/2023 11:01:25-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Luiz Ricardo Nakamura (Coorientador - UFLA)

Documento assinado digitalmente
 **DIANA FRANCISCA ADAMATTI**
Data: 05/06/2023 11:15:37-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Diana Francisca Adamatti (FURG)

Documento assinado digitalmente
 **TIAGO DA CRUZ ASMUS**
Data: 02/06/2023 14:04:26-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Tiago da Cruz Asmus (FURG)

ANA JULIA
RIGHETTO:34468810858

Assinado de forma digital por ANA JULIA
RIGHETTO:34468810858
Dados: 2023.06.05 10:11:23 -03'00'

Profa. Dra. Ana Julia Righetto (ALVAZ Agritech)

Rio Grande - RS
2023

Ficha Catalográfica

S587e Silveira, Anderson Garcia.

Estratégia para definição do número de defasagens na aplicação do teste ADF / Anderson Garcia Silveira. – 2023.

119 f.

Tese (doutorado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2023.

Orientadora: Dra. Viviane Leite Dias de Mattos.

Coorientador: Dr. Luiz Ricardo Nakamura.

1. Séries temporais 2. Previsão 3. ADF 4. Seleção de defasagens
5. Raiz unitária I. Mattos, Viviane Leite Dias de II. Nakamura, Luiz Ricardo III. Título.

CDU 004.4

Catálogo na Fonte: Bibliotecária Vanessa Ceiglinski Nunes CRB 10/2174

À minha esposa e minha filha.

AGRADECIMENTOS

Agradeço aos meus pais Rosa Maria Silva Garcia e Henrique da Silva Silveira por me mostrarem o caminho a seguir, mesmo que por eles nunca traçado.

À minha esposa Mariane Coelho Amaral, pelo suporte, companheirismo e incentivo durante toda a jornada.

À minha filha Martina Amaral Silveira, dona dos olhos onde encontrei forças para continuar nos momentos mais difíceis.

À minha orientadora Viviane Leite Dias de Mattos, pela orientação, paciência e incansável busca por melhores soluções para os problemas encontrados.

Ao meu coorientador Luiz Ricardo Nakamura, pela grande contribuição e por indicar caminhos diante das dificuldades.

Aos profissionais que compuseram a banca de avaliação, que apresentaram valorosas contribuições para a melhoria do trabalho.

Aos meus colegas de trabalho do IFSUL e do IFRS, que compreenderam minha dupla jornada e muitas vezes contribuíram para que eu pudesse me dedicar a este trabalho.

À Universidade Federal do Rio Grande (FURG) e aos professores do Programa de Pós-Graduação em Modelagem Computacional, por me proporcionarem realizar este estudo.

A todos que, de alguma forma, contribuíram para o meu crescimento pessoal e acadêmico.

RESUMO

A energia elétrica é um fator de grande importância para o desenvolvimento de uma nação, logo, adequar sua geração à demanda é vital para evitar crises de falta de energia ou mesmo capacidade de geração ociosa. Diante disso, é importante modelar o seu consumo a fim de planejar a capacidade de geração, adequando-a à demanda. Dentre os métodos mais utilizados de modelar séries temporais de consumo de energia elétrica, está a abordagem econométrica. Nesse contexto, a ausência de estacionariedade faz com que a série não seja adequada para a realização de previsões. Os testes que são capazes de identificar a presença ou ausência desta propriedade em uma série temporal são os testes de raiz unitária, entre os quais, um dos mais utilizados na literatura é o teste de Dickey-Fuller Aumentado (ADF). Para a aplicação do teste ADF, é necessário selecionar o número de defasagens a ser considerado no modelo, tarefa que não apresenta procedimento consensual na literatura. Diante disso, este trabalho realiza experimentos para identificar a melhor maneira de selecionar o número máximo de defasagens a ser considerado no teste ADF, bem como determinar ideal de defasagens, considerando quatro fatores: influência da estrutura do modelo da série temporal, critério de informação utilizado para identificar o modelo mais parcimonioso, critério de determinação da defasagem máxima e nível de significância. Assim, por simulação, são geradas 648 séries temporais com diferentes características, que são replicadas 10000 vezes e tomadas as proporções de acerto do teste. Para avaliar a influência de cada fator nestas proporções é aplicado um modelo aditivo generalizado para localização, escala e forma (GAMLSS). Ao fim, melhores resultados foram encontrados quando o critério de seleção de defasagens máxima foi determinado por l_4 e o número de defasagens foi selecionado por BIC.

Palavras-chaves: Séries Temporais; Previsão; ADF; Seleção de defasagens; Raiz Unitária.

ABSTRACT

Electricity is a crucial factor in the development of a nation, and adjusting its generation to meet demand is vital to avoid energy shortages or idle generation capacity. In this regard, it is important to model electricity consumption to plan generation capacity, aligning it with demand. One of the most commonly used methods for modeling time series of electricity consumption is the econometric approach. However, the absence of stationarity makes the series unsuitable for making predictions. Unit root tests, such as the Augmented Dickey-Fuller (ADF) test, can identify the presence or absence of this property in a time series. To apply the ADF test, it is necessary to select the number of lags to be considered in the model, which is not a consensual procedure in the literature. This study conducts experiments to identify the best way to select the maximum number of lags to be considered in the ADF test and determine the optimal lags, considering four factors: the influence of the model structure of the time series, the information criterion used to identify the most parsimonious model, the criterion for determining the maximum lag, and the significance level. Thus, 648 time series with different characteristics are generated by simulation, replicated 10,000 times, and the test's accuracy proportions are calculated. To evaluate the influence of each factor on these proportions, a Generalized Additive Model for Location, Scale, and Shape (GAMLSS) is applied. Ultimately, better results are found when the maximum lag selection criterion is determined by l_4 , and the number of lags is selected by BIC.

Key-word: Time Series; Forecasting; ADF; Lag selection; Unit root.

SUMÁRIO

1	INTRODUÇÃO	17
1.1	O tema e sua importância	17
1.2	A modelagem	19
1.3	Hipótese da Pesquisa	20
1.4	Objetivos	20
1.4.1	Objetivo geral	20
1.4.2	Objetivos específicos	20
1.5	Justificativa	21
1.6	Estrutura do trabalho	21
2	REVISÃO DE LITERATURA	23
2.1	Estacionariedade	23
2.2	Testes DF e ADF	25
2.3	Modelagem GAMLSS	28
2.4	Revisão Sistemática	32
2.4.1	Análise Bibliométrica	32
2.4.2	Metassíntese	36
3	METODOLOGIA	46
3.1	Simulações preliminares	46
3.2	Simulações	47
3.2.1	Planejamento dos experimentos	47
3.2.2	Análise dos experimentos	48
3.3	Simulações Complementares	49
4	RESULTADOS E DISCUSSÕES	50
4.1	Análise dos resultados do estudo preliminar	50
4.2	Análise dos resultados das simulações	54
4.2.1	Séries Não Estacionárias($\phi = 1$)	54
4.2.2	Séries Estacionárias($\phi < 1$)	74
4.3	Análise dos resultados das simulações complementares	94
4.4	A estratégia	106
5	ESTUDO DE CASO	107
6	CONSIDERAÇÕES FINAIS	112
	REFERÊNCIAS	114

LISTA DE FIGURAS

Figura 2.1: Fluxograma para a aplicação do teste ADF	28
Figura 2.2: Histórico de produção anual	33
Figura 2.3: Publicações por Periódico	33
Figura 2.4: Autores mais relevantes	34
Figura 2.5: Produção por autor ao longo do tempo	34
Figura 2.6: Produção por país	34
Figura 2.7: Produção por Universidade	35
Figura 2.8: Citações por Documentos	35
Figura 2.9: Referências mais citadas	36
Figura 2.10: Palavras mais frequentes	36
Figura 4.1: Gráficos de caixa da proporção de casos de rejeição do hipótese nula, de acordo com a quantidade de defasagens e componentes do modelo no primeiro cenário com simulações considerando $\phi = 1$	51
Figura 4.2: Gráficos de caixa da proporção de casos de rejeição do hipótese nula, de acordo com a quantidade de defasagens e componentes do modelo no segundo cenário com simulações considerando $\phi < 1$	53
Figura 4.3: Representação da proporção de casos em que a hipótese nula é rejeitada e as interações entre as defasagens e: (a), (b), (c) número de componentes (com constante; ambos – com constante e tendência; nenhum – sem constante, nem tendência); (d), (e), (f) intensidade de ϕ	55
Figura 4.4: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 25$	56
Figura 4.5: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 25$	58
Figura 4.6: Interações entre critério de informação e tipo de modelo considerando a presença de raiz unitária e $T = 25$	59
Figura 4.7: Interações entre critérios de seleção e de determinação da defasagem máxima considerando a presença de raiz unitária e $T = 25$	59
Figura 4.8: Interações entre tipo de modelo e nível de significância considerando a presença de raiz unitária e $T = 25$	60
Figura 4.9: Interações entre o tipo de modelo e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 25$	60
Figura 4.10: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 25$	61
Figura 4.11: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 50$	63
Figura 4.12: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 50$	65

Figura 4.13: Interações entre critério de informação e tipo de modelo, considerando a presença de raiz unitária e $T=50$	65
Figura 4.14: Interações entre critério de informação e de determinação da defasagem máxima, considerando a presença de raiz unitária e $T=50$	66
Figura 4.15: Interações entre tipo de modelo e nível de significância, considerando a presença de raiz unitária e $T=50$	66
Figura 4.16: Interações entre o tipo de modelo e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=50$	67
Figura 4.17: Interações entre o nível de significância e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=50$	67
Figura 4.18: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 100$	69
Figura 4.19: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 100$	71
Figura 4.20: Interações entre critérios de seleção e de determinação da defasagem máxima, considerando a presença de raiz unitária e $T=100$	71
Figura 4.21: Interações entre tipo de modelo e nível de significância, considerando a presença de raiz unitária e $T=100$	72
Figura 4.22: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 25$	75
Figura 4.23: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a série estacionária e $T = 25$	77
Figura 4.24: Interações entre nível de significância e tipo de modelo considerando a série estacionária e $T = 25$	78
Figura 4.25: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 25$	78
Figura 4.26: Interações entre tipo de modelo e nível de significância considerando a presença de raiz unitária e $T = 25$	79
Figura 4.27: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 25$	79
Figura 4.28: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 25$	80
Figura 4.29: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 50$	82
Figura 4.30: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a série estacionária e $T = 50$	84
Figura 4.31: Interações entre nível de significância e tipo de modelo considerando a série estacionária e $T = 50$	84
Figura 4.32: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 50$	85

Figura 4.33: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 50$	86
Figura 4.34: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 50$	86
Figura 4.35: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 100$	88
Figura 4.36: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a série estacionária e $T = 100$	90
Figura 4.37: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 100$	90
Figura 4.38: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 100$	91
Figura 4.39: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 100$	91
Figura 4.40: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 40$	95
Figura 4.41: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 40$	97
Figura 4.42: Interações entre critério de informação e tipo de modelo, considerando a presença de raiz unitária e $T=40$	97
Figura 4.43: Interações entre critérios de seleção e de determinação da defasagem máxima, considerando a presença de raiz unitária e $T=40$	98
Figura 4.44: Interações entre tipo de modelo e nível de significância, considerando a presença de raiz unitária e $T=40$	98
Figura 4.45: Interações entre o tipo de modelo e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=40$	99
Figura 4.46: Interações entre o nível de significância e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=40$	99
Figura 4.47: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 40$	101
Figura 4.48: <i>Worm plot</i> gerado para o modelo GAMLSS considerando a série estacionária e $T = 40$	103
Figura 4.49: Interações entre nível de significância e tipo de modelo considerando a série estacionária e $T = 40$	103
Figura 4.50: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 40$	104
Figura 4.51: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 40$	104
Figura 4.52: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 40$	105

Figura 5.1: Box plot da série de consumo de eletricidade	107
Figura 5.2: Consumo de energia elétrica no Brasil, 1952-2022	108
Figura 5.3: Box plot da série de consumo de eletricidade	110
Figura 5.4: Consumo de energia elétrica no Brasil, 1952-2022 (em primeira diferença)	110

LISTA DE TABELAS

Tabela 2.1: Interpretações do <i>worm plot</i>	31
Tabela 2.2: Resultados das combinações nas bases de dados	32
Tabela 3.1: Fatores e respectivos níveis considerados nos experimentos realizados com a presença de raiz unitária	47
Tabela 3.2: Fatores e respectivos níveis considerados nos experimentos realizados com a ausência de raiz unitária	48
Tabela 4.1: Média (e desvio padrão) da proporção de casos em que a hipótese nula é rejeitada, no primeiro cenário com simulações considerando $\phi = 1$	50
Tabela 4.2: Resultados da análise de variância no primeiro cenário com simulações considerando $\phi = 1$	51
Tabela 4.3: Média (e desvio padrão) da proporção de casos em que a hipótese nula é rejeitada, no segundo cenário com simulações considerando $\phi < 1$	52
Tabela 4.4: Resultados da análise de variância no primeiro cenário com simulações considerando $\phi = 1$	54
Tabela 4.5: Média e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 25$	56
Tabela 4.6: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 25$	57
Tabela 4.7: Média e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 50$	62
Tabela 4.8: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 50$	64
Tabela 4.9: Média e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 100$	68
Tabela 4.10: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 100$	70
Tabela 4.11: Resumo dos resultados na presença de raiz unitária	73
Tabela 4.12: Média e o desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 25$	74
Tabela 4.13: Análise do efeito principal dos fatores e respectivas interações na taxa rejeição da hipótese nula em experimentos com séries estacionárias e $T = 25$	76
Tabela 4.14: Média e o desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 50$	81
Tabela 4.15: Análise do efeito principal dos fatores e respectivas interações na taxa rejeição da hipótese nula em experimentos com séries estacionárias e $T = 50$	83

Tabela 4.16 Média e o desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 100$	87
Tabela 4.17 Análise do efeito principal dos fatores e respectivas interações na taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 100$	89
Tabela 4.18 Resumo dos resultados para séries estacionárias	92
Tabela 4.19 Média, mediana e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 40$	94
Tabela 4.20 Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 40$	96
Tabela 4.21 Média, Mediana e desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 40$	100
Tabela 4.22 Análise do efeito principal dos fatores e respectivas interações na taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 40$	102
Tabela 5.1: Medidas resumo do consumo de energia elétrica no Brasil, 1952-2022 (em nível)	107
Tabela 5.2: Resultados do teste ADF para a série de consumo de eletricidade em nível	109
Tabela 5.3: Medidas resumo do consumo de energia elétrica no Brasil, 1952-2022 (em primeira diferença)	109
Tabela 5.4: Resultados do teste ADF para a série de consumo de eletricidade em primeira diferença	110

LISTA DE SÍMBOLOS

H_0	Hipótese nula
H_1	Hipótese alternativa
int	Valor inteiro
k	número da defasagem
l_4	Critério de seleção da defasagem máxima de Schwert
l_{12}	Critério de seleção da defasagem máxima de Schwert
l_h	Critério de seleção da defasagem máxima de Hyndman
L_n	Logaritmo natural
R^2	Coefficiente de determinação
u_t	Ruído branco
y	Série temporal
g_k	Função de ligação do GAMLSS
X	Matriz com termos aditivos do modelo
$\hat{r}$	Resíduos de Dunn e Smyth
$\bar{u}_t$	resíduos dos mínimos quadrados
T	Comprimento da série temporal

Símbolos gregos

β_1	Intercepto
$\bar{\beta}_1$	Estimador de β
β_2	Coefficiente da tendência linear
ϵ_t	Resíduos do Modelo AR(1)
σ	Parâmetro de escala
σ^2	Variância
$\hat{\sigma}_u^2$	Variância estimada dos resíduos
μ	Parâmetro de locação
τ	Estatística de teste dos testes DF e ADF
ϕ	Coefficiente autorregressivo
$\bar{\phi}$	Estimador de ϕ
v	Parâmetro associado à assimetria
λ	Parâmetro associado à curtose
η	Preditor do parâmetro do GAMLSS
Φ^{-1}	função de distribuição cumulativa inversa de uma variável normal padrão
χ^2	qui-quadrado

Sub índices

t	instante de tempo
max	máximo
min	mínimo

LISTA DE ABREVIATURAS

ADF	Dickey-Fuller Aumentado
AIC	Critério de Informação de Akaike
ANEEL	Agência Nacional de Energia Elétrica
AR	Autorregressivo
ARIMA	Autorregressivo integrado de médias móveis
ARMA	Autorregressivo de médias móveis
BIC	Critério de Informação Bayesiano
C	Constante
DF	Dickey-Fuller
EPE	Empresa de Pesquisa Energética
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
GAMLSS	Modelos Aditivos Generalizados para Localização, Escala e Forma
GLM	Modelos Lineares Generalizados
GLS	Mínimos Quadrados Generalizados
GSC	Geral para específico
HQC	Critério de Informação de Hannan-Quinn
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
MA	Médias Móveis
MAIC	Critério de Akaike Modificado
MBIC	Critério Bayesiano Modificado
MQO	Mínimos Quadrados Ordinários
N	Sem constante e sem tendência
NP	Nelsson e Plosser
OCDE	Organização para Cooperação e Desenvolvimento Econômico
ONS	Operador Nacional do Sistema
PP	Phillips-Perron
CT	Constante e Tendência

1 INTRODUÇÃO

1.1 O tema e sua importância

De acordo com ANEEL (2008), a energia é fator preponderante para o desenvolvimento econômico e social de um país, pois fornece apoio mecânico, térmico e elétrico às ações humanas. Conforme EPE (2019), o fato do consumo de energia estar relacionado à atividade econômica do país faz com que a análise econômica seja relevante para compreender os movimentos de variação de consumo.

Os países desenvolvidos que compõe a Organização para Cooperação e Desenvolvimento Econômico (OCDE) historicamente são os maiores consumidores de energia do mundo. Entretanto sua participação no consumo total mundial vêm diminuindo ao longo do tempo, ao contrário de países em desenvolvimento que apresentam grande taxa de crescimento de consumo de energia. Isso ocorre porque em países desenvolvidos há pouco espaço para grandes aumentos na produção industrial, ou no consumo de bens que pressionam a absorção de energia, uma vez que grande parte de suas populações já adquiriu estes bens na segunda metade do século XX. Além disso, em países ricos existe a tendência de utilização de equipamentos mais eficientes, tanto no setor industrial, quanto no residencial. Por outro lado, países em desenvolvimento ainda consomem muita energia através de indústrias energointensivas, principalmente (ANEEL, 2008).

Durante o século XX havia oferta abundante de energia, principalmente obtida através de combustíveis fósseis, e isso deu suporte para o crescimento econômico mundial. Já nos primeiros anos deste século, passou a ter maior importância o uso de energias de fontes sustentáveis. Nesse contexto, a produção de energia elétrica ingressa no século XXI em busca do desenvolvimento sustentável (ANEEL, 2008).

Dados de EPE (2020) mostram que entre os anos de 2001 e 2002 há uma alteração na série do consumo por eletricidade no Brasil. Essa modificação é justificada pelo racionamento ocorrido nesses anos. Além disso, segundo EPE (2019), em 2001 foi publicada a Lei de Eficiência Energética (nº 10.295/2001), o que tornou possível a definição de índices mínimos de eficiência energética para equipamentos.

De acordo com EPE (2019), em 2009, a oferta de energia no Brasil caiu 3,4% em função da crise financeira global. Entre os anos de 2010 e 2013, devido às políticas sociais dos anos anteriores, houve um aumento da renda média da população e diminuição da desigualdade. Este fato, associado à facilitação do crédito, fez com que itens como eletrodomésticos tivessem aumento de posse e uso pela população. Como resultado, foi observado um aumento médio anual de 5,1% do consumo energético, enquanto o PIB apresentou crescimento médio anual de 4,1%. Já no período de 2014 a 2018, houve um descolamento da demanda energética em relação ao PIB, pois mesmo em meio a retração econômica houve um aumento de intensidade energética devido ao aumento de participação de segmentos energointensivos no setor industrial (EPE, 2019).

Em EPE (2020) pode ser observado que a matriz energética brasileira é predominantemente de fonte hidráulica. Este fato faz com que grande parte da geração brasileira esteja sujeita ao clima, uma vez que esta fonte é diretamente dependente de fenômenos naturais. Em caso de grandes secas, por exemplo, os reservatórios ficam em níveis baixos, debilitando a produção de energia na região atingida.

O Operador Nacional do Sistema Elétrico (ONS) é responsável por controlar e operar o Sistema Interligado Nacional, que abrange as regiões Sul, Sudeste, Centro-Oeste, Nordeste e parte do Norte do país, podendo realizar troca de energia entre regiões, caso seja necessário. A integração permite também que as usinas hidrelétricas e termelétricas possam operar em regime de complementaridade. Assim, o ONS avalia quais usinas devem operar e quais devem permanecer em espera, para que o volume de produção seja adequado ao consumo. Cabe ressaltar que a preferência se dá para a energia hidrelétrica, que apresenta menor custo ao consumidor, além de ser mais abundante (ANEEL, 2008).

A energia elétrica constitui um importante segmento de infraestrutura do país e está diretamente relacionada ao crescimento econômico, como foi observado nas últimas décadas. Isso faz com que a sua produção adequada à demanda represente um fator relevante para a manutenção do crescimento da economia. No Brasil, existe a preocupação em elevar suas bases geradoras de energia, o que é evidenciado com a realização de fóruns nacionais, compostos por agentes ligados à questão energética, onde são discutidas medidas e metas para o setor (Viana e Silva, 2014).

Além disso, o custo de armazenamento da eletricidade torna inviável tal prática, o que faz com que seja necessário que a geração seja produzida conforme a demanda. Havendo disponibilidade de geração, problemas de escassez (como a crise do apagão ocorrida entre os anos de 2001 e 2002) podem ser evitados. Porém, uma grande capacidade de geração ociosa indica que ocorreu um erro de planejamento e o investimento feito na unidade geradora levará muito tempo para ser recuperado.

Em De Assis Cabral et al. (2017) é enfatizada a importância de se prever corretamente a demanda por eletricidade, onde é mostrado que as empresas geradoras devem garantir o fornecimento demandado com uma tolerância de 3% para mais ou para menos, sob pena de serem multadas pela agência reguladora. Assim, estas empresas devem se utilizar das melhores ferramentas disponíveis para realizar previsão da demanda de energia.

Portanto, o conhecimento prévio da evolução do consumo de energia elétrica pode, além de evitar problemas de abastecimento, servir de base para o planejamento do setor energético de um país e nortear políticas de eficiência energética, de modo que é fundamental que se estude a demanda de energia e os fatores que a influenciam. Para isso, é de suma importância buscar as melhores técnicas para que se realize a modelagem da demanda por eletricidade, identificando modelos com boa capacidade preditiva.

1.2 A modelagem

Dentre as metodologias mais utilizadas para analisar a demanda por energia elétrica, bem como realizar previsões, encontra-se a abordagem econométrica, que faz uso de técnicas estatísticas para análise de séries temporais. Um conjunto de observações, ordenadas no tempo, de qualquer fenômeno aleatório pode ser definido como uma série temporal (Bayer, 2010). Observar e entender o seu comportamento se faz importante para tomar decisões, e até mesmo realizar previsões.

Nos últimos anos, trabalhos como os de Schmidt e Lima (2004), Irffi et al. (2009), Viana e Silva (2014) e De Oliveira et al. (2018) utilizaram a abordagem econométrica para realizar estudo da demanda por eletricidade no Brasil. No cenário mundial, esse tipo de abordagem foi recentemente utilizada nos trabalhos de Al-Musaylh et al. (2018), Jamil (2020) e Murshed (2021), por exemplo.

Nessa abordagem, a presença de tendência estocástica (ou raiz unitária) na série, o que caracteriza a não estacionariedade, faz com que as previsões se tornem mais imprecisas conforme o distanciamento do último ponto da amostra aumente (Silveira et al., 2019). Segundo Gujarati (2011), uma série é estacionária se sua média e sua variância são invariantes ao longo do tempo e o valor da covariância entre dois períodos depende somente do grau de defasagens entre as observações e não do período efetivo de tempo em que a covariância é calculada.

Deste modo, é importante identificar a presença de raiz unitária em uma série temporal a fim de encontrar a técnica que melhor se aplica às características da série estudada. Para isso são utilizados os denominados testes de raiz unitária. Dentre estes testes, um dos mais populares, é o proposto por Dickey e Fuller (1979), conhecido como teste DF, capaz de testar a presença de raiz unitária em modelos autorregressivos (AR). Cinco anos mais tarde, Said e Dickey (1984) propuseram uma modificação do teste DF, que passou a ser denominado teste ADF, onde são incluídos valores defasados da diferença da série trabalhada. Com isso, o teste ADF pode ser aplicado em séries que apresentem correlação nos resíduos.

Os testes DF e ADF consistem em testes de hipóteses em que a hipótese nula indica a presença de raiz unitária, implicando na não estacionariedade. Na verificação da veracidade desta hipótese nula, é utilizada a estatística τ e a decisão é tomada pela comparação de um valor calculado com um valor tabelado, obtido mediante simulações de Monte Carlo. Assim, de acordo com um nível de significância pré-estabelecido, toma-se a decisão de que há (ou não) indícios de presença de raiz unitária na série.

Na aplicação do teste ADF, existe a necessidade de selecionar o número de defasagens a serem aplicadas para realização do teste. No entanto, ainda não existe um consenso na literatura, tanto para a determinação da máxima defasagem, quanto para a determinação do número de defasagens ideal. Portanto, cabe um maior aprofundamento nesta etapa de aplicação do teste ADF.

Para avaliar o desempenho do teste, podem ser construídos experimentos com diferentes características e tomar seus resultados. Deste modo, para que se possa avaliar os resultados

e realizar comparações é pertinente optar pela construção de um modelo que seja capaz de representar os resultados do teste em função de seus parâmetros.

Na literatura existem diversos tipos de modelos que podem ser utilizados para tal fim, começando pela regressão linear, que tem pressupostos um tanto rígidos para a variável de estudo, como a normalidade, a independência das observações e que as variâncias sejam constantes. Essas condições nem sempre são satisfeitas, fazendo-se necessário recorrer a alternativas. Dentre elas, pode-se optar pela transformação dos dados e pela utilização de estatísticas não paramétricas, como por exemplo a distribuição qui-quadrado. Outra alternativa é trabalhar com os modelos lineares generalizados (GLM), que trabalham com funções de ligação e permitem considerar diversas distribuições para a variável resposta, desde que pertençam a família exponencial. Por fim, os modelos aditivos generalizados para localização, escala e forma (*Generalized Additive Models for Position, Scale and Shape* - GAMLSS) permitem modelar todos os parâmetros da variável resposta, e não apenas os parâmetros de localização. Assim, é possível trabalhar com uma ampla gama de distribuições, independentemente de pertencer a família exponencial, desde que as observações sejam independentes.

1.3 Hipótese da Pesquisa

Diante do exposto, a hipótese de pesquisa proposta é a seguinte: É possível desenvolver uma estratégia para determinar a seleção de defasagens do teste ADF, em séries que possam ser modeladas por AR(1), de acordo com técnicas apresentadas na literatura, baseada nas características da série, que seja capaz de melhorar o desempenho do teste.

1.4 Objetivos

Os objetivos do trabalho são organizados em objetivo geral e objetivos específicos e serão detalhados nas próximas subseções.

1.4.1 Objetivo geral

Definir uma estratégia a ser utilizada para a determinação do número máximo e ótimo de defasagens na identificação da presença de raiz unitária pelo teste ADF, com base em técnicas descritas na literatura.

1.4.2 Objetivos específicos

- Analisar o teste ADF, identificar pontos fracos e pontos fortes do teste, por meio de análise bibliométrica;
- Comprovar, através de simulações, a influência do número de defasagens na proporção de rejeição da hipótese nula do teste ADF;

- Analisar o resultado do teste ADF em diferentes condições;
- Definir uma estratégia de seleção de defasagens na aplicação do teste ADF;
- Aplicar o teste, com a metodologia proposta, em uma série de dados real, relacionada ao consumo de eletricidade, e avaliar o resultado.

1.5 Justificativa

A determinação do número de defasagens na aplicação do teste ADF pode afetar diretamente a tomada de decisão sobre o resultado do teste. Mesmo que sejam mantidos todos os fatores inalterados, se apenas o critério de seleção de defasagens for alterado, o resultado do teste pode ser modificado. Diante da importância deste teste na modelagem de séries temporais, se faz necessário tomar a decisão sobre seu resultado da maneira mais assertiva possível.

Conforme o exposto, considerando que existem diversas maneiras de determinar os números ótimo e máximo de defasagem no teste ADF e nenhuma delas é considerada unanimidade na literatura, se faz necessário encontrar qual estratégia de seleção de defasagens apresenta a maior taxa de acerto no resultado do teste, levando em consideração as características da série estudada.

1.6 Estrutura do trabalho

O trabalho está estruturado em seis capítulos, conforme a seguir:

Capítulo 1 - apresenta a definição e a importância do tema, a modelagem, a hipótese da pesquisa, bem como os objetivos geral e específico. Além disso, é apresentada a justificativa pela escolha do tema.

Capítulo 2 - mostra a revisão de literatura, onde são abordados os conceitos de estacionariedade, os testes DF e ADF e a modelagem realizada no trabalho. Ainda neste capítulo, é apresentada a revisão sistemática, com análise bibliométrica e metassíntese.

Capítulo 3 - trata da metodologia proposta para alcançar os objetivos. Neste capítulo são elaboradas as simulações preliminares e as simulações utilizadas para a proposição da estratégia, com o planejamento e a análise dos experimentos. Também é apresentada uma proposta para simulações complementares.

Capítulo 4 - expõe e discute os resultados encontrados na aplicação da metodologia. São apresentados os resultados das simulações preliminares, das simulações para séries não estacionárias e estacionárias, bem como das simulações complementares. Ao final é apresentada uma estratégia para a definição de defasagens na aplicação do teste ADF.

Capítulo 5 - relata um estudo de caso considerando a aplicação da estratégia proposta. O teste ADF é aplicado em uma série de consumo residencial de eletricidade, com os dados em nível e em primeira diferença.

Capítulo 6 - manifesta as considerações finais sobre o desenvolvimento da tese e apresenta recomendações para a realização de trabalhos futuros.

2 REVISÃO DE LITERATURA

Esta seção está dividida em quatro subseções. A primeira aborda a conceituação de estacionariedade, a segunda trata diretamente dos testes DF e ADF, enquanto a terceira trata da modelagem utilizada para a análise dos resultados. Na quarta subseção, é apresentada a revisão sistemática.

2.1 Estacionariedade

Em Morettin e Toloi (2018) um processo estocástico é definido como uma família de variáveis aleatórias supondo-as definidas em um mesmo espaço de probabilidade, sendo controlado por leis probabilísticas. Em linhas gerais, um processo aleatório é estacionário se sua média e sua variância são invariantes ao longo do tempo e o valor da covariância entre dois períodos depende somente do grau de defasagens entre as observações e não do período efetivo de tempo em que a covariância é calculada. Ainda, na literatura este tipo de processo é denominado processo estocástico fracamente estacionário ou estacionário no sentido amplo.

A série temporal será denominada estritamente estacionária quando todos os momentos de sua distribuição de probabilidade, e não apenas a média e a variância, são invariantes ao longo do tempo. Entretanto se o processo estacionário for normal, o processo estocástico fracamente estacionário também será estritamente estacionário (Gujarati, 2011). Para Shumway e Stoffer (2000), se uma série temporal for estritamente estacionária, todas as funções de distribuição multivariada para subconjuntos de variáveis devem concordar com suas contrapartes, considerando um conjunto deslocado para todos os valores de um parâmetro de deslocamento h .

Além disso, Gujarati (2011) define uma série como não estacionária quando ela não satisfaz a definição anterior. Ou seja, sua média, sua variância ou ambas variam em função do tempo. Em Bueno (2008) é afirmado que uma série não estacionária não pode ser trivialmente estimada, pois é impossível estimar todos os momentos da série.

De acordo com Bueno (2008), a variância não condicional de um modelo AR(1) pode ser dada por: $var(y_t) = \frac{1}{1-\phi^2}$, onde y_t é a série a ser modelada e ϕ o coeficiente de y_{t-1} . Deste modo, é simples perceber que se o valor de ϕ tender a 1, a variância tenderá ao infinito. Quando $\phi = 1$ a série é caracterizada pela presença de raiz unitária, assim, a série é dita não estacionária. Em Bueno (2008), ainda é abordado o caso em que a série pode ser estacionária em torno de uma tendência determinística, neste caso basta adicionar uma variável explicativa t ao modelo.

Segundo Shumway e Stoffer (2000), é possível forçar uma série a se tornar estacionária através de um procedimento de diferenciação, o que frequentemente é necessário quando são utilizadas técnicas estatísticas, pois, de acordo com Morettin e Toloi (2018), a maioria dos procedimentos de análise estatística de séries temporais faz esta suposição. Por outro lado, o trabalho de Engle e Granger (1987) afirma que este procedimento remove toda a lógica de curto prazo das séries econômicas.

Quando a série é não estacionária torna-se muito importante a utilização de teoria econômica

para explicar a relação entre variável explicada e variável explicativa. Em Granger e Newbold (1974) é abordado o tema da regressão espúria, em que pode se encontrar relação econométrica entre duas séries, com alto valor de R^2 (coeficiente de determinação) e valor-t (relativo a uma distribuição t) significativo, mesmo sem qualquer significado econômico. O que implicaria em estimativas dos coeficientes de regressão ineficientes, previsões baseadas nas equações de regressão sub-ótimas e nos testes usuais de significância dos coeficientes se tornando inválidos. Este fenômeno é mais comum em séries com raiz unitária. Em Gujarati (2011) é afirmado que mesmo para amostras muito grandes o fenômeno da regressão espúria pode persistir se a série não for estacionária, e se deve suspeitar que está ocorrendo este fenômeno quando a regressão apresentar um valor muito baixo para o teste de Durbin-Watson (teste utilizado para verificar a independência dos resíduos).

Dentre as maneiras utilizadas para identificar a presença de raiz unitária em uma série temporal, pode ser citada a análise gráfica, com base em gráficos em linhas e das funções de autocorrelação (FAC) e autocorrelação parcial (FACP), entretanto, segundo (Arltová e Fedorová, 2016), esta abordagem deve ser utilizada por pessoas experientes capazes de identificar a existência da propriedade na série. Por outro lado, técnicas analíticas apresentam melhores resultados e contam com a vantagem de não dependerem da experiência do analista.

As técnicas analíticas para identificar a presença de raiz unitária em séries temporais vêm sendo tema de diversos estudos ao longo do tempo. Estes estudos se baseiam em elaborar testes para verificar a presença (ou ausência) de raiz unitária nas séries temporais, os chamados testes de raiz unitária.

Os testes de DF e ADF, que serão detalhados na próxima seção, foram uns dos primeiros a surgirem na literatura e continuam sendo largamente utilizados em trabalhos atuais. Entretanto, existem outros testes de raiz unitária já tradicionais, como, por exemplo, o teste de Phillips-Perron, proposto em Phillips e Perron (1988) que testa a presença de raiz unitária através de métodos estatísticos não paramétricos, podendo assim tratar da correlação serial nos termos de erro sem adicionar os termos de diferença defasados.

Outro teste bastante utilizado é o teste KPSS, proposto em Kwiatkowski et al. (1992) e que tem a hipótese nula oposta a do ADF e do PP, sendo a estacionariedade. Os autores trabalham com essa hipótese para evitar o problema de baixo poder do teste ADF, de modo que o teste KPSS possa ser utilizado como um teste complementar ao ADF.

No estudo de Elliott, Rothenberg e Stock (1992) são retirados, através de uma modificação, os termos determinísticos do teste ADF, pois os autores argumentam que assim poderiam aumentar o poder do teste. Deste modo, o denominado teste DF-GLS consiste em um procedimento para remoção de constante e tendência da equação do teste ADF e na estimação uma nova equação.

2.2 Testes DF e ADF

O testes DF e ADF, propostos em Dickey e Fuller (1979) e Said e Dickey (1984) respectivamente, vêm sendo amplamente utilizados na literatura de séries temporais a fim de identificar a presença de raiz unitária. Os procedimentos para a aplicação destes testes são apresentados a seguir.

Consideremos o modelo AR(1) apresentado na Equação 2.1:

$$y_t = \phi y_{t-1} + u_t, \quad (2.1)$$

onde u_t é um ruído branco de média zero e variância constante, $u_t \sim RB(0, \sigma^2)$.

O Teste de Dickey-Fuller (Teste DF) utiliza o modelo da Eq.2.1 e considera a hipótese nula da presença de raiz unitária, conforme a Eq. 2.2:

$$H_0 : \phi = 1, \quad (2.2)$$

e como hipótese alternativa a Eq. 2.3:

$$H_1 : |\phi| < 1. \quad (2.3)$$

Considerando as mesmas hipóteses, podem ser incluídos termos determinísticos na Eq.2.1, resultando na Eq.2.4 que considera a presença de um intercepto, e na Eq.2.5 que considera a presença de um intercepto e de uma tendência linear:

$$y_t = \phi y_{t-1} + \beta_1 + u_t, \quad (2.4)$$

$$y_t = \phi y_{t-1} + \beta_1 + \beta_2 t + u_t, \quad (2.5)$$

onde a constante β_1 representa o intercepto e a constante β_2 o coeficiente associado à tendência.

Em Choi (2015) é suposto um processo de geração de dados que é estacionário e estimado por mínimos quadrados, $y_t = \bar{\beta}_1 + \bar{\phi} y_{t-1} + \bar{u}_t$ onde $\bar{\beta}_1$ e $\bar{\phi}$ são estimadores de mínimos quadrados, T é o comprimento da série, e $\bar{u}_t$ são os resíduos dos mínimos quadrados. Então, considerando $y_- = \frac{1}{T-1} \sum_{t=2}^T y_{t-1}$ e $\bar{y} = \frac{1}{T-1} \sum_{t=2}^T y_t$, é mostrado que

$$\begin{aligned} \bar{\phi} &= \frac{\sum_{t=2}^T (y_{t-1} - \bar{y}_-)(y_t - \bar{y})}{\sum_{t=2}^T (y_{t-1} - \bar{y}_-)^2} \\ &= \frac{\sum_{t=2}^T ((t-1) - \frac{T}{2})(t - \frac{(T+2)}{2})}{\sum_{t=2}^T ((t-1) - \frac{T}{2})^2} + o_p(1) \xrightarrow{p} 1. \end{aligned}$$

Isso indica que o estimador do coeficiente do AR(1) converge para 1 em probabilidade mesmo sobre a hipótese alternativa. Então, realizar o teste de raiz unitária sobre $\bar{\phi}$ tende a ser inconsistente. Deste modo, o trabalho de Said e Dickey (1984) considera as estatísticas $T(\hat{\phi}-1)$, $T(\bar{\phi}-1)$ e $T(\phi^*-1)$, para os modelos das equações 2.1, 2.4 e 2.5, respectivamente.

Adicionalmente, são consideradas as estatísticas da Eq.2.6

$$\hat{\tau} = \frac{\hat{\phi} - 1}{\sqrt{\hat{\sigma}_u^2 (\sum_{t=2}^T y_{t-1}^2)^{-1}}}, \quad \bar{\tau} = \frac{\bar{\phi} - 1}{\sqrt{\bar{\sigma}_u^2 (\sum_{t=2}^T \bar{y}_{t-1}^2)^{-1}}}, \quad \tau^* = \frac{\phi^* - 1}{\sqrt{\sigma_u^{*2} (\sum_{t=2}^T y_{t-1}^{*2})^{-1}}}, \quad (2.6)$$

para os modelos das equações 2.1, 2.4 e 2.5, respectivamente.

Ainda, a distribuição assintótica das estatísticas são dadas nas equações 2.7, 2.8 e 2.9, considerando o movimento Browniano $\{W(r), 0 \leq r \leq 1\}$:

$$\hat{\tau} = \frac{\int_0^1 W(r) dW(r)}{\sqrt{\int_0^1 W^2(r) dr}}, \quad (2.7)$$

$$\bar{\tau} = \frac{\int_0^1 \bar{W}(r) dW(r)}{\sqrt{\int_0^1 \bar{W}^2(r) dr}} \quad (2.8)$$

$$\tau^* = \frac{\int_0^1 W^*(r) dW(r)}{\sqrt{\int_0^1 W^{*2}(r) dr}}. \quad (2.9)$$

Os valores empíricos das estatísticas são tabelados por Dickey e Fuller (1979) para os níveis de significância de 1, 5 e 10%, e de acordo com os autores, as estatísticas apresentadas nas equações 2.7, 2.8 e 2.9 tendem a ter maior poder se comparadas às estatísticas t de Student. Assim, se o valor calculado para a estatística for menor do que o valor crítico encontrado na tabela, a hipótese nula da presença de raiz unitária deve ser rejeitada.

Conforme apresentado, o teste DF considera u_t como um ruído branco, todavia essa consideração não pode ser tomada como verdadeira em grande parte das séries de dados reais. Segundo Bueno (2008), este problema pode causar distorção no poder do teste, entretanto a questão pode ser resolvida sem maiores problemas.

Consideremos um processo autorregressivo de ordem k com a presença de raiz unitária da Eq. 2.10:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_{k-1} y_{t-k+1} + \phi_k y_{t-k} + u_t. \quad (2.10)$$

Segundo Bueno (2008), para a aplicação do teste, deve-se considerar tantas variáveis autorregressivas quanto forem necessárias para que o teste de resíduos não rejeite a hipótese nula de que este se trate de um ruído branco. Assim, partindo da Eq. 2.10 é adicionado e subtraído $\phi_k y_{t-k+1}$, e em seguida $(\phi_{k-1} + \phi_k) y_{t-2}$, até que o procedimento seja realizado p vezes e obtida a Eq. 2.11, onde $\Delta y_t = y_t - y_{t-1}$

$$\Delta y_t = \alpha y_{t-1} + \sum_{i=1}^{k-1} \lambda_i \Delta y_{t-i} + u_t, \quad (2.11)$$

em que: $\alpha = -(1 - \sum_{i=1}^p \phi_i)$ e $\lambda_i = -\left(-\sum_{j=i}^{p-1} \phi_{j+1}\right)$.

Deste modo, na Eq. 2.11 podem ser incluídos os termos determinísticos e o teste pode ser feito utilizando os valores críticos encontrados em Dickey e Fuller (1979), neste caso é denominado Teste de Dickey-Fuller Aumentado (Teste ADF).

De acordo com Bueno (2008) o procedimento anterior pode ser adotado também para o caso de modelos ARIMA(m,1,n) uma vez que é possível transformar um MA(q) em um AR(∞). Entretanto, a estimação de um modelo com infinitas defasagens seria um problema, assim, em Said e Dickey (1984) é mostrado que um modelo ARIMA(m,1,n) pode ser aproximado por um modelo ARIMA(k,1,0) sem que a distribuição assintótica da estatística de teste do ADF seja alterada, desde que $k \leq T^{1/3}$.

Dessa maneira, k deve tender ao infinito a medida que o T o faz, porém k tende ao infinito a uma taxa menor do que T. Além disso, segundo Choi (2015), k deve ser grande o suficiente para que seja fornecida uma aproximação satisfatória do modelo de dimensão infinita. Diante disso, o autor afirma que um valor elevado de k tende a gerar um bom tamanho do teste, porém baixo poder, enquanto um valor muito pequeno de k pode gerar um tamanho empírico pobre e um alto poder do teste.

A seleção do número de defasagens a ser incluído no modelo é tema de diversos estudos nos últimos anos. Grande parte destes trabalhos segue a metodologia de definir um número máximo de defasagens, e realizar algum tipo de teste para selecionar o número ideal a ser incluído no modelo. Em relação ao número máximo, as equações 2.12 e 2.13 apresentam os critérios propostos por Schwert (1989), aqui denominados l_4 e l_{12} .

$$l_4 = \text{int}[4(T/100)^{1/4}], \quad (2.12)$$

$$l_{12} = \text{int}[12(T/100)^{1/4}]. \quad (2.13)$$

Já o trabalho de Hyndman e Athanasopoulos (2018) apresenta o critério aqui chamado de l_h , da Eq. 2.14 para definir o número máximo de defasagens.

$$l_h = \text{int}[4(T/100)^{2/9}]. \quad (2.14)$$

Após a definição do número máximo de defasagens, busca-se determinar a defasagem ótima, a fim de encontrar a melhor relação entre o poder do teste e seu tamanho empírico. Trabalhos recentes utilizam critérios de informação para a determinação do número ótimo de defasagens a ser incluído no modelo. Em Cavaliere et al. (2015), por exemplo, é indicado que a seleção de defasagens através de critérios de informação apresenta melhor desempenho em relação a sequenciais testes-t.

Os critérios de informação mais utilizados na literatura são o critério de Akaike (AIC) proposto em Akaike (1973) e representado pela Eq. 2.15; critério de informação Bayesiano (BIC), apresentado em Schwartz (1978), mostrado na Eq. 2.16; e o critério de informação de Hannan-Quinn (HQC), relatado em Hannan e Quinn (1979) e que pode ser visualizado na Eq.

2.17:

$$AIC(k) = Ln(\hat{\sigma}_u^2(k)) + 2k/T, \quad (2.15)$$

$$BIC(k) = Ln(\hat{\sigma}_u^2(k)) + kLn(T)/T, \quad (2.16)$$

$$HQC(k) = Ln(\hat{\sigma}_u^2(k)) + 2kLn(Ln(T))/T, \quad (2.17)$$

onde $\hat{\sigma}_u^2(k)$ é a estimativa da variância dos resíduos ao utilizar k defasagens.

A Fig.2.1 apresenta um fluxograma da aplicação do teste.

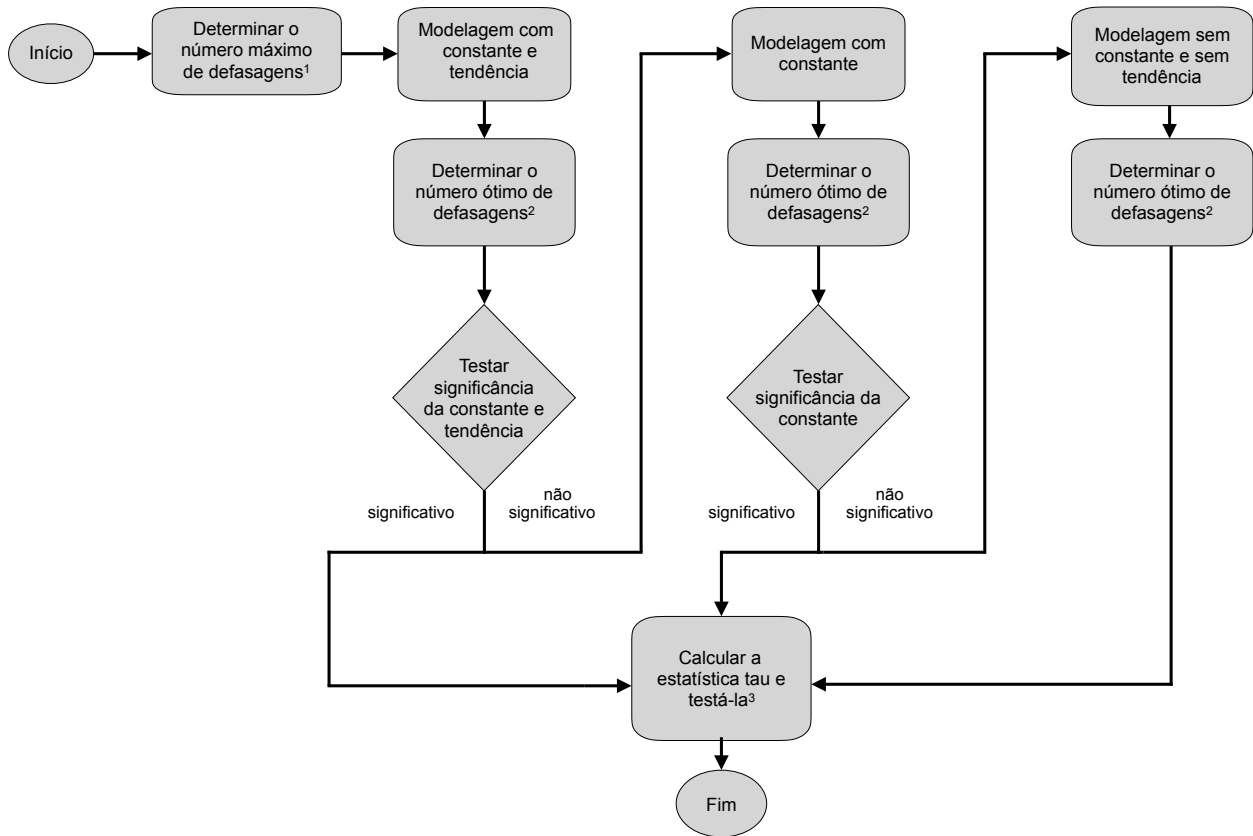


Figura 2.1: Fluxograma para a aplicação do teste ADF

1. Pode ser calculado através de as equações 2.12, 2.13 ou 2.14; 2. Pode ser determinado através das equações 2.15, 2.16 ou 2.17; 3. Ao rejeitar a hipótese nula, assume-se a estacionariedade.

2.3 Modelagem GAMLSS

A modelagem GAMLSS (modelos aditivos generalizados para locação, escala e forma) aborda tipos de modelos de regressão semiparamétricos. Esse tipo de modelagem, proposto por Rigby e Stasinopoulos (2005), a fim de superar limitações dos modelos lineares generalizados e dos modelos aditivos generalizados, permite que todos os parâmetros da distribuição da

variável resposta sejam modelados por meio de funções lineares ou não lineares das variáveis explicativas.

De acordo com Roquim (2018), os GAMLSS são um modelo mais avançado e flexível que não só propõe uma nova classe de modelos, mas também inclui modelos existentes, assumindo como premissa que as observações das variáveis são independentes. A modelagem GAMLSS é especialmente útil para dados com distribuições incomuns ou que mudam com as variáveis explicativas, uma vez que o ajuste de modelos aceita qualquer distribuição da variável resposta, independente de pertencer à alguma família de distribuições. Em Rigby et al. (2019) a forma semi-paramétrica do GAMLSS pode ser definida pela Equação 2.18, onde $Y_1, Y_2, Y_3, \dots, Y_n$ são observações independentes da variável resposta de modo que $Y_i \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{\lambda})$ em que D é a distribuição dos quatro parâmetros, que normalmente são definidos como: μ sendo um parâmetro de locação, σ um parâmetro de escala, v e λ são parâmetros de forma da distribuição, geralmente associados à assimetria e à curtose, respectivamente. Diante disso, pode-se escrever na forma de vetor:

$$Y \stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{\lambda}),$$

onde

$$\begin{aligned} g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(\mathbf{x}_{11}) + \dots + s_{1j_1}(\mathbf{x}_{1j_1}) \\ g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(\mathbf{x}_{21}) + \dots + s_{2j_2}(\mathbf{x}_{2j_2}) \\ g_3(\boldsymbol{v}) &= \boldsymbol{\eta}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(\mathbf{x}_{31}) + \dots + s_{3j_3}(\mathbf{x}_{3j_3}) \\ g_4(\boldsymbol{\lambda}) &= \boldsymbol{\eta}_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(\mathbf{x}_{41}) + \dots + s_{4j_4}(\mathbf{x}_{4j_4}), \end{aligned} \tag{2.18}$$

em que $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, $\boldsymbol{\eta}_3$ e $\boldsymbol{\eta}_4$ são chamados de preditores de μ, σ, v e λ , respectivamente; g_k são as funções de ligação; $\mathbf{X}_k$ são as matrizes que incorporam os termos aditivos lineares no modelo; e $\boldsymbol{\beta}_k$ são os coeficientes lineares dos parâmetros. Por fim, $s_k j_k(\mathbf{x}_k j_k)$ representam as funções de suavização. Cabe destacar que cada parâmetro pode receber uma função de ligação diferente.

Considerando a Eq. 2.18, para o caso em que só é modelado $g_1(\boldsymbol{\mu})$, o GAMLSS é reduzido a um modelo aditivo generalizado (GAM), apresentado em Hastie e Tibshirani (1990). Conforme os autores, existe a necessidade da distribuição da variável resposta pertencer à família exponencial. Todavia, como afirmado por Fernandes et al. (2019), esta suposição pode ser relaxada utilizando-se os algoritmos de estimação presentes no pacote `gamlss`, do software `r`, a saber: algoritmo RS e algoritmo CG. Ainda, se os termos $s_{11}(\mathbf{x}_{11}) + \dots + s_{1j_1}(\mathbf{x}_{1j_1})$ forem nulos, o modelo é reduzido a um GLM, mostrado em Nelder e Wedderburn (1972), relaxando-se novamente a suposição de que a resposta deve pertencer à família exponencial.

Segundo Roquim (2018), os modelos semi-paramétricos são difíceis de descrever em termos matemáticos. Diante disso, informações detalhadas sobre a modelagem matemática podem ser encontradas em Rigby e Stasinopoulos (2005). Ainda conforme Roquim (2018), para se

trabalhar com GAMLSS, é necessário utilizar ferramentas capazes de lidar com a diversidade de distribuições e termos presentes no modelo, bem como diferentes conjuntos de variáveis explicativas e funções de ligação para cada parâmetro do modelo.

Dentre as famílias de distribuições que se aplicam ao caso de proporções, encontra-se a Logit Normal (LOGITNO), que é uma distribuição de probabilidade contínua. A LOGITNO é uma transformação da distribuição normal que é restrita ao intervalo (0, 1) e é construída através da transformação logit dos parâmetros da distribuição normal.

Em Rigby et al. (2019) é definida a função densidade de probabilidade da distribuição logit normal, $LOGITNO(\mu, \sigma)$, conforme a Equação 2.19:

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y(1-y)} \exp\left(\frac{\{Ln[y/(1-y)] - Ln[\mu/(1-\mu)]\}^2}{2\sigma^2}\right), \quad (2.19)$$

para $0 < y < 1$, onde $0 < \mu < 1$ e $\sigma > 0$.

Cabe ressaltar que, conforme a Equação 2.19, é possível notar que a distribuição LOGITNO utiliza apenas os parâmetros de locação e de escala, não sendo utilizados os parâmetros de forma.

Em um modelo de regressão linear simples os resíduos podem ser obtidos pela simples diferença entre os valores observados e estimados. O problema é que é difícil de generalizar essa análise para qualquer distribuição além da normal. Para modelos GAMLSS, são usados os resíduos de quantil normalizados (aleatorizados), definidos em Dunn e Smyth (1996). Segundo Stasinopoulos et al. (2017), a principal vantagem dos resíduos de quantil normalizados (aleatorizados) é que, mesmo com diferentes distribuições da variável de resposta, seus valores seguem uma distribuição normal padrão, assumindo que o modelo esteja correto. Como a verificação da normalidade é bem estabelecida na estatística, esses resíduos permitem uma verificação fácil da adequação de um modelo GAMLSS.

Considerando que a distribuição $f(y; \theta)$ é ajustada às observações y_i , para $i = 1, 2, \dots, n$, os resíduos normalizados (aleatorizados) de quantil ajustados, propostos por Dunn e Smyth (1996), são dados por $\hat{r}_i = \Phi^{-1}(\hat{u}_i)$, em que Φ^{-1} é a função de distribuição cumulativa inversa de uma variável normal padrão e $(\hat{u}_i)$ representam os resíduos quantílicos. O processo de transformação de probabilidade integral é demonstrado graficamente em Stasinopoulos et al. (2017), para variáveis contínuas e discretas.

Os *worm plots*, ou gráficos de minhoca, apresentados em Buuren e Fredriks (2001) estão entre as principais maneiras de analisar os resíduos de um GAMLSS. Segundo Buuren (2007), o gráfico de minhoca, que consiste em uma coleção de gráficos Q-Q sem tendência, é uma ferramenta útil para avaliar o quão bem um modelo estatístico se ajusta aos dados, identificar áreas onde o ajuste pode ser aprimorado e comparar o ajuste de diferentes modelos.

Em Stasinopoulos et al. (2017) são apresentadas diferentes formas do *worm plot*, a deficiência relacionada aos resíduos e onde deve ser o ajuste na distribuição, conforme consta na Tabela 2.1.

Tabela 2.1: Interpretações do *worm plot*

Formato do Gráfico	Resíduos	Parâmetro ajustado
nível: acima da origem	média muito alta	locação subestimada
nível: abaixo da origem	média muito alta	locação superestimada
linha: inclinação positiva	variância muito alta	escala subestimada
linha: inclinação negativa	variância muito baixa	escala superestimada
U	assimetria positiva	assimetria subestimada
U invertido	assimetria negativa	assimetria superestimada
S com curva da esquerda para cima	leptocúrtica	curtose subestimada
S com curva da esquerda para baixo	platicúrtica	curtose superestimada
Fonte: Stasinopoulos et al. (2017)		

Considerando que os pontos estejam dentro das bandas de confiança e o gráfico não apresente nenhum dos problemas descritos na Tabela 2.1, existem indícios de que o modelo está bem ajustado.

2.4 Revisão Sistemática

Nesta seção é apresentada a revisão sistemática, onde é investigada a produção científica sobre o tema estudado. Através desta revisão é possível identificar, selecionar e avaliar trabalhos relevantes e atualizados relacionados à área da pesquisa. Inicialmente foi realizada a análise bibliométrica, onde foram definidos critérios para a seleção das publicações potencialmente relevantes para o estudo da seleção de defasagens no teste ADF. Em seguida, foi realizada a metassíntese, etapa em que é realizado um estudo mais aprofundado dos artigos a fim de identificar lacunas na área de estudo pesquisada.

2.4.1 Análise Bibliométrica

Foram escolhidas duas das maiores bases de dados de documentos científicos para a realização da análise bibliométrica: Web of Science e Scopus. Na base Web of Science a busca por documentos contendo as combinações de palavras-chave ocorreu por “*Tópico*”, enquanto na Scopus, a busca pelas palavras-chave ocorreu por “*Article title, Abstract, Keywords*”.

Para a pesquisa, foram selecionadas cinco combinações de palavras-chave. A Tabela 2.2 apresenta as combinações de palavras, bem como o número de documentos encontrado em cada base de dados para cada uma delas.

Tabela 2.2: Resultados das combinações nas bases de dados

Palavras-Chave	Web of Science	Scopus
unit root x electricity demand	60	21
unit root x lag order	37	24
unit root x lag selection	17	21
unit root x adf x lag	47	56
unit root x adf x stationarity	89	98

Deste modo, foram encontrados 250 resultados na Web of Science e 220 resultados na Scopus, totalizando 470 documentos. Em seguida, foram retirados os documentos duplicados em cada base de dados, resultando em 231 na Web of Science e 196 na Scopus. Logo foram deixados apenas os documentos caracterizados como *articles* e *proceedings papers*, em cada base de dados, restando 229 resultados na Web of Science e 191 resultados na Scopus. Na sequência, foram selecionados os artigos escritos nos idiomas inglês e espanhol, restando 225 na Web of Science e 185 na Scopus.

Dando continuidade, foram cruzados os artigos das duas bases de dados com o auxílio da biblioteca *bibliometrix* do *software R*. Com isso, foram eliminados os arquivos duplicados entre as bases de dados, obtendo um total de 305 artigos.

A próxima etapa foi a leitura e filtragem dos 305 documentos a fim de encontrar os que se enquadram diretamente com o tema da pesquisa. Após essa tarefa, permaneceram 29 arquivos, aproximadamente 10% do total dos que foram examinados na etapa.

Os 29 artigos selecionados para download e leitura integral possuem algumas características que cabem ser mencionadas. A Figura 2.2 apresenta a distribuição dos artigos ao longo dos anos, onde é possível notar que, dentre os artigos selecionados, a ocorrência de três artigos em um mesmo ano se fez presente apenas em 2013, 2018 e 2019, evidenciando que o tema está sendo escolhido para pesquisa nos últimos anos.

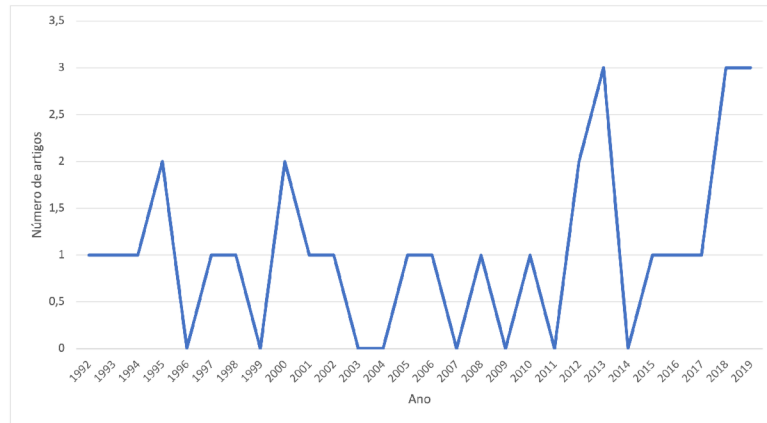


Figura 2.2: Histórico de produção anual

De acordo com a Figura 2.3, dentre os periódicos de maior relevância, destaca-se o *Oxford Bulletin of Economics and Statistics*, com três publicações. Seguido de *Applied Economics*, *Communications in Statistics: Simulation and Computation*, *Economics Reviews* e *Economics Letters*, com duas publicações cada.

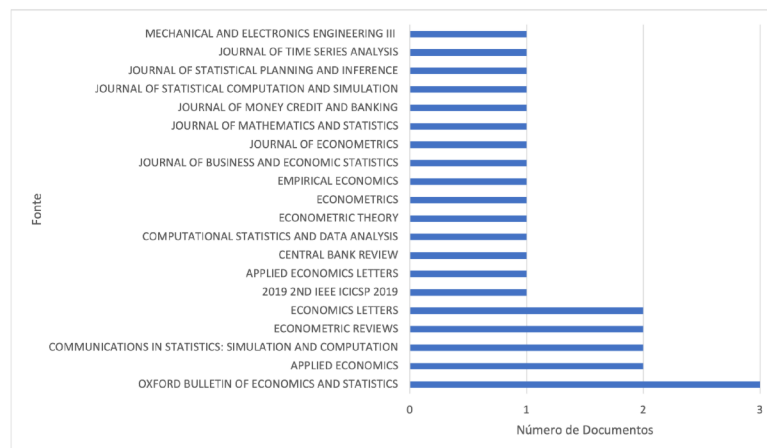


Figura 2.3: Publicações por Periódico

Em relação aos autores, a Figura 2.4 mostra que os que mais publicaram são Cheung Y, Lai K, e Taylor A, com três publicações cada. Seguidos por Cavaliere G e Smeekes S, com duas cada um.

Na Figura 2.5 é apresentada a produção de cada autor ao longo do tempo, neste quesito destaca-se Taylor A. que desenvolve estudos sobre o tema desde o ano de 2000.

Dentre os países com mais publicações sobre o tema, conforme a Figura 2.6 destacam-se Estados Unidos, com 10, seguido por Canadá e Reino Unido, com 4 cada, e pela República

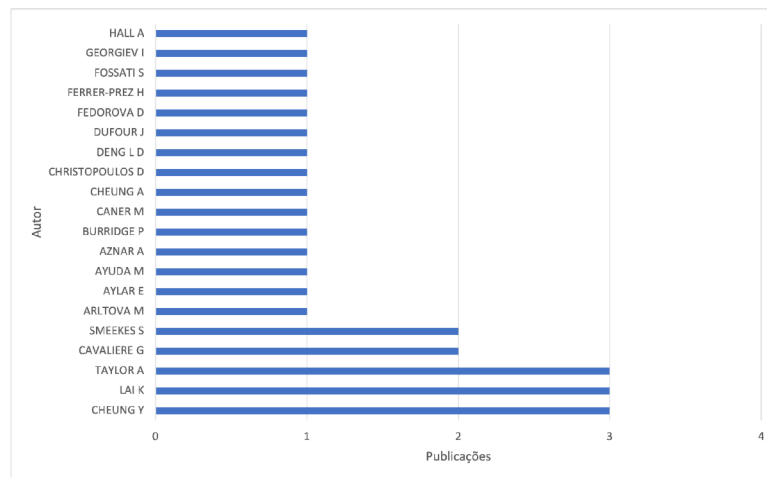


Figura 2.4: Autores mais relevantes

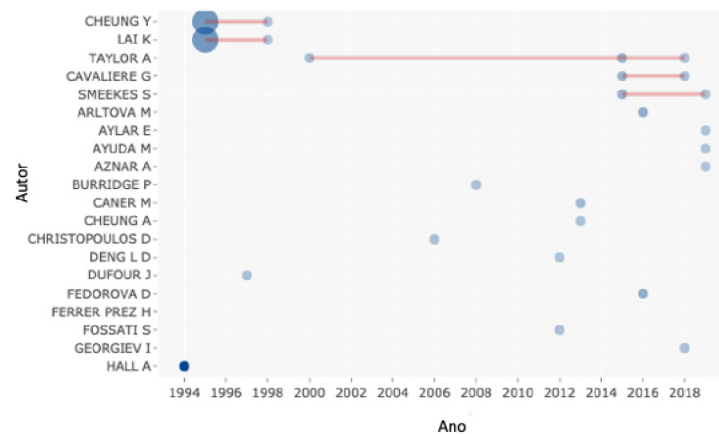


Figura 2.5: Produção por autor ao longo do tempo

Tcheca com 3 publicações. Cabe observar que estes quatro países representam 74,4% das publicações sobre o tema.

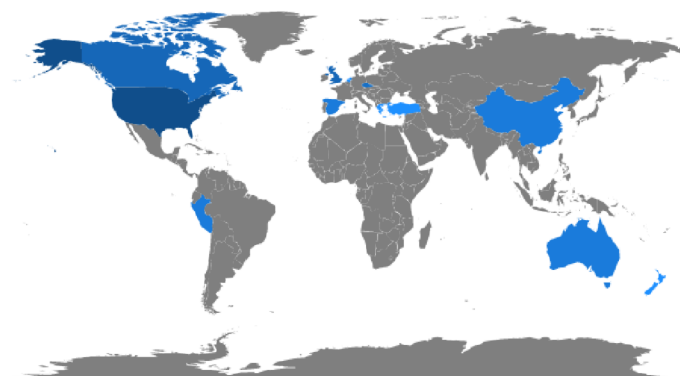


Figura 2.6: Produção por país

Em relação às universidades que publicaram mais artigos, conforme a Figura 2.7 destaca-se

Universidade de Lund, Universidade de Maastricht, Universidade de Montreal, Universidade de Bolonha e Universidade de Essex, com duas publicações cada.



Figura 2.7: Produção por Universidade

Os artigos mais citados são mostrados na Figura 2.8. Neste caso cabe observar que o artigo de Hall (1994) é o que apresenta maior número de citações: 386. Os demais trabalhos aparecem com um número menor de citações, cabendo menção a Cheung e Lai (1995) e Lopez et al. (2005).

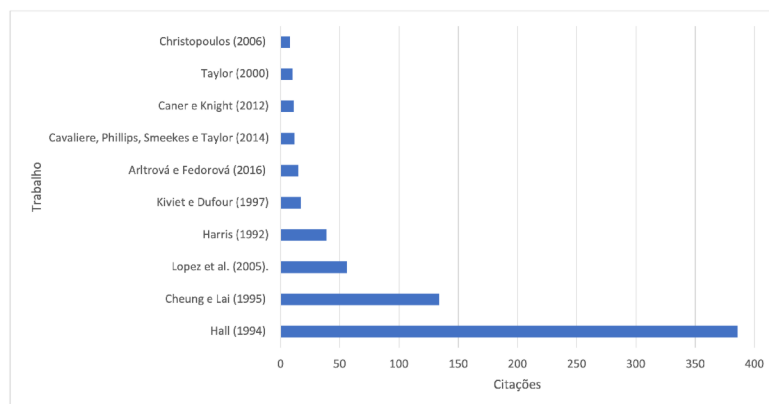


Figura 2.8: Citações por Documentos

Em relação a análise das referências dos artigos selecionados, o trabalho de Perron (2001) foi encontrado em doze artigos, seguido por Dickey e Fuller (1979), Elliott, Rothenberg e Stock (1992), Hall (1994), NG (2001) e Phillips e Perron (1988). Estes dados podem ser observados na Figura 2.9.

Na continuidade, foi feita a verificação de quais palavras ocorreram com mais frequência nos resumos dos artigos. Os resultados, compilados na Figura 2.10, mostram que as palavras *test* e *lag* foram as que mais se fizeram presentes, com 65 e 53 ocorrências, o que está de acordo com o tema pesquisado.

Cabe ainda salientar alguns dados relevantes: os artigos são de 23 periódicos diferentes, tiveram em média 25,55 citações por ano, foram elaborados por 45 autores diferentes, sendo 12

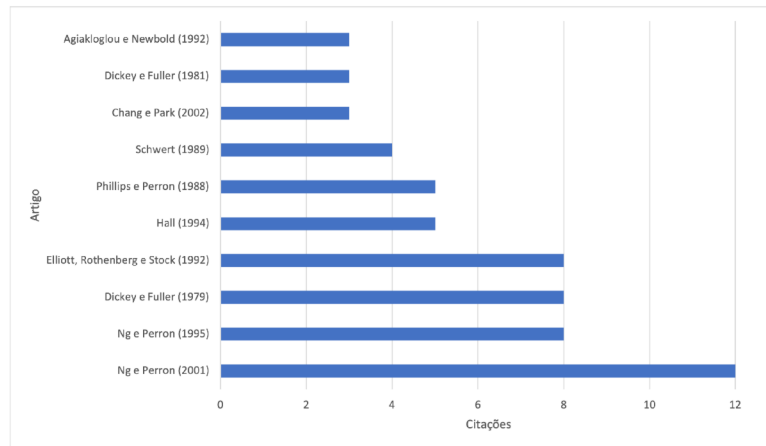


Figura 2.9: Referências mais citadas

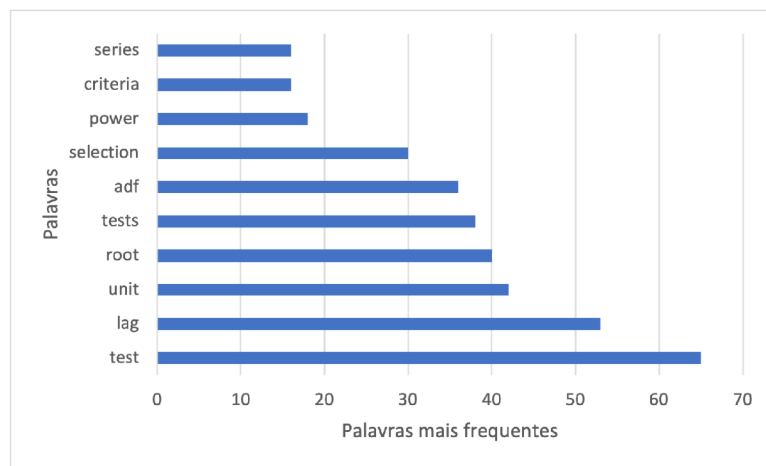


Figura 2.10: Palavras mais frequentes

destes em trabalho de único autor e 33 trabalhos de multiautores, resultando em 0,644 trabalho por autor.

2.4.2 Metassíntese

Em Harris (1992) são utilizados modelos gerados por ARIMA misto para avaliar o número de defasagens a ser utilizado na aplicação do teste ADF. É avaliada a influência do número de defasagens quando o teste ADF consistentemente rejeita a hipótese nula e alternativa quando ela é verdadeira. O autor compara a determinação do número de defasagens por AIC e um modo que considera o valor em que R^2 seja maximizado. Além disso, cada experimento é replicado 2000 vezes. Concluem que ao permitir que a ordem de autorregressão cresça com o tamanho da amostra, é possível obter um teste ADF com tamanho próximo ao nominal e poder adequado.

Em Hall (1994) é considerado o impacto de pré-testes baseados na seleção do modelo nos processos de seleção de defasagem do teste ADF. São utilizados critérios de informação e GSC, além disso, também são atribuídos valores fixos para k . São gerados modelos AR e MA e são

calculados valores de níveis de significância de acordo com o valor de defasagem selecionado ou critério adotado. Os resultados são apresentados para $T=100$ e 250 , com os valores de ϕ sendo $0,9$, $0,95$ e $1,00$. Por fim, os testes são aplicados em dados reais de inventário. O autor conclui que para o caso de modelos $ARIMA(p,1,0)$ seria mais satisfatório utilizar GSC, HQC, AIC ou BIC. Para os modelos AR em que nenhum dos valores é nulo, é recomendado utilizar HQC ou BIC. O autor conclui que é mais prudente estimar k por vários critérios diferentes antes de tomar uma decisão sobre a presença de raiz unitária.

O trabalho de Cheung e Lai (1995) tem por objetivo demonstrar o efeito da seleção de defasagens nos valores críticos do teste DF-GLS para amostras finitas, além de estimar os valores críticos corretos considerando o efeito da ordem de defasagem. Para isso, são utilizadas equações de superfícies de respostas e são utilizadas 296 combinações de T (valores entre 28 e 600) e k (valores de 0 a 8), considerando valores críticos a 5% e 25.000 replicações. São geradas $T + 50$ observações, onde as 50 primeiras são descartadas. Em seguida são calculados os valores críticos e colocados em uma tabela com cada valor de defasagem. Na continuidade, são aplicados os testes com os valores críticos estimados em séries reais de PIB per capita e taxas de câmbio. Os autores concluem que a seleção de defasagens pode afetar o resultado do teste ADF-GLS para o caso de amostras finitas.

Em Cheung e Lai (1995) é afirmado que embora a estatística de teste do ADF não dependa da ordem do lag assintoticamente, para amostras finitas, a estatística pode ser sensível ao número de lags. Então os autores abordam casos de amostras finitas com diferentes tamanhos para determinar a ordem do lag na aplicação do teste. O estudo indica que os valores críticos disponíveis até então são viesados por ignorar a dependência da ordem do lag selecionada. São realizados experimentos com T variando de 18 a 500 e k de 1 a 9, onde para T menor do que 25 é usado k menor do que 5 e são realizados 30000 experimentos, ou 40000 se T for maior do que 30. Com isso, são construídas superfícies de respostas através de uma equação para a estimação do valor crítico, de k e de T . Por fim, os autores medem a velocidade com que os valores de amostra finita conseguem se aproximar de seus valores assintóticos, concluindo que para os casos em que há a presença de constante e tendência a aproximação ocorre mais lentamente. O estudo mostra que a ordem do lag, além do tamanho da amostra, pode afetar o comportamento do teste ADF para amostras finitas.

Em Kiviet e Dufour (1997) são implementados testes de raiz unitária em modelos ARX de regressão dinâmica, os testes incluem a seleção do lag, presença de quebras estruturais, para séries com diferentes características. Após os testes, o estudo indica que os testes apresentam especificação adequada apenas assintoticamente e são não exatos para amostras finitas. É desenvolvido um processo alternativo de inferência para modelos lineares de regressão dinâmica e demonstrada sua viabilidade através de modelos empíricos, sendo denominados testes exatos. Por fim, os autores apresentam suas próprias estatísticas de testes obtidas através de simulações de Monte Carlo. Então concluem que os testes padrões podem apresentar p-valores menores do que os exatos, porém isso não implicaria em maior poder de teste. Se a distribuição dos erros não apresentar normalidade, a precisão dos modelos assintóticos é muito pobre para amostras

finitas, enquanto a precisão dos testes exatos se apresentou melhor, porém é mais sensível a vários tipos de distribuições.

O trabalho de Cheung e Lai (1998) examina o problema de como a seleção de defasagens usando critério de informação padrão pode ajudar a melhorar o poder de teste do ADF. São construídas curvas empíricas do poder de teste sob diferentes métodos de seleção de defasagens. O estudo reavalia o impacto de critérios de informação baseado nos dados no teste ADF pela correção para o viés para amostras finitas. Para isso, são simulados modelos AR e MA com k variando de 1 a 10, ϕ de 0,7 a 1 com passos de 0,025, e T assumindo os valores 50, 100 e 150, além da variação dos coeficientes de AR e MA. O número máximo de defasagens é calculado por $4(T/100)^{1/4}$ e são aplicados testes com e sem tendência. Como critério de seleção são utilizados AIC e BIC e são replicadas 40.000 vezes cada experimento. Os níveis de significância adotados são 5 e 10%. Para o modelo AR, são apresentados graficamente os resultados através do poder do teste em relação ao valor de ϕ para valores fixados diferentes de k e valores selecionados por AIC, com e sem tendência. Os mesmos resultados, nas mesmas condições são apresentados para o modelo MA. Ao final, concluem que para os casos em que as séries apresentam comportamento AR os critérios padrões são úteis, porém na presença de erros MA os resultados não são encorajadores, uma vez que a performance neste caso não se apresentou superior à seleção via um número fixo defasagens.

Os autores Montañes e Reyes (2000) examinam o comportamento assintótico dos testes ADF e PP na presença de quebras estruturais. Quando a mudança na tendência da série é omitida, as probabilidades de rejeição da hipótese nula são leves, porém essa característica só é observada para amostras finitas, não assintoticamente. Os autores consideram $T = 100$, são replicadas 10.000 vezes cada experimento para os testes ADF e PP considerando $k = 1, 2, 3, 4, 5, 6$ e 12. São abordados os casos com perturbação positiva, sem perturbação e com perturbação negativa. Em seguida é elaborada uma tabela com o poder do teste para cada condição experimental e cada valor de k . É mostrado que a magnitude da quebra afeta as duas estatísticas fazendo com que sejam deslocadas para a zona de aceitação, conforme aumenta a magnitude da quebra. Em relação ao número de defasagens, para o teste ADF, valores altos induzem o pesquisador a aceitar a presença de raiz unitária, mesmo para quebras de baixa magnitude. Já para o teste PP, o número de defasagens não mostrou influência significativa na aceitação da hipótese nula.

O trabalho de Otero e Smith (2000) estuda os efeitos de aumentar a frequência de observação e da extensão dos dados no teste de integração de Johansen. Para isso são criadas duas séries com 1188 observações mensais. São testados os valores de poder do teste para λ_{max} considerando que as séries sejam mensais, trimestrais ou anuais. Em seguida, os testes são aplicados em dados reais de variação de taxa de juros. Os autores concluem que é preferível ter dados coletados por um longo período de tempo do que ter um grande número de observações em um período pequeno.

O trabalho de Taylor (2000) encontrou que, para amostras finitas, o teste ADF é altamente sensível às duas formas de inclusão de variáveis determinísticas e à estrutura de defasagens

adotada. Em relação à seleção por meio de testes sequenciais, é demonstrado que este efeito aparentemente é atribuível ao viés sistemático de tendência a zero dos lags estimados, que é acentuado na presença de termos determinísticos. Para os experimentos, são geradas séries ARIMA(p,1,q), com 10.000 replicações e são considerados T como 50, 100 e 200, e valores de pares de k_{max} e k_{min} de (5,0), (10,0) e (10,4). As séries são geradas com e sem termos determinísticos, e são utilizados três critérios de seleção de defasagens: AIC, BIC e GSC, considerando o nível de significância de 5%. Os autores encontram evidências de que os modelos baseados nos dados com termos determinísticos tendem a subestimar o número de defasagens nos testes ADF, deste modo a rotina de inclusão de uma tendência linear na regressão do ADF não é aconselhada. Uma solução para este problema poderia ser utilizar métodos de seleção que tendem a selecionar menos lags. Outra hipótese seria aumentar o nível de significância para a tomada de decisão, como por exemplo para 0,20 e aplicar os testes sequenciais. Por fim, outra sugestão seria estimar o modelo sem os termos determinísticos e utilizar algum algoritmo baseado nos dados para estimar o valor de k, depois reestimar a regressão com os termos determinísticos.

O trabalho de Weber (2001) trabalha com a seleção de defasagens para os testes de raiz unitária das séries de Nelson-Plosser, no caso em que a hipótese alternativa inclui uma quebra de tendência. O autor diz que diferentes escolhas do número de defasagens podem afetar o poder do teste ADF de maneira significativa, para as séries abordadas, a presença de raiz unitária pode ser altamente sensível ao método de seleção de lags. São comparados cinco diferentes critérios de seleção de lags. O autor conclui que para as séries macroeconômicas estudadas, quando o critério geral para específico é utilizado, são encontradas menos evidências contra a hipótese da presença de raiz unitária em torno de uma quebra de tendência. Em vários casos a evidência contra a raiz unitária enfraquece consideravelmente, para salários reais e Produto Nacional Bruto a estatística ADF é mais insignificante ou somente significativa a um nível de 0.10. Quando a quebra é determinada endogenamente, como em Zivot e Andrews, os resultados obtidos pelo critério alternativo de seleção de lags fornecem ainda menos evidência contra a presença de raiz unitária para as séries de Nelson-Plosser. Por fim, o autor sugere que ao trabalhar com séries macroeconômicas os pesquisadores devem checar cuidadosamente o número de defasagens selecionado ao testar a presença de raiz unitária.

Em Aylar et al (2002) é mostrado que aumentar o número de lags ótimo selecionados para o teste ADF está associado à diminuição do poder do teste. Os autores buscam avaliar a distribuição local assintótica do teste ADF quando os erros seguem um processo linear dirigido por inovações de diferença de Martingale. Para isso, fazem simulações, considerando que os erros seguem um processo linear com inovações de diferença de Martingale, o que pode fazer com que apresentem heterocedasticidade. São apresentados fundamentos matemáticos das distribuições para um modelo AR. São utilizadas simulações com T sendo 50, 100, 200, 400, 800, 1600 e $k = T^a$, com a variando de 0,05 to 0,49 em passos de 0,04. Não é realizada a análise individual de cada caso. Quanto à presença dos termos determinísticos, são utilizadas técnicas para remover o efeito dos mesmos de modo que não altere o resultado do teste. É apresentado que a influência

do número de lags tende a diminuir conforme aumenta o número de elementos na série. Os autores afirmam que o teste ADF é o teste mais popular de raiz unitária e apresenta incontáveis aplicações. Porém, o baixo poder do teste é o mais conhecido dos problemas relacionados ao teste ADF.

O artigo de Vougas (2002) utiliza as séries de Nelson e Plosser (NP) para aplicação do teste DF. O autor afirma que em NP foi escolhido o número de defasagens de maneira incorreta e que havia presença de heterocedasticidade condicional. Justificando que isso se deve ao fato de que em NP o número de lags é selecionado através de comparação da autocorrelação da série em nível e em primeira diferença. Porém uma abordagem mais correta seria a utilização de critérios de informação, como por exemplo o de Akaike (AIC). Assim, são utilizadas as séries de NP considerando o AIC como método de seleção de defasagens, além disso, para aliviar o efeito de heterocedasticidade são construídas matrizes de covariância heterocedasticidade consistente-t. São construídas tabelas com os resultados e realizadas comparações entre eles. O autor encontrou resultados diferentes de NP em quatro das quatorze séries e conclui que a não rejeição da hipótese nula deve ser tratada com cuidado, pois pode estar relacionada a erro de especificação no teste

Em Lopez et al. (2005), é estudada a aplicação de testes de raiz unitária em séries de Paridade de Poder de Compra (PPC), dando enfoque na seleção ótima de defasagens no resultado dos testes. São utilizadas séries anuais de 16 países com tamanhos diferentes para a aplicação dos testes ADF e DF-GLS. São utilizados os critérios de seleção de defasagens de Akaike Modificado, testes-t sequenciais e LM. São apresentados os resultados dos testes e o número de defasagens selecionado para as condições experimentais com e sem tendência. Em seguida, é elaborada uma tabela com o número de rejeições para cada teste para fins de comparação. Por fim, os autores aplicam os testes escolhendo o número de defasagens que maximiza o valor p do teste de raiz unitária. O resultado do artigo apresentou maior taxa de rejeição da hipótese nula tanto para o ADF e quanto para o DF-GLS, se comparados ao estudo anterior. Os autores concluem que o número de defasagens pode afetar o resultado dos testes em séries de PPC.

No trabalho de Christopoulos (2006) são utilizados modelos não lineares para investigar a estacionariedade do PIB per capita dos sete países da OCDE. É apresentado o modelo autorregressivo de TAR que une as hipóteses de não linearidade e estacionariedade. Também é aplicado o teste ADF, utilizando critério de Schwert (1989) e testes-t sequenciais para o PIB de sete países. Para os mesmos países, membros da OCDE, são aplicados os testes TAR e estimado o modelo de Fourier. Em seguida são comparados os resultados dos testes ADF e TAR, concluindo que os testes não lineares se aplicam melhor às séries de PIB.

Em Burrridge e Hristova (2008) são estudados os estimadores de mínimos quadrados com a possibilidade de um processo não estacionário com no máximo uma raiz autorregressiva unitária. São realizados experimentos, onde o número de defasagens é escolhido pelo BIC e pelo AIC. Os tamanhos de amostra são de 100, 250, 500 e 1000. Para isso são utilizados modelos de ruído branco, bem como modelos AR, com o intuito de verificar se os critérios selecionam a ordem certa dos modelos. Os autores concluíram que uma escolha segura para aplicação dos testes seria

realizar a seleção de defasagens proporcional ao valor obtido pela equação $C(T) = \ln(T)/T$, porém não deve ser adotada essa regra de maneira acriticamente.

O autor Wu (2010) aborda a temática da seleção de lags no teste DF-GLS, que é uma modificação do ADF onde os dados têm sua tendência removida. O artigo reinvestiga o problema de seleção de lags no teste citado, mostrando a importância de utilizar valores críticos ajustados para o tamanho da amostra na seleção de lags para a aplicação do teste. Para isso, são estimados os modelos de DF-GLS e é utilizado k_{max} como $\text{int}[12(T/100)^{1/4}]$. Os níveis de significância trabalhados são 5 e 10%. Assim, são testados os critérios de informação modificados de Akaike estimados por OLS e GLS, bem como testes sequenciais t. Foram selecionadas as defasagens através de cada teste e cada método para taxa de câmbio de 20 países. Assim, é mostrado o número de defasagens selecionado bem como o resultado de cada teste. Em seguida são gerados um modelo MA e um AR (com $T = 100, 150$ e 250) e é avaliado o tamanho do teste em cada caso. Em seguida é gerado um novo modelo, com mais semelhanças às séries de taxas de juros a fim de avaliar o tamanho do teste nesta situação para cada um dos testes. No caso em que os dados imitam a estrutura de séries de câmbio, a seleção de lags através de testes sequenciais t, para o teste ADF-GLS considerando o nível de significância de 10% apresentou o melhor resultado. Por outro lado, quando a série apresenta grandes coeficientes negativos MA, como por exemplo, -0,5 ou -0,8, o critério de seleção por critério de informação apresentou-se vantajoso. Pelo fato de nenhum dos critérios ter sido superior para diferentes tipos de dados, o autor sugere a utilização deles em conjunto.

O artigo de Fossati (2012) analisa a seleção de lags para testes de raiz unitária covariáveis. Para isso, são gerados pares de séries com diferentes características para cada par. Em seguida, são aplicados os testes EJ e CADF-GLS. Para cada um dos testes são utilizados diferentes critérios de seleção de defasagens e os resultados comparados entre si. São apresentados os resultados para $T = 100$. Em seguida, é utilizada a série de inflação dos EUA, utilizando como covariáveis o PIB e a taxa de desemprego. O autor mostra que os critérios de informação de AIC e BIC selecionam lags muito pequenos que resultam em testes com grandes distorções. Um critério de informação modificado pode ser utilizado para construir um teste de raiz unitária covariável com bom tamanho e poder. São aplicados os critérios Modificado e BIC em um conjunto de dados que contém a inflação nos EUA combinada com outras variáveis econômicas, onde segundo o critério Modificado, a hipótese nula não é rejeitada, encontrando resposta diferente do BIC.

No artigo de Deng (2012) são simulados modelos ARIMA(0,1,q) e comparada a influência de diferentes métodos de seleção de defasagens e o poder do teste ADF. São geradas séries ARIMA(0,1,q) e aplicado o teste ADF utilizando os critérios AIC, BIC, MAIC, MBIC e GSC. Em seguida são simulados diferentes valores de θ , com T e k_{max} fixados em 100 e 20, respectivamente. São realizadas 10.000 simulações para cada caso, e os resultados apresentam o lag médio selecionado por cada método, bem como o percentual de rejeição da hipótese nula em cada um. O artigo mostra que os critérios de GSC, AIC e MAIC tendem a selecionar mais lags do que os SIC e MBIC. Além disso, o número de lags selecionado aumenta quando o valor da

raiz tende à raiz unitária. Ao final conclui que o critério de informação modificado apresentou melhor tamanho do teste, enquanto o critério geral para específico mostrou mais robustez nas propriedades do teste ADF.

O trabalho de Caner e Kight (2013) apresenta um modo alternativo de diferenciar a raiz unitária utilizando os chamados estimadores Bridge. Inicialmente são apresentadas as equações, bem como a representação matemática dos estimadores. A seleção de parâmetros se dá por meio do MBIC. Em seguida, são comparados os resultados dos testes DFGLS e ADF com o Bridge, para casos com e sem tendência. Não existe grande ênfase na seleção de defasagens, entretanto são comparados os valores críticos considerando k_{max} como 24 e 36, com $T=100$. É mostrado que os estimadores Bridge podem selecionar o modelo correto com probabilidade tendendo a um. Por fim, o trabalho utiliza o estimador Bridge para determinar os parâmetros que são zero e diferentes zero, assim o modelo diferencia a presença ou não de raiz unitária e seleciona o lag ótimo simultaneamente.

O autor Tam (2013) apresenta uma análise de superfície de resposta para o teste ADF com otimização da seleção da ordem do lag. As simulações mostram que para amostras finitas o resultado do teste depende diretamente da regra de seleção das defasagens. O teste apresenta distorção de tamanho quando a ordem do lag é otimizada, mas uma distribuição inapropriada é empregada na implementação do teste. São propostas superfícies resposta para distribuições de amostras finitas com otimização da seleção de lags. São realizados dez experimentos para cada cruzamento de T com k_{max} para a realização dos experimentos. Assim, são geradas séries com tamanho $T+100$ e são descartadas as primeiras 100 observações a fim de descartar o efeito do valor inicial. O autor conclui que o comportamento do teste ADF para amostras finitas é sensível ao método de seleção de lags aplicado.

Em Cavaliere et al. (2015) é apontado que grande parte dos trabalhos sobre o tema determinam o tamanho do lag em relação ao tamanho da série, porém poderia ser mais eficiente se fosse em relação às características dos dados. Os autores examinam o impacto de heteroscedasticidade incondicional, para amostras finitas, nos métodos que dependem dos dados para a seleção de lags. É proposto um novo método de seleção de lags que pode ser utilizado para heteroscedasticidade incondicional, deste modo, implementado um algoritmo wild-bootstrap. Os modelos testados são com $\phi = -c/T$, com $c = 0$ e $c = 7$. São realizadas simulações com diversos modelos variando os parâmetros de AR e MA, com erros homocedásticos e heterocedásticos, séries com tamanhos de 150 e 250. Por fim, é apresentado graficamente o poder do teste em função dos lags selecionados. Para os casos em que existe tendência, esta é removida através de algoritmo *wild bootstrap*. É observado que os métodos padrões de seleção de lags tendem a superestimar a ordem do lag em caso de heterocedasticidade, resultando em uma significativa perda do poder de teste sob a hipótese alternativa. Enquanto o novo método de seleção de lags se mostrou capaz de evitar este problema, entregando testes com propriedades para amostras finitas quase idênticas aos correspondentes baseados nos métodos tradicionais, na presença de choques homocedásticos. A modificação proposta consiste em redimensionar os dados através de um estimador paramétrico do processo de volatilidade antes de computar a informação do

critério de interesse. Foi descoberto que a volatilidade não estacionária pode ter um impacto significativo na seleção de lags em amostras finitas. Os autores encontraram indícios de que, para diversos modelos, a seleção de lags tem sido altamente dependente do lag máximo permitido em certos casos. Enquanto isso, o critério proposto pelos autores não apresentou essa característica, porém na ausência de heterocedasticidade, apresentou resultados semelhantes aos testes tradicionais.

O trabalho de Arltová e Fedorová (2016) tem por objetivo resolver o dilema de aconselhar qual teste de raiz unitária utilizar de acordo com critérios específicos. Para isso são apresentados os testes ADF, PP, ADF-GLS, NGP e KPSS. São geradas séries AR(1), sem termos determinísticos, com o ϕ variando de 0,01 a 0,99, com T sendo 25, 50, e 100 e realizadas 3.000 replicações com nível de significância de 5%. Em seguida, são aplicados os testes de raiz unitária e comparados através do poder do teste. Também são realizadas simulações com $\phi = 1$ para efeito de complementação. Para cada teste é mostrado o poder do teste em função do valor do parâmetro ϕ para as séries dos diferentes tamanhos. A seguir é feita a comparação entre os testes para cada comprimento de série. Também é mostrada a estimação de erro do tipo II para cada situação do caso em que a série é não estacionária. Os autores formam uma tabela indicando a aplicação dos testes de acordo com os valores de ϕ e com o comprimento da série. Ao fim, concluem que o trabalho fornece recomendações para a aplicação dos testes de raiz unitária de acordo com cada condição exposta.

O artigo de Cavaliere et al. (2016) realiza os testes ADF para o caso em que os choques seguem um processo linear conduzido por inovação de variância infinita. Para isso, é demonstrado matematicamente o que ocorre com a distribuição assintótica do teste para o caso em que apresenta variância infinita. Então é comparada a aplicação de Estimadores M com MQO para a estimação do modelo. Logo, é aplicada uma técnica de reamostragem, através de um algoritmo Wild Bootstrap. O modelo simulado é um MA, e a determinação do número de lags se dá através do critério de Schwert (1989). Os modelos testados são com $\phi = -c/T$, com $c = 7$ e $\phi = 0$. Para a determinação do lag ótimo é utilizada a fórmula $T^{1/3}$. Também estabelecem uma medida de consistência e distribuição assintótica do MQO que estima a regressão do ADF. Por fim, demonstram a validade do algoritmo wild bootstrap para testes ADF, de modo que as estatísticas do wild bootstrap ADF alcançam os mesmos limites de distribuição que o teste ADF original, condicionadas à magnitude das inovações.

O trabalho proposto por Paparoditis e Politis (2016) propõe uma nova maneira de determinar o número de lags a ser considerado no teste. Após realização de equacionamento matemático, é utilizada uma série real da variação de temperatura global anual. Para as simulações, é utilizada a função *adf.test* do *software R*. Em seguida, é montado um experimento utilizando um modelo ARMA. É proposta uma fórmula para a determinação do número de lags: $k = n^a$, sendo testados diferentes valores de a , variando de 0,05 a 0,49 e com séries de tamanhos entre 50 e 1600. Ao fim, concluem que o valor mais apropriado para a é 0,3. Em seguida apresentam prova matemática para sua conclusão. Os autores comparam sua proposta com o AIC, chegando a conclusão que quando existe raiz unitária essa fórmula pode ser melhor,

porém quando não há, é melhor usar o AIC. Porém, ressaltam que o AIC apresenta o problema de rejeitar a hipótese nula, mesmo quando verdadeira. Os autores observam que a razão intuitiva para o reduzido poder de teste do ADF é que como p tende ao infinito, os regressores de p se tornam assintoticamente colineares. A potência do teste é uma função decrescente da ordem autoregressiva p escolhida.

O trabalho de Quineche e Rodríguez (2017) apresenta um estudo baseado na simulação do tamanho e no comportamento excêntrico que pode ser apresentado pelo teste MGLS no contexto de mudanças estruturais. São gerados modelos AR(1) e MA(1) com séries de tamanho 100 e 250 e considerando a presença de raiz unitária. São apresentados, para cada tamanho de série, os resultados do poder do teste para nove critérios diferentes de seleção de defasagens, além de seis valores fixos de k . Os resultados mostraram uma grande relação entre valores explosivos negativos de MGLS e o valor da defasagem selecionado. Quando há quebra na série, os autores indicam a utilização do método supremo, selecionando as defasagens por MIC.

Os autores Özel e Ilalan (2018) propõem uma metodologia capaz de identificar a estacionariedade em séries de taxas de juros. Os autores convertem um processo OU em uma estrutura de regressão linear para calcular as estatísticas t pertencentes ao teste de reversão à média, e estes valores são comparados com os valores críticos obtidos por simulações de Monte Carlo. As séries estudadas são de taxas de juros diária correspondente a dez anos da Turquia e dos Estados Unidos. São aplicados os testes OU, ADF e PP e são comparados os resultados, inclusive com a presença de tendência. Em alguns casos, o teste OU indicou estacionariedade nas séries, ao contrário do ADF e do PP. Houveram outros casos em que todos os testes indicaram estacionariedade, nestes o teste OU apresentou maiores valores- p . Entretanto, os autores indicam o teste apenas para os casos em que for necessário detectar reversão de média em séries de taxa de juros, e deve ser aplicado juntamente com o ADF para a detecção de reversão à média.

O trabalho de Ferrer-Pérez et al. (2019) apresenta o efeito da observação inicial da série econômica nos testes de raiz unitária, buscando examinar o efeito da observação inicial da série no resultado de testes de raiz unitária. Os autores afirmam que nenhum estudo abordou a seleção de lags para testes de raiz unitária nessas configurações. Para modelagem, é considerado um modelo ARMA onde os parâmetros são variados, e os valores do coeficiente ϕ são (1; 0,95; 0,9; 0,8), onde 1 caracteriza a presença de raiz unitária. É assumida a presença de um outlier na condição inicial e são testadas séries com os erros considerados em três situações: ruído branco, processo autoregressivo e processo de médias móveis. Os autores apresentam os resultados para tendência determinística, com $\phi = 1$ e 0,9 e afirmam que os resultados em demais casos são similares. São realizadas simulações com $T=100$ e $T=250$. É avaliado o poder do teste em cada condição experimental, na aplicação de seis testes diferentes, com cinco métodos de seleção de lags distintos. O trabalho tem dois objetivos principais: i. identificar como a condição inicial afeta a seleção de lags; e ii. qual a implicação na performance do poder do teste. O lag máximo é determinado por $k = \text{int}[12(T/100)^{1/4}]$, e a seleção dos lags é feita pelos métodos propostos pelo artigo: DF_{OLS} , DF_{PQ} , EM , HL_{05} , HL_{06} , UR_{SA} que são baseados nos critérios de MAIC,

MBIC, BIC, AIC e HQC. Ao fim, encontram que nenhum dos métodos se apresenta superior para todas as circunstâncias, porém recomendam o uso de HL_{05} e UR_{SA} , deixando claro que o assunto ainda precisa de mais pesquisas. Para trabalhos futuros, os autores afirmam que o uso de critérios tradicionais de seleção de lags podem causar distorções de tamanho e que essa relação de distorção de tamanho e poder de teste ainda deve ser mais pesquisada.

O artigo de Zuo (2019) compara as características das equações dos testes DF, ADF, PP e KPSS através do conteúdo de vinte materiais. O autor apresenta equações, características de testes e aborda as hipóteses consideradas. Deste modo, o autor busca analisar as equações, aplicações e distribuições dos testes através de levantamento bibliográfico.

A metassíntese mostrou que o tema de seleção de defasagens na aplicação dos testes de raiz unitária vêm sendo tema de muitos estudos nos últimos anos. Grande parte destes estudos utiliza o teste ADF para identificar a presença de raiz unitária, evidenciando sua popularidade. Entretanto, diversos trabalhos propõe novos testes e estes geralmente não são capazes de superar o ADF nos resultados, salvo em algumas situações em específico. É possível observar frequentes relatos de baixo poder do teste ADF, bem como uma possível sensibilidade do resultado do teste tanto com as defasagens selecionadas, quanto com a inclusão de termos determinísticos.

3 METODOLOGIA

3.1 Simulações preliminares

Para a realização um estudo preliminar sobre a influência do número de defasagens selecionado para a aplicação do teste ADF no resultado deste, foi proposto um experimento de um modelo AR(1), com e sem a presença dos termos determinísticos. Para o valor do parâmetro ϕ , foram considerados os valores de 0,55, 0,65, 0,75, 0,85, 0,95 e 1,0. Quando presentes, os valores de β_1 e β_2 foram, respectivamente, 10 e 1. Além disso, u_t foi considerado um ruído branco gaussiano, isto é $u_t \sim N(0, 1)$ e as séries tiveram tamanhos $T = 30, 50$ e 100 .

Para a aplicação do teste, o valor de k_{max} foi fixado a partir valor calculado na Eq. 2.13 e foram utilizados modelos com defasagens pertencentes ao intervalo $[0, k_{max} + 2]$. Diante disso, foram realizados 1000 experimentos para cada combinação de parâmetros e foi tomada a proporção de rejeição da hipótese nula ($\hat{p}$) para cada defasagem. Em seguida foram gerados gráficos de linha para análise visual do comportamento de cada resultado do teste em relação ao número de defasagens.

Para efeito de análise comparativa, foram tomados os ensaios com ϕ igual a 0,55, 0,75, 0,95 e 1,0, considerando a inclusão ou não de termos determinísticos da seguinte maneira: modelo com constante e tendência (CT), modelo somente com constante (C) e modelo sem constante nem tendência (N).

Segundo Zar (1999), as proporções entre zero e um podem se afastar muito de uma distribuição normal, assim foi realizada a transformação de $\hat{p}$ em $\arcsen\sqrt{\hat{p}}$ de modo que a variância seja estabilizada e os resultados tendam a apresentar uma distribuição próxima à normal. Para analisar o efeito de cada fator na proporção, foi realizada uma análise de variância, através da estatística qui-quadrado. Além disso, a transformação matemática estabiliza a variância do erro em $1/(4Ns)$, sendo Ns a quantidade de simulações. A significância do resultado pode ser avaliada por meio da estatística de teste $\chi^2 = \frac{SQ_{\text{fonte de variação}}}{\sigma_{\text{erro}}^2}$, que apresenta distribuição qui-quadrado com $gl_{\text{fonte de variação}}$ graus de liberdade.

Em um primeiro cenário foi considerado $\phi = 1$, o que indica a presença de raiz unitária na série. No segundo cenário, são abordadas as séries com $\phi < 1$ (isto é, 0,55, 0,75 e 0,95), casos em que as séries são estacionárias.

Os procedimentos foram realizados com o auxílio do *software R* (R Core Team, 2020), com a utilização dos pacotes *forecast* (Hyndman et al., 2008), *urca* (Pfaff, 2008), *lmtest* (Zeileis e Hothorn, 2002) e *agricolae* (De Mendiburu, 2020).

3.2 Simulações

3.2.1 Planejamento dos experimentos

Inicialmente foi realizada uma análise bibliométrica onde foi encontrado o estado da arte da aplicação dos testes, bem como foram identificadas lacunas de pesquisa. Em seguida, foi proposta uma estratégia para selecionar o número de defasagens capaz de obter a maior taxa de acerto na aplicação do teste ADF. Para isso foram construídos experimentos através do processo de geração de dados de um modelo AR(1): $y_t = \beta_1 + \beta_2 t + \phi y_{t-1} + \epsilon_t$. Assim, foram consideradas as condições experimentais com $\beta_1 = \beta_2 = 0$, que corresponde ao modelo sem constante e sem tendência, a condição com apenas $\beta_2 = 0$, correspondendo ao modelo com apenas constante, além da situação da equação completa representando modelo com constante e com tendência. As séries foram geradas com diferentes comprimentos: 25, 50 e 100. Em relação ao valor de ϕ , foram utilizados os valores de 0,95, 0,75 e 0,55 onde foi abordada a condição das séries estacionárias. Além disso, também foi considerado o caso em que existe a presença de raiz unitária, isto é $\phi = 1$. Após a geração dos dados, foram executadas 10.000 replicações para cada condição experimental, onde foi avaliada a capacidade do teste de identificar a presença de raiz unitária. Foram aplicados testes considerando os níveis de significância de 5 e 10%, assim como três critérios diferentes de determinação da defasagem máxima (k_{max}), dois propostos por Schwert (1989) e outro por Hyndman e Athanasopoulos (2018). Além disso, foram aplicados três critérios diferentes de seleção da defasagem a ser aplicada no teste: AIC, BIC e HQC. Assim, com a construção dos experimentos, contando com a alteração de diversos fatores, foram gerados cálculos e gráficos de percentual de rejeição da hipótese nula quando ela era falsa, além de percentual de não rejeição da hipótese nula para o caso em que ela era verdadeira. Deste modo, através das 648 séries diferentes geradas, buscou-se encontrar uma metodologia a ser indicada na aplicação do teste ADF de acordo com a característica da série a ser testada.

As tabelas 3.1 e 3.2 apresentam o planejamento dos experimentos realizados, identificando os diferentes fatores e respectivos níveis, onde foram realizados experimentos com séries temporais de comprimentos iguais a 25, 50 e 100.

Tabela 3.1: Fatores e respectivos níveis considerados nos experimentos realizados com a presença de raiz unitária

Fator	Atribuições
Estrutura do modelo	$\beta_1 = \beta_2 = 0$, $\beta_1 \neq 0$ e $\beta_2 = 0$ $\beta_1 \neq 0$ e $\beta_2 \neq 0$
Coefficiente ϕ	1,0
Nível de significância	5% e 10%
Lag máximo (k_{max})	Equações 2.12, 2.13 e 2.14
Lag ótimo (k)	AIC, BIC e HQC
Interpretação	Proporção de Rejeição de H_0

Tabela 3.2: Fatores e respectivos níveis considerados nos experimentos realizados com a ausência de raiz unitária

Fator	Atribuições
Estrutura do modelo	$\beta_1 = \beta_2 = 0$, $\beta_1 \neq 0$ e $\beta_2 = 0$ $\beta_1 \neq 0$ e $\beta_2 \neq 0$
Coefficiente ϕ	0,55, 0,75 e 0,95
Nível de significância	5% e 10%
Lag máximo (k_{max})	Equações 2.12, 2.13 e 2.14
Lag ótimo (k)	AIC, BIC e HQC
Interpretação	Proporção de Rejeição de H_0

3.2.2 Análise dos experimentos

A fim de verificar a influência do efeito principal de cada um dos fatores considerados no experimento no resultado do teste ADF, bem como a interação destes dois a dois, foram aplicadas algumas técnicas conforme descrito a seguir.

Uma das maneiras mais utilizadas para realizar essa tarefa é por meio de uma Análise de Variância (ANOVA) com a utilização do teste F. Entretanto, esta abordagem pressupõe que as observações sejam independentes, homoscedásticas e apresentem distribuição normal. Devido às características dos experimentos construídos, estes pressupostos não puderam ser satisfeitos em todos os casos. Mesmo com a realização de transformações nos dados, não foi possível encontrar resultados satisfatórios.

Em seguida, como alternativa, foi aplicada uma análise de variância com a estatística qui-quadrado, após a estabilização da variância por meio de uma transformação matemática, onde o resultado encontrado foi aceitável. Entretanto, houve dificuldade em encontrar um teste para realizar as comparações múltiplas para fatores que apresentassem mais de dois níveis.

Como alternativa, optou-se pelos Modelos Lineares Generalizados (GLM). Neste caso não foi encontrada uma distribuição que se adequasse suficientemente bem às características dos dados. Diante disso, optou-se por trabalhar com a modelagem GAMLSS, apresentada na Seção 2.3.

Inicialmente foi feita uma reordenação dos fatores que possuem três níveis em função de suas medianas, de modo que o nível intermediário fosse sempre tomado como referência. Quanto às funções de ligação, estas foram selecionadas de modo a respeitar o espaço paramétrico. Ou seja, neste trabalho, foi utilizada a função logit para o parâmetro μ , uma vez que $0 < \mu < 1$, e a função logarítmica para $\sigma > 0$. Por questão de parcimônia, foi realizada a modelagem apenas com o primeiro parâmetro (μ) e testado o ajuste do modelo, caso houvesse a necessidade de melhoria, seria realizada a modelagem incluindo o parâmetro σ e feita nova verificação do ajuste. A seleção das covariáveis em um GAMLSS normalmente é realizada por meio de um procedimento baseado no stepwise. Um detalhamento deste procedimento pode ser encontrado em Ramires et al. (2021). Neste trabalho, todas as covariáveis foram mantidas, uma vez que o objetivo é verificar quais os fatores que influenciam no resultado do teste. Por fim, foram

gerados modelos com diferentes famílias de distribuições (GB1, LOGITNO, BE, BEo, BEINF0) e foi feita a comparação dos ajustes de cada um através do *worm plot*.

Os procedimentos foram realizados com o auxílio do *software R* (R Core Team, 2020), com a utilização dos pacotes *gamlss*, *gamlss.dist*.

3.3 Simulações Complementares

Uma vez que nas simulações preliminares os resultados se mostraram diferentes para as séries com $T=30$ e $T=50$ e que na análise bibliométrica foram encontrados indícios de problemas no teste ADF para séries de tamanhos pequenos, considerou-se ser interessante a realização de experimento complementar, considerando séries temporais com comprimento intermediário entre 25 e 50. Diante disso, foram realizadas simulações complementares de experimentos com fatores e respectivos níveis conforme o descrito na Tabelas 3.1 e na Tabela 3.2, considerando $T=40$. A análise de seus resultados foi realizada por meio da modelagem GAMLSS.

4 RESULTADOS E DISCUSSÕES

4.1 Análise dos resultados do estudo preliminar

Para o cenário em que $\phi = 1$ é esperado que o teste identifique a presença da raiz unitária e a proporção de rejeição da hipótese nula seja inferior a 5%, uma vez que 0,05 foi o nível de significância adotado para os experimentos. A Tabela 4.1 apresenta as medidas de resumo para as séries geradas e é possível notar que para cada estrutura, o valor-p tende a diminuir com o aumento da amplitude da série, com exceção da série sem constante nem tendência, com $T=100$.

Tabela 4.1: Média (e desvio padrão) da proporção de casos em que a hipótese nula é rejeitada, no primeiro cenário com simulações considerando $\phi = 1$

T	Constante	Constante e tendência	Sem constante e sem tendência
30	0,04011 (0,00411)	0,04091 (0,00630)	0,05555 (0,00446)
50	0,03085 (0,00428)	0,00862 (0,00409)	0,04723 (0,00588)
100	0,01506 (0,00282)	0,00435 (0,00154)	0,05918 (0,00484)

Já na Figura 4.1 são mostradas as proporções de casos em que H_0 é rejeitada, considerando as séries com diferentes estruturas e comprimentos. É possível observar que a proporção de casos em que há a rejeição da hipótese nula parece diminuir com o aumento da amplitude da série, enquanto em relação ao número de defasagens parece ocorrer certa estabilidade. Em relação à estrutura da série, nota-se que as séries sem constante e sem tendência apresentam maior taxa de rejeição.

A Tabela 4.2 apresenta o resultado da análise de variância, onde não foram encontrados indícios de que a quantidade de defasagens interfere na proporção de casos de rejeição da hipótese nula para as séries de comprimento 30 e 50, o mesmo não ocorrendo para $T=100$ (embora o valor-p tenha ficado próximo a 0,05). Também foram encontradas evidências de que a estrutura do modelo interfere na proporção de casos em que a hipótese nula é rejeitada, sendo o caso de modelo sem constante e sem tendência o que apresenta proporções mais elevadas.

Já para o cenário em que $\phi < 1$, indicando que as séries são estacionárias. Neste contexto, espera-se encontrar a proporção de rejeições elevadas. A Tabela 4.3 apresenta o resumo das proporções encontradas para cada valor de ϕ , considerando as diferentes amplitudes e características, onde é possível notar que, para cada estrutura considerada, a proporção de rejeições de H_0 tende a aumentar com o aumento da amplitude da série, com a diminuição do valor do coeficiente, e com a não inclusão de termos determinísticos.

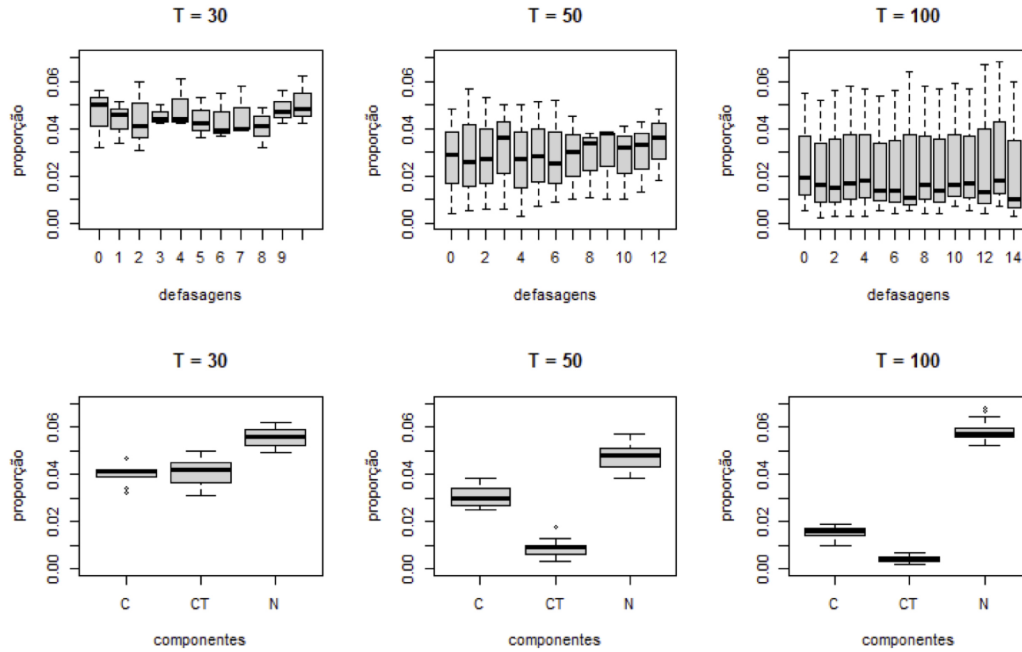


Figura 4.1: Gráficos de caixa da proporção de casos de rejeição do hipótese nula, de acordo com a quantidade de defasagens e componentes do modelo no primeiro cenário com simulações considerando $\phi = 1$

Tabela 4.2: Resultados da análise de variância no primeiro cenário com simulações considerando $\phi = 1$

T	Fonte de variação	gl	SQ	X^2_{calc}	valor- p
30	Defasagens	10	0,00146	5,84	0,0817
	Componentes	2	0,00927	37,08	<0,0001
50	Defasagens	12	0,00235	9,40	0,0869
	Componentes	2	0,11075	443,00	<0,0001
100	Defasagens	14	0,00192	7,68	0,0478
	Componentes	2	0,25189	1.007,56	<0,0001

Os gráficos da Figura 4.2 apresentam a distribuição dos resultados referentes à rejeição da hipótese nula nas diferentes séries testadas. Nos gráficos é possível notar que essa proporção tende a diminuir com o aumento do número de defasagens, além disso o modelo sem constante e sem tendência apresenta maior proporção de rejeição da hipótese nula. Ainda, conforme aumenta o valor do coeficiente ϕ , a proporção de rejeição da hipótese nula também aumenta. Por fim, conforme aumenta a amplitude da série, mais evidente se tornam as conclusões anteriores.

Os resultados da análise de variância podem ser encontrados na Tabela 4.4, onde é possível notar que a quantidade de defasagens, a estrutura do modelo e a intensidade do coeficiente ϕ interferem na proporção de rejeição de H_0 , o que está de acordo com a análise gráfica.

Em relação a interação entre o número de componentes e de defasagens, a Figura 4.3 (a)-(c)

Tabela 4.3: Média (e desvio padrão) da proporção de casos em que a hipótese nula é rejeitada, no segundo cenário com simulações considerando $\phi < 1$

T	ϕ	Constante	Constante e tendência	Sem constante e sem tendência	Total
30	0,95	0,05736 (0,00642)	0,04554 (0,00706)	0,28318 (0,10915)	0,12870 (0,12681)
	0,75	0,27339 (0,27055)	0,07782 (0,04468)	0,53309 (0,33477)	0,29509 (0,30720)
	0,55	0,33618 (0,35251)	0,21045 (0,25953)	0,54191 (0,35732)	0,36285 (0,35401)
	Total	0,22248 (0,27648)	0,11127 (0,16413)	0,45273 (0,30574)	0,26216 (0,29104)
50	0,95	0,06431 (0,01212)	0,08477 (0,02396)	0,35623 (0,12262)	0,16844 (0,15214)
	0,75	0,37669 (0,32106)	0,26400 (0,23523)	0,67008 (0,28933)	0,4369 (0,32639)
	0,55	0,43977 (0,36722)	0,41300 (0,35827)	0,70185 (0,29068)	0,51821 (0,35669)
	Total	0,29359 (0,32068)	0,25392 (0,27689)	0,57605 (0,28781)	0,37452 (0,32663)
100	0,95	0,10607 (0,02086)	0,21727 (0,06903)	0,54647 (0,11764)	0,28993 (0,20813)
	0,75	0,63773 (0,28566)	0,56573 (0,27417)	0,89433 (0,13363)	0,69927 (0,28273)
	0,55	0,71113 (0,28211)	0,69440 (0,28655)	0,92780 (0,10584)	0,77778 (0,25899)
	Total	0,58753 (0,31349)	0,56950 (0,28203)	0,88103 (0,12968)	0,59694 (0,32952)

mostra que para modelos sem constante e sem tendência a proporção de rejeição da hipótese nula apresenta valores mais elevados. Entretanto, no que se refere à interação entre o número de defasagens e a magnitude do parâmetro ϕ , os dados mostrados na Figura 4.3 (d)-(f) mostram que valores mais elevados da proporção de rejeições da hipótese nula são encontrados para coeficientes de maior intensidade. Por outro lado, como a diminuição dos valores é mais acentuada para valores menores do coeficiente, estas proporções se igualam a partir da oitava defasagem para as séries de amplitude 30, e a partir da 12^a para séries de amplitude 50, ou seja, $k_{max} + 2$, e o mesmo ocorre para a série de 100 elementos.

Os resultados obtidos mostraram que, se a série temporal possui raiz unitária, a quantidade de defasagens pouco interfere no resultado, independentemente da estrutura do modelo. Entretanto, se a série temporal não possui raiz unitária, a quantidade de defasagens interfere no resultado desta proporção, e consequentemente no valor-p, diminuindo com o aumento do

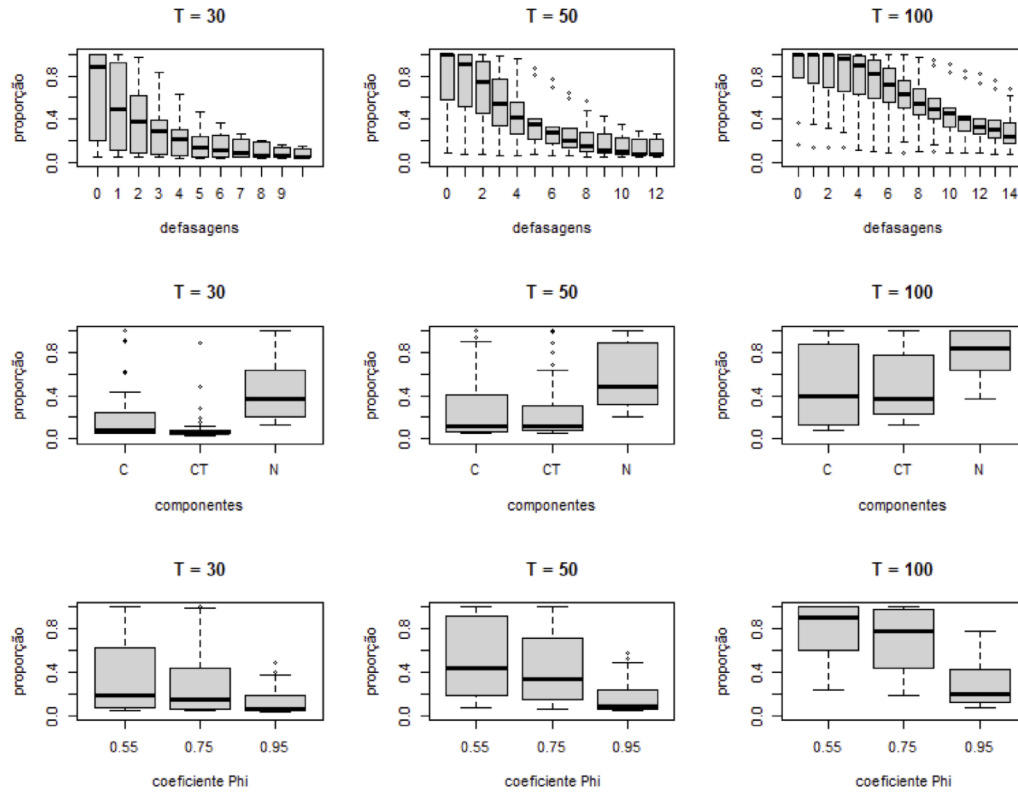


Figura 4.2: Gráficos de caixa da proporção de casos de rejeição do hipótese nula, de acordo com a quantidade de defasagens e componentes do modelo no segundo cenário com simulações considerando $\phi < 1$

número de defasagens, o que indica uma diminuição do poder do teste. Menores proporções são encontradas em modelos com constante e tendência, assim como também em modelos com coeficiente ϕ próximo de 1,0.

Não obstante, os resultados obtidos ao relação ao parâmetro ϕ estão de acordo com os encontrados em Arltová e Fedorová (2016), em que, para séries com 50 observações, o poder do teste é muito baixo para $\phi > 0,5$ enquanto que para séries de 100 observações o poder do teste decresce rapidamente quando $\phi > 0,8$. Além disso, como pode-se observar, para pequenas amostras, o poder do teste decresce com o aumento do número de defasagens selecionado, corroborando assim com Aylar et al (2002) e Paparoditis e Politis (2016).

Em relação aos termos determinísticos, em Taylor (2000), para o caso do teste em que há a inclusão de tendência linear os resultados indicam que há um aumento na probabilidade de rejeitar a hipótese nula quando esta é verdadeira, associando isso ao fato de que a inclusão da tendência leva a subestimar o número de defasagens a serem incluídas na regressão do teste ADF. Estes achados não estão de acordo com o que foi encontrado neste trabalho, uma vez que aqui os indícios mostram que a inclusão de termos determinísticos reduz ligeiramente a proporção de rejeição da hipótese nula na presença da raiz unitária. Entretanto, cabe salientar que os dados estudados em Taylor (2000) seguiam um processo ARIMA, diferentemente dos dados aqui simulados.

Tabela 4.4: Resultados da análise de variância no primeiro cenário com simulações considerando $\phi = 1$

T	Fonte de variação	gl	SQ	X^2_{calc}	valor- p
30	Defasagens	10	5,068	20,272	<0,0001
	Componente	2	3,549	14,196	<0,0001
	Coeficiente	2	1,640	6,560	<0,0001
	Componente $\times$ Coeficiente	4	0,212	848	<0,0001
	Defasagens $\times$ Componente	20	0,860	3,440	<0,0001
	Defasagens $\times$ Coeficiente	20	1,912	7,648	<0,0001
50	Defasagens	12	8,585	34,340	<0,0001
	Componente	2	3,874	15,496	<0,0001
	Coeficiente	2	4,328	17,312	<0,0001
	Componente $\times$ Coeficiente	4	0,157	628	<0,0001
	Defasagens $\times$ Componente	20	0,258	1,032	<0,0001
	Defasagens $\times$ Coeficiente	20	2,767	11,068	<0,0001
100	Defasagens	14	7,546	30,184	<0,0001
	Componente	2	4,172	16,688	<0,0001
	Coeficiente	2	10,179	40,716	<0,0001
	Componente $\times$ Coeficiente	4	0,328	1,312	<0,0001
	Defasagens $\times$ Componente	20	0,298	1,192	<0,0001
	Defasagens $\times$ Coeficiente	20	1,649	6,596	<0,0001

4.2 Análise dos resultados das simulações

Para efeito de análise, as séries foram divididas em não estacionárias e estacionárias, e em experimentos complementares. No primeiro caso foi observada a proporção de não rejeição da hipótese nula, ao passo que no segundo, foi tomada a proporção de rejeição da hipótese nula, pois estes são os resultados esperados em cada caso. Ainda, em cada tipo de série, os resultados dos experimentos foram analisados separadamente para cada uma das diferentes amplitudes consideradas para as séries temporais. Cabe ressaltar que os critérios de seleção de defasagem máxima, l_{12} e l_h , para os valores T deste trabalho, selecionaram o mesmo número de k_{max} , deste modo serão apresentados os resultados para l_{12} , que podem ser estendidos ao l_h . Para analisar a influência de cada fator e de seus níveis no resultado do teste, foi realizada a modelagem GAMLSS.

4.2.1 Séries Não Estacionárias($\phi = 1$)

Neste caso, como já relatado, espera-se que o teste não rejeite a hipótese nula da presença de raiz unitária. Deste modo, é tomada a taxa de não rejeição de H_0 para a realização das análises.

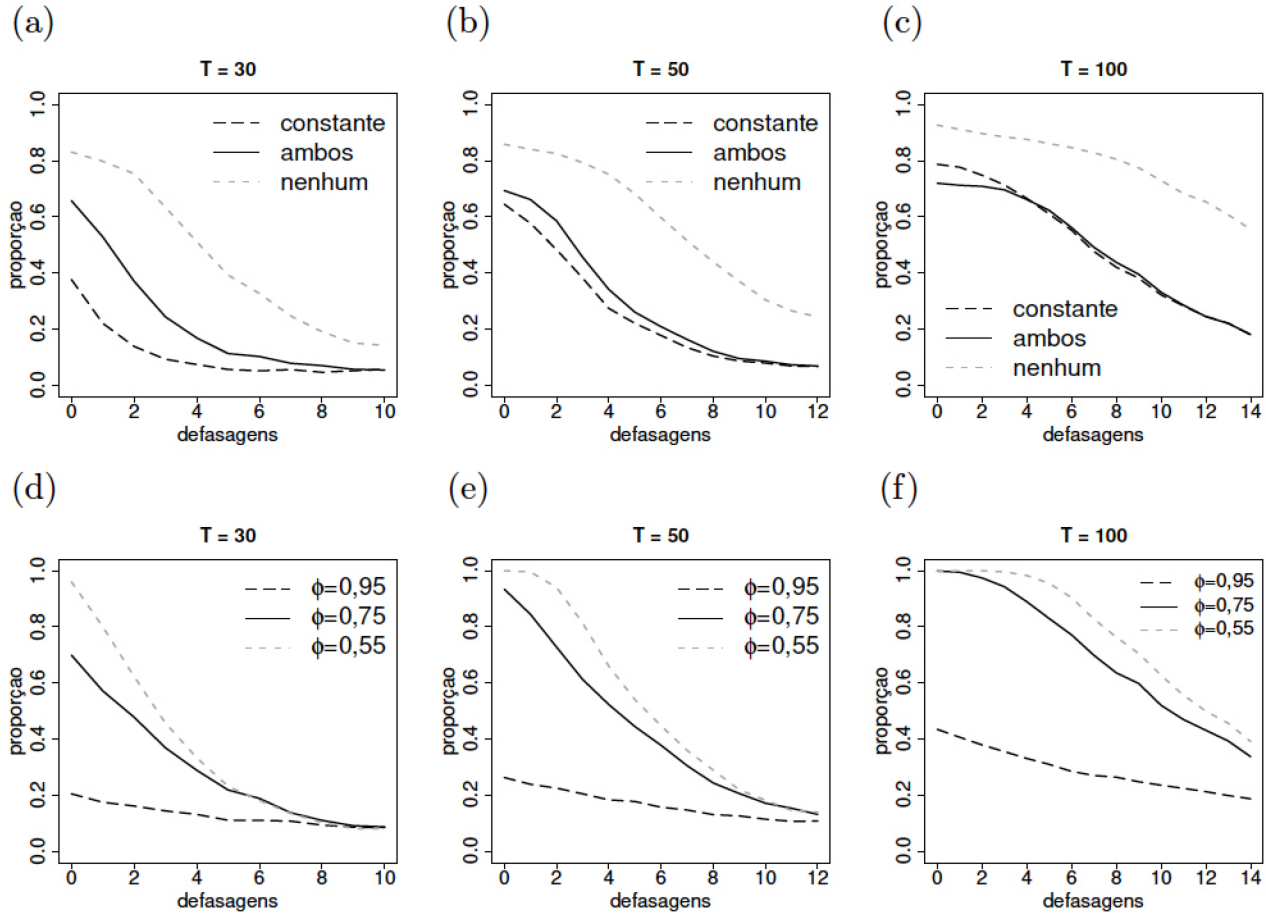


Figura 4.3: Representação da proporção de casos em que a hipótese nula é rejeitada e as interações entre as defasagens e: (a), (b), (c) número de componentes (com constante; ambos – com constante e tendência; nenhum – sem constante, nem tendência); (d), (e), (f) intensidade de ϕ

i. $T=25$

Na Tabela 4.5 são apresentados a média, a mediana e o desvio padrão da taxa de não rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, enquanto nos gráficos da Figura 4.4, box plots que informam a mediana, os quartis e valores máximo e mínimo, além da presença de valor fora do padrão e de assimetria.

De acordo com as informações apresentadas na Tabela 4.5, na amostra analisada, o melhor desempenho do teste ADF, por apresentar maior tendência central da taxa de não rejeição da hipótese nula, ocorre em modelos CT, selecionados pelo BIC, com determinação da defasagem máxima pelo l_4 e utilização do nível de significância 0,05. Observa-se também que a menor variabilidade desta taxa ocorre em modelos com tendência e constante, enquanto a maior variabilidade ocorre para nível de significância 0,10.

No gráfico a da Figura 4.4 é possível observar que o modelo sem constante e sem tendência (N) apresenta menores taxas de não rejeição, enquanto no gráfico b desta mesma figura os critérios de seleção apresentam comportamentos semelhantes em relação a esta taxa, apesar do BIC apresentar menor variabilidade. De acordo com o gráfico c, em relação ao nível de

Tabela 4.5: Média e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 25$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,90685	0,91165	0,07081
	BIC	0,91523	0,91790	0,06168
	HQC	0,90734	0,91245	0,07047
Defasagem Máxima	l12	0,87061	0,88465	0,05041
	l4	0,94900	0,97815	0,05619
Estrutura do modelo	C	0,93369	0,94375	0,04989
	N	0,84659	0,84330	0,05159
	CT	0,94879	0,95625	0,04574
Nível de Significância	0,05	0,92070	0,93071	0,04968
	0,10	0,88890	0,88465	0,07455

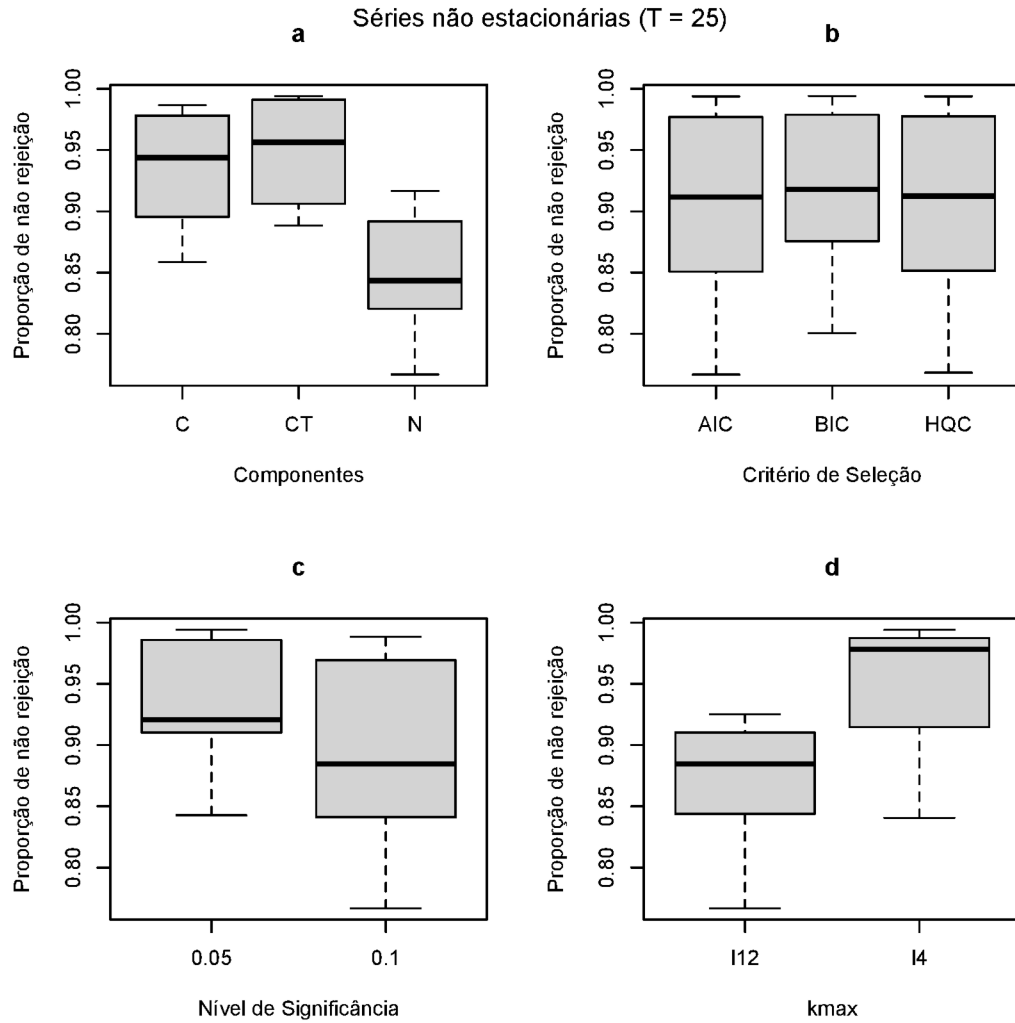


Figura 4.4: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 25$

significância, quando considerado 0,05, a proporção de não rejeição de H_0 assume maior valor e menor variabilidade, sendo bastante assimétrica principalmente entre os resultados centrais. Já quanto ao critério para selecionar o número máximo de defasagens, a maior proporção de não rejeição de H_0 é encontrada no l_4 , conforme gráfico d, que também apresenta resultados com distribuição bastante assimétrica.

Em nenhum dos gráficos da Figura 4.4 são encontrados valores fora do padrão, sendo a assimetria aparentemente mais forte encontrada para nível de significância de 0,05, com maior concentração dos valores centrais menores e critério l_4 para determinar defasagem máxima, com maior concentração dos valores maiores.

Após, o efeito de cada fator e respectivas interações são analisados por meio da modelagem GAMLSS, cujos resultados são apresentados na Tabela 4.6, seguida do wormplot para verificação das suposições teóricas, o que é feito nos resíduos do modelo, podendo ser visualizado na Figura 4.5. Nesta análise foi tomado como referência o critério de informação HQC, modelo de estrutura C, nível de significância 0.05 e critério de determinação da defasagem máxima l_{12} .

Ao observar o gráfico da Figura 4.5, nota-se que todos os pontos estão dentro da banda de confiança, indicando que o modelo está bem ajustado. Diante disso, parte-se para a interpretação dos resultados presentes na Tabela 4.6.

Tabela 4.6: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 25$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	2,32559	0,02256	0,00000
criterioAIC	-0,01209	0,02763	0,66782
criterioBIC	0,12857	0,02763	0,00031
modeloN	-0,62953	0,02763	0,00000
modeloCT	0,16403	0,02763	0,00003
nivelsig0,1	-0,51047	0,02471	0,00000
kmaxl4	1,90276	0,02471	0,00000
criterioAIC.modeloN	0,00597	0,03027	0,84640
criterioBIC.modeloN	0,00815	0,03027	0,79147
criterioAIC.modeloCT	0,01952	0,03027	0,52873
criterioBIC.modeloCT	-0,08739	0,03027	0,01129
criterioAIC.nivelsig0.1	0,00153	0,02471	0,95131
criterioBIC.nivelsig0.1	0,01643	0,02471	0,51620
criterioAIC.kmaxl4	-0,00817	0,02471	0,74541
criterioBIC.kmaxl4	-0,05958	0,02471	0,02921
modeloN.nivelsig0.1	0,04113	0,02471	0,11679
modeloCT.nivelsig0.1	0,07806	0,02471	0,00649
modeloN.kmaxl4	-1,22816	0,02471	0,00000
modeloCT.kmaxl4	0,70135	0,02471	0,00000
nivelsig0.1.kmaxl4	-0,27095	0,02018	0,00000

Os resultados apresentados na Tabela 4.6 indicam haver evidências de melhor desempenho

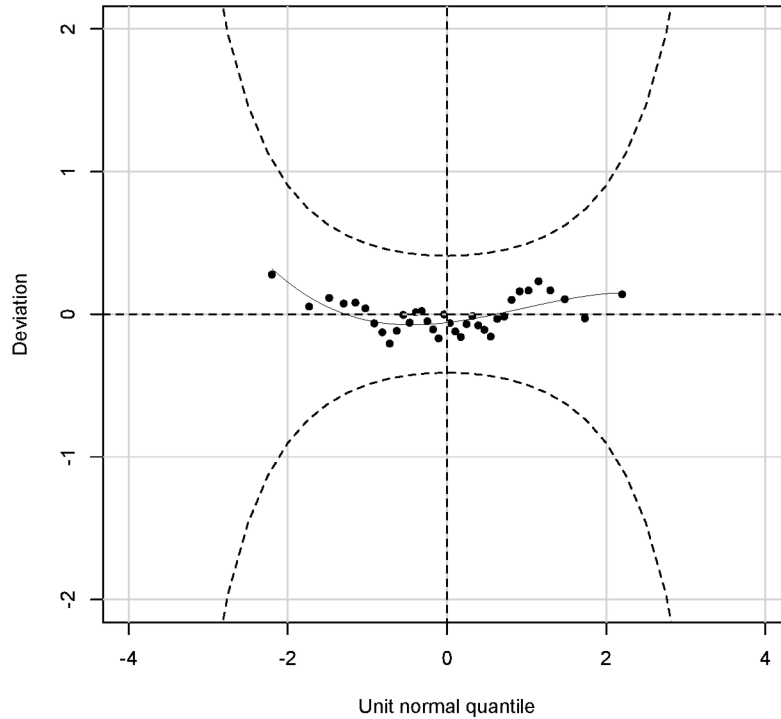


Figura 4.5: *Worm plot* gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 25$

do teste ADF quando utilizado o critério de informação BIC, pois há diferença significativa entre os resultados deste teste, quando faz uso dos critérios HQC e BIC, o mesmo não ocorrendo entre HQC e AIC. Em relação à estrutura do modelo, os resultados indicam haver evidências de melhor desempenho do teste ADF em modelo com estrutura CT e pior desempenho em modelo com estrutura N, pois há diferença significativa entre estrutura C e estrutura CT, o mesmo acontecendo para estrutura C e estrutura N. Também foram encontradas evidências de que o teste ADF apresenta melhor desempenho com nível de significância 0,05 e quando o critério de determinação da defasagem máxima é l_4 .

Em relação às interações, há evidências de sua presença entre critério de informação e diferentes tipos de modelos, conforme indicado na Tabela 4.6, pois o efeito de BICxCT mostrou-se significativo. Conforme pode ser visto na Figura 4.6, a proporção de não rejeição da hipótese nula, quando a parcimonia do modelo é mensurada pelo BIC, se mantém praticamente constante em modelos tipo CT, diferentemente do que acontece quando é usado o HQC, onde há um aumento na proporção embora esta proporção seja sempre maior com uso do BIC.

Conforme Tabela 4.6, também há evidência de interação entre critério de informação e critério para determinação da defasagem máxima. Conforme pode ser visto na Figura 4.7, a proporção de não rejeição da hipótese nula, quando mensurada pelo BIC, não aumenta com o uso do critério l_4 na determinação da defasagem máxima, diferentemente do que acontece quando é usado l_{12} , embora esta proporção seja sempre maior com uso do l_4 .

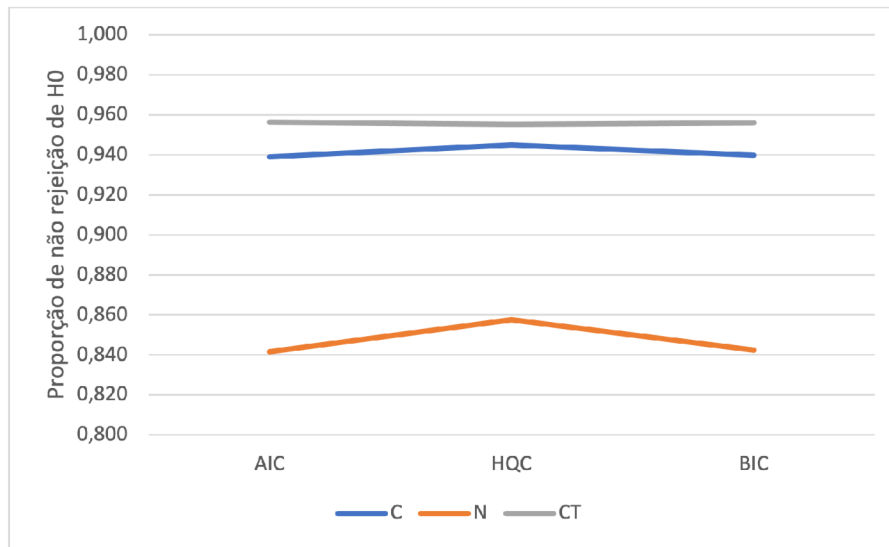


Figura 4.6: Interações entre critério de informação e tipo de modelo considerando a presença de raiz unitária e $T = 25$

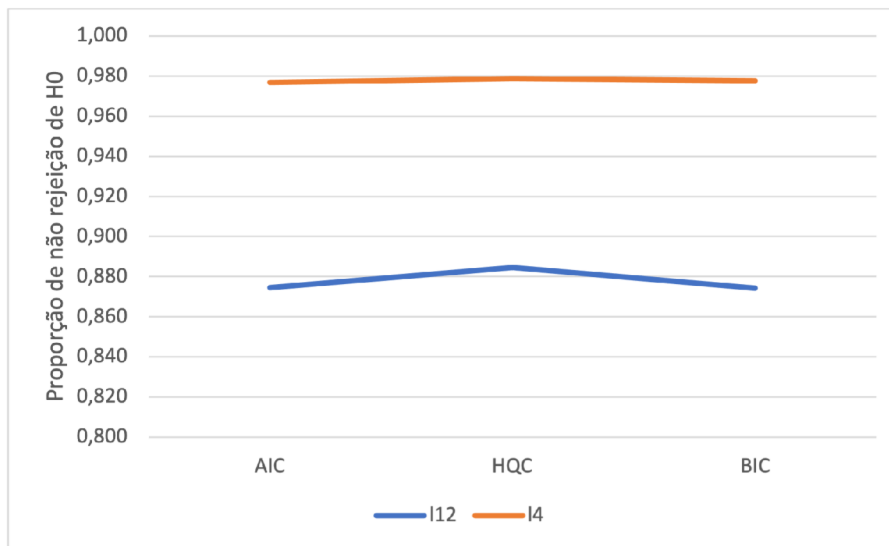


Figura 4.7: Interações entre critérios de seleção e de determinação da defasagem máxima considerando a presença de raiz unitária e $T = 25$

De acordo com Tabela 4.6, o tipo de estrutura do modelo também interage com nível de significância e com critério de determinação da defasagem máxima. De acordo com a Figura 4.8, em modelo CT, o crescimento da proporção de não rejeição da hipótese nula é maior em relação ao modelo C quando o nível de significância é 0,10, do que quando o nível de significância é 0,05, embora esta proporção seja sempre maior com uso deste último. Já a Figura 4.9 mostra que, quando a defasagem máxima é determinada por l_4 , a proporção de não rejeição da hipótese nula tem uma queda maior em modelo N, quando comparado ao C, e um crescimento maior em modelo CT, do que quando é usado o l_{12} , embora esta proporção seja sempre maior com uso do l_4 .

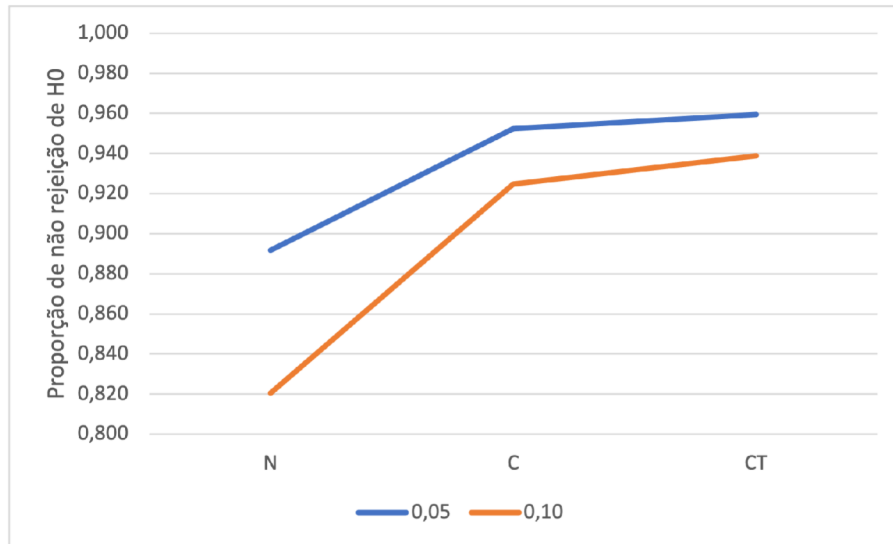


Figura 4.8: Interações entre tipo de modelo e nível de significância considerando a presença de raiz unitária e $T = 25$

O critério de determinação da defasagem máxima também interage com o nível de significância. De acordo com a 4.10, o crescimento da proporção de não rejeição da hipótese nula de l_{12} para l_4 é maior quando usado o nível de significância 0,10 do que quando usado 0,05, embora esta proporção seja sempre maior com uso do 0,05.

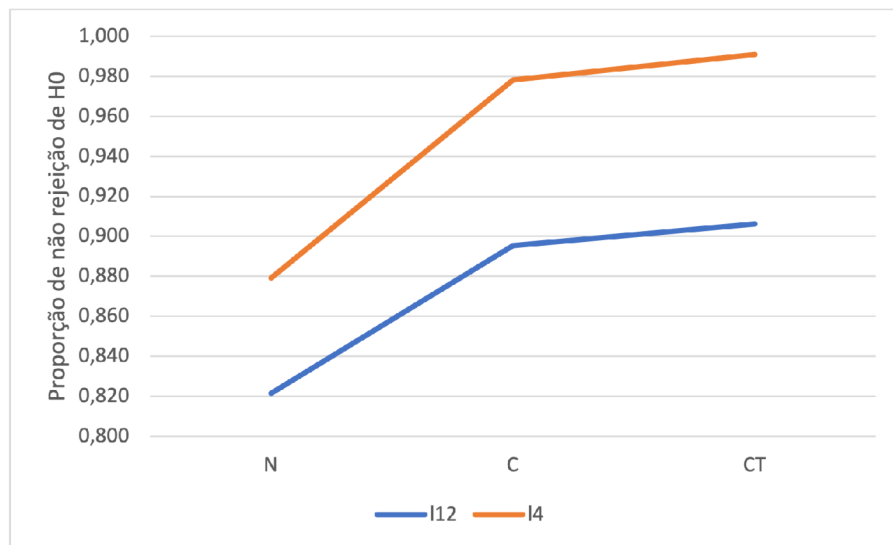


Figura 4.9: Interações entre o tipo de modelo e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 25$

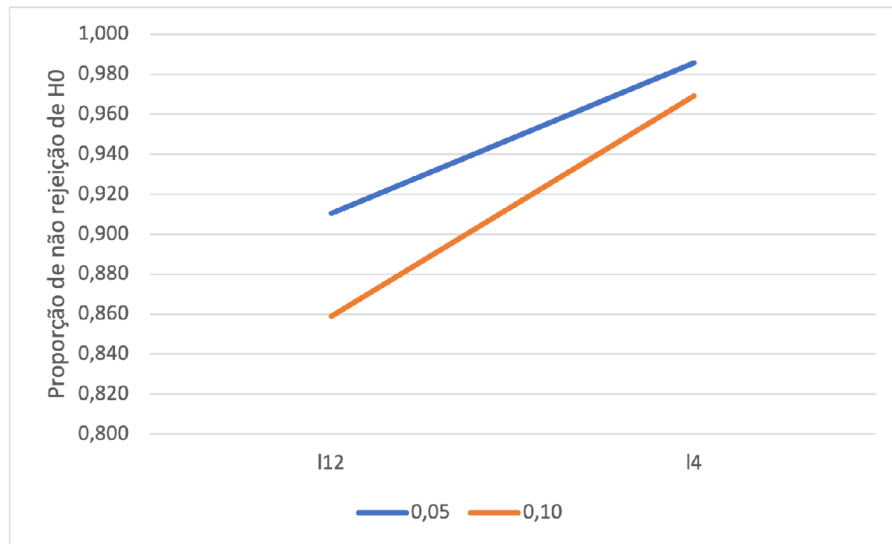


Figura 4.10: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 25$

ii. T=50

Para a amostra de séries com comprimento $T = 50$, na Tabela 4.7 são apresentadas algumas medidas descritivas da taxa de não rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, e os respectivos box plots são mostrados na Figura 4.11.

Tabela 4.7: Média e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 50$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,94679	0,97390	0,05494
	BIC	0,95709	0,98490	0,05216
	HQC	0,95223	0,97975	0,05269
Defasagem Máxima	l_{12}	0,94704	0,97390	0,05229
	l_4	0,95704	0,98575	0,05254
Estrutura do modelo	C	0,97978	0,98020	0,01033
	N	0,88655	0,87965	0,03663
	CT	0,98978	0,99230	0,00728
Nível de Significância	0,05	0,96644	0,98580	0,03424
	0,10	0,93763	0,97550	0,06274

As medidas apresentadas na Tabela 4.7, indicam que na amostra analisada, os resultados se assemelham aos encontrados para os caso já analisado em que $T=25$, apresentado o mesmo comportamento dos fatores em relação à taxa de não rejeição de H_0 e à variabilidade.

No gráfico a da Figura 4.11, embora apresente semelhanças ao caso de $T=25$, é possível notar que o modelo N apresenta variabilidade muito maior que os demais níveis e menor proporção de não rejeição de H_0 . Além disso, o CT é o que apresenta a maior proporção de não rejeição. No gráfico b, observa-se maior variabilidade quanto é adotado o AIC e maiores valores da proporção de não rejeição de H_0 , para o caso do critério BIC. Em relação ao gráfico c, destaca-se que o nível de significância de 0,10 apresenta consideravelmente maior variabilidade e menor taxa de rejeição de H_0 , já em d verifica-se que melhores resultados são encontrados quando adotado o critério l_4 . Nos casos de b, c e d pode ser observada assimetria à parte superior, além disso, assim como em $T=25$, nenhum dos gráficos mostrou a presença de outliers.

Em seguida foi gerado o modelo através da modelagem GAMLSS, em que resposta obtida pode ser visualizada na Tabela 4.8: Nesta análise é tomado como referência os mesmos níveis dos fatores considerados nos experimentos para $T = 25$, ou seja: critério de informação HQC, modelo de estrutura tipo C, nível de significância 0.05 e critério de determinação da defasagem máxima l_{12} .

A Figura 4.12 mostra o wormplot para a verificação das suposições teóricas do modelo em seus resíduos.

Ao observar o gráfico da Figura 4.12, verifica-se que não há pontos fora da banda de confiança, apresentando evidências de que o modelo está bem ajustado. Assim, parte-se para a

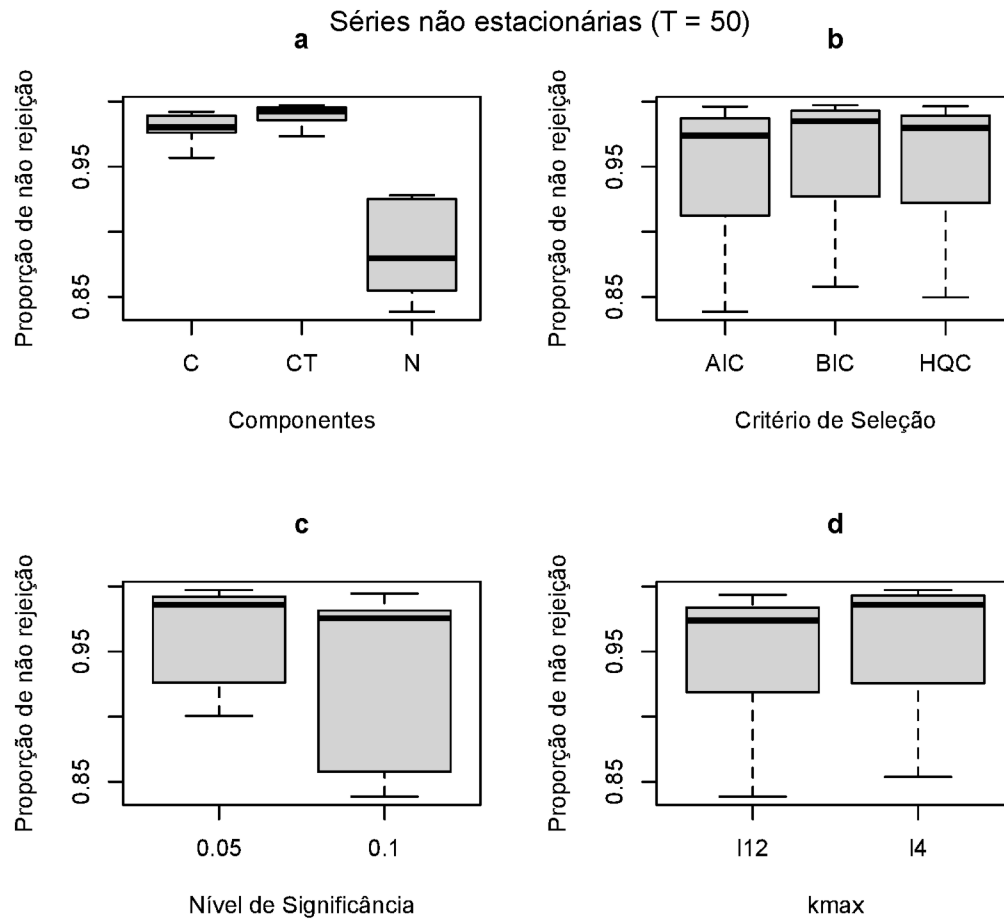


Figura 4.11: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 50$

interpretação dos resultados presentes na Tabela 4.8.

Nos resultados apresentados na Tabela 4.8, todos os fatores se apresentam com efeito significativo. Assim como na série de comprimento 25, o melhor desempenho foi encontrado para BIC entre os critérios de informação avaliados, CT entre as estruturas de modelo investigadas, 0,05 entre os níveis de significância considerados e l_4 como critério de determinação da defasagem máxima a ser usada na modelagem.

Já em relação às interações, a Tabela 4.8 apresenta evidências de que há diferença significativa no desempenho do teste ADF por critério de informação em diferentes tipos de estrutura. No gráfico da Figura 4.13 é possível notar que a diferença entre a proporção de não rejeição da hipótese nula entre os modelos N e C é maior quando o critério adotado é o AIC e menor quando o critério de informação para a seleção de defasagens é o HQC, visto que há um aumento na proporção de rejeição de H_0 no modelo N quando o critério adotado é o HQC. Além disso, a diferença entre os modelos N e CT é sutilmente superior, embora significativa, quando o critério adotado é o HQC, uma vez que há um crescimento da proporção no modelo CT quando há a alteração de de HQC para BIC.

Também, segundo a Tabela 4.8, há evidências de interação significativa entre os resultados

Tabela 4.8: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 50$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	3,99319	0,03373	0,00000
criterioAIC	-0,29201	0,04131	0,00000
criterioBIC	0,42066	0,04131	0,00000
modeloN	-1,63271	0,04131	0,00000
modeloCT	0,43982	0,04131	0,00000
nivelsig0,1	-0,62113	0,03695	0,00000
kmaxl4	0,66658	0,03695	0,00000
criterioAIC.modeloN	0,10469	0,04525	0,03528
criterioBIC.modeloN	-0,22788	0,04525	0,00015
criterioAIC.modeloCT	-0,02052	0,04525	0,65670
criterioBIC.modeloCT	0,16306	0,04525	0,00261
criterioAIC.nivelsig0.1	0,01158	0,03695	0,75819
criterioBIC.nivelsig0.1	-0,03302	0,03695	0,38558
criterioAIC.kmaxl4	0,19191	0,03695	0,00011
criterioBIC.kmaxl4	-0,24030	0,03695	0,00001
modeloN.nivelsig0.1	0,01589	0,03695	0,67319
modeloCT.nivelsig0.1	0,11775	0,03695	0,00612
modeloN.kmaxl4	-0,46037	0,03695	0,00000
modeloCT.kmaxl4	0,49838	0,03695	0,00000
nivelsig0.1.kmaxl4	-0,17974	0,03017	0,00003

do teste ADF com uso de diferentes critérios de informação e diferentes critérios de definição da defasagem máxima, ilustrada na Figura 4.14. Mesmo que o desempenho do teste ADF sempre tenha sido melhor com defasagem máxima determinada por l_4 , quando é adotado o l_{12} , há um crescimento da taxa de não rejeição de H_0 quando há a alteração de AIC para HQC, ocorrendo o mesmo quando há a troca de HQC para BIC, ficando, neste caso, os valores da proporção próximos aos de quando é escolhido l_4 . Na amostra de séries com $T = 25$, diferentemente do ocorrido para $T = 50$, o AIC apresentou o mesmo comportamento do HQC.

Em relação à interação entre estrutura do modelo e nível de significância, como pode ser observado na Figura 4.15, houve interação significativa apontada na Tabela 4.8. Ao nível de significância de 0,10, nota-se um crescimento da proporção de não rejeição de H_0 quando o modelo é alterado de C para CT maior do que quando o nível de significância é 0,05, concordando com o achado em séries com $T = 25$. Este crescimento também ocorre ao nível de 0,05, porém de maneira muito mais sutil, cabendo salientar este nível apresentou desempenho superior em todos os modelos considerados.

Por fim, analisando a Tabela 4.8, assim como nos itens anteriores, é possível verificar indícios de interação entre a estrutura do modelo e o critério de seleção da defasagem máxima (Figura 4.16) e da interação entre o nível de significância e o critério de seleção da defasagem máxima, que pode ser observada na Figura 4.17. Nestes casos, os achados são os mesmos que para o

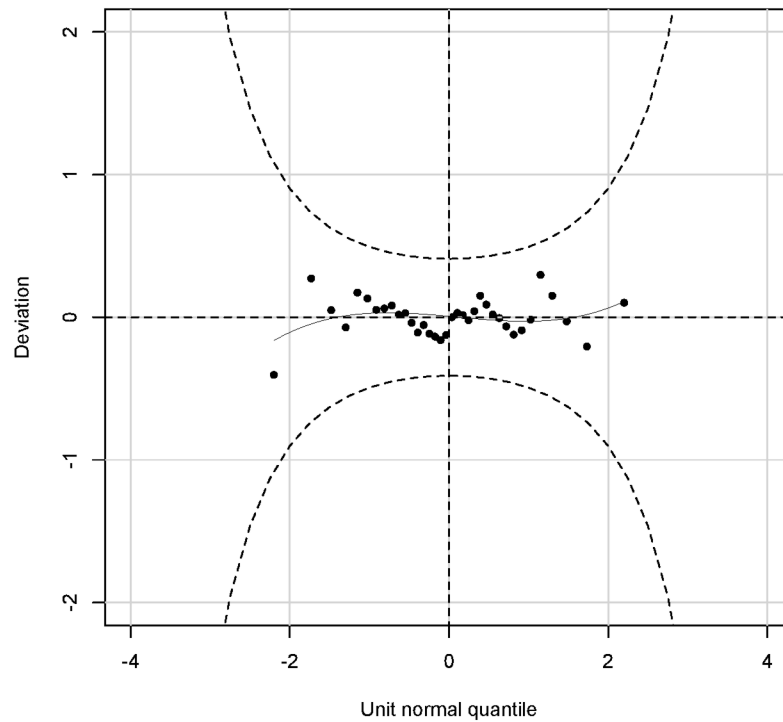


Figura 4.12: *Worm plot* gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 50$

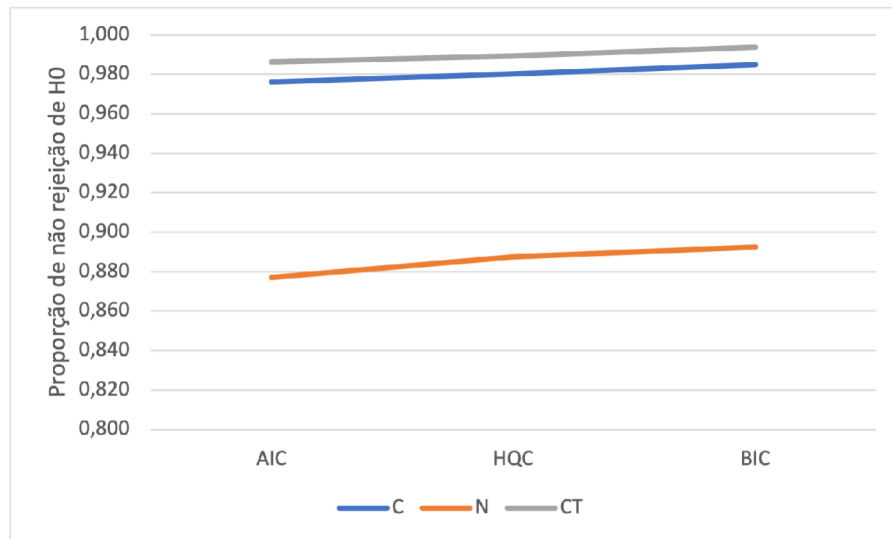


Figura 4.13: Interações entre critério de informação e tipo de modelo, considerando a presença de raiz unitária e $T=50$

caso de $T=25$, embora caiba salientar que considerando a Figura 4.16, no experimento atual a proporção de não rejeição de H_0 fique bastante próxima para os níveis de significância de 0,05 e 0,10, quando o modelo considerado é o CT. Além disso, no caso atual, o crescimento da taxa de não rejeição, quando há a troca de l_4 por l_{12} apresenta um aumento de menor intensidade do que no item anterior.

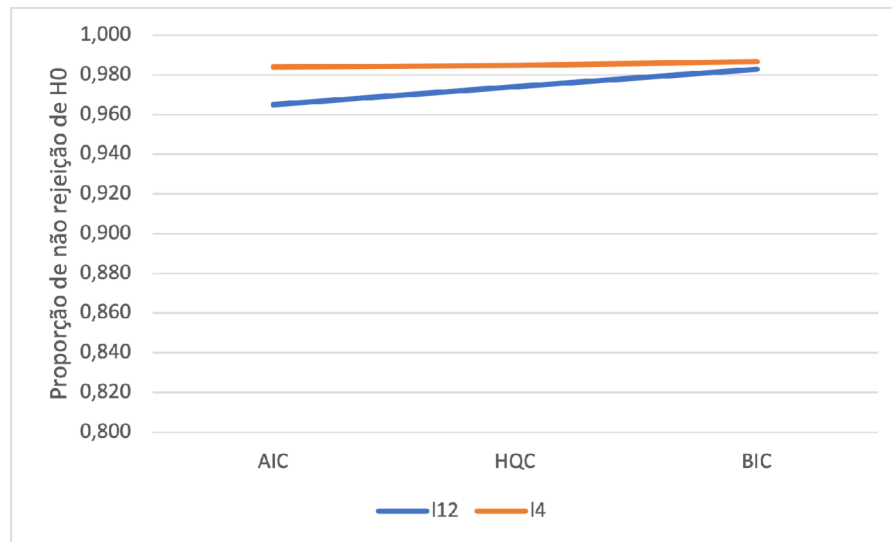


Figura 4.14: Interações entre critério de informação e de determinação da defasagem máxima, considerando a presença de raiz unitária e $T=50$

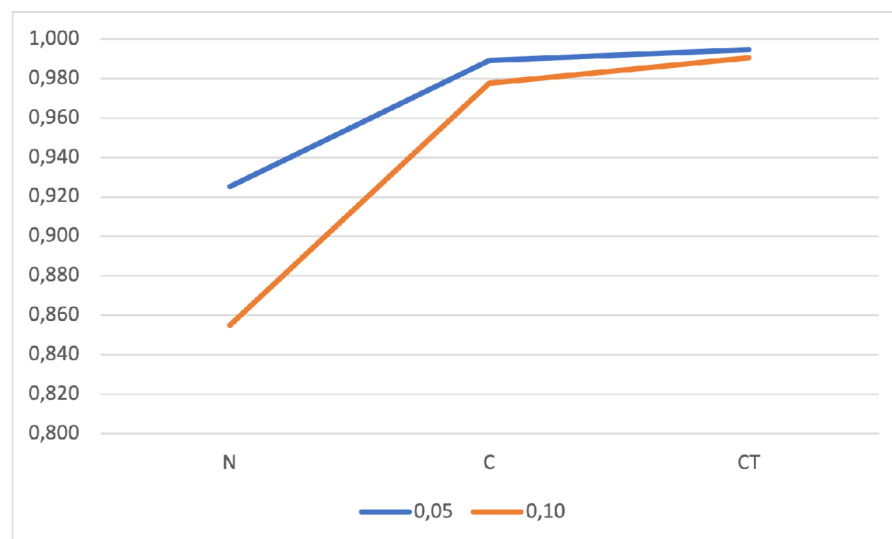


Figura 4.15: Interações entre tipo de modelo e nível de significância, considerando a presença de raiz unitária e $T=50$

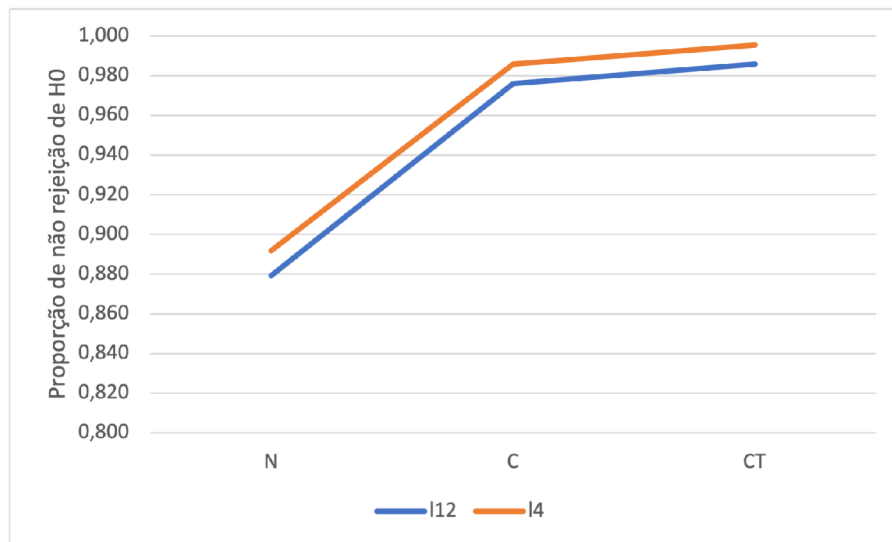


Figura 4.16: Interações entre o tipo de modelo e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=50$

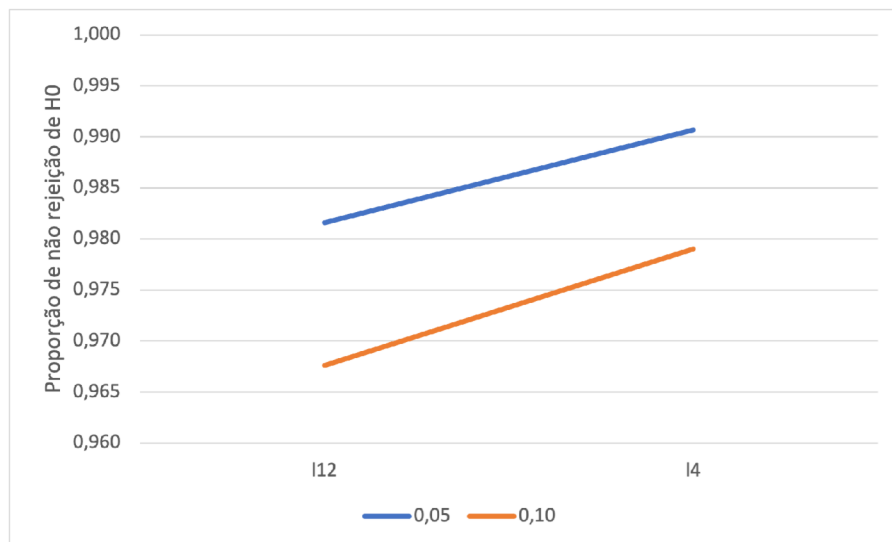


Figura 4.17: Interações entre o nível de significância e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=50$

iii. T=100

Para a amostra de séries com comprimento $T = 100$, na Tabela 4.9, são apresentadas medidas descritivas da taxa de não rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, enquanto os box plots são mostrados na Figura 4.18.

Tabela 4.9: Média e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 100$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,93636	0,95395	0,06023
	BIC	0,94188	0,96100	0,05810
	HQC	0,94008	0,95870	0,05860
Defasagem Máxima	l_2	0,93920	0,95640	0,05787
	l_4	0,93967	0,95905	0,05845
Estrutura do modelo	C	0,98798	0,98780	0,00495
	N	0,89891	0,89940	0,03514
	CT	0,93143	0,93285	0,06902
Nível de Significância	0,05	0,97394	0,99235	0,03035
	0,10	0,90493	0,86760	0,05764

As medidas apresentadas na Tabela 4.9, indicam que na amostra analisada, o melhor desempenho do teste ADF ocorre em modelos C, selecionados pelo BIC, com determinação da defasagem máxima pelo l_4 e utilização do nível de significância 0,05. Neste caso, apenas a estrutura do modelo apresentou achados diferentes das séries de outros tamanhos já analisadas. Com relação aos resultados das medidas de variabilidade, é possível notar que a variação da taxa apresentou menor valor em modelos C e maior valor nos CT, diferentemente dos achados anteriores.

Ao analisar o gráfico a da Figura 4.18 é possível notar que o modelo CT apresenta uma variabilidade muito maior que os demais. Além disso, o modelo com C é o que apresenta a maior proporção de não rejeição e a menor variabilidade. No gráfico b, observa-se caixas semelhantes para todos os critérios, cabendo destaque à variabilidade suavemente maior no AIC. Em relação ao gráfico c, destaca-se que o nível de significância de 0,10 apresenta consideravelmente maior variabilidade, cabendo observar que há assimetria com maior concentração de resultados na parte superior em 0,05 e, embora também haja assimetria quando o nível de significância é 0,10, a maior concentração ocorre entre os valores inferiores. No gráfico d, os diagramas de caixa são semelhantes para l_4 e l_{12} .

Em seguida foi gerado o modelo através da modelagem GAMLSS, em que resposta obtida pode ser visualizada na Tabela 4.10. Nesta análise é tomado como referência para o fator tipo de estrutura do modelo o nível CT, diferentemente do nível considerado nas análises anteriores. Os níveis referência dos demais fatores permanecem os mesmos. São eles: critério de informação HQC, nível de significância 0.05 e critério de determinação da defasagem máxima l_{12} .

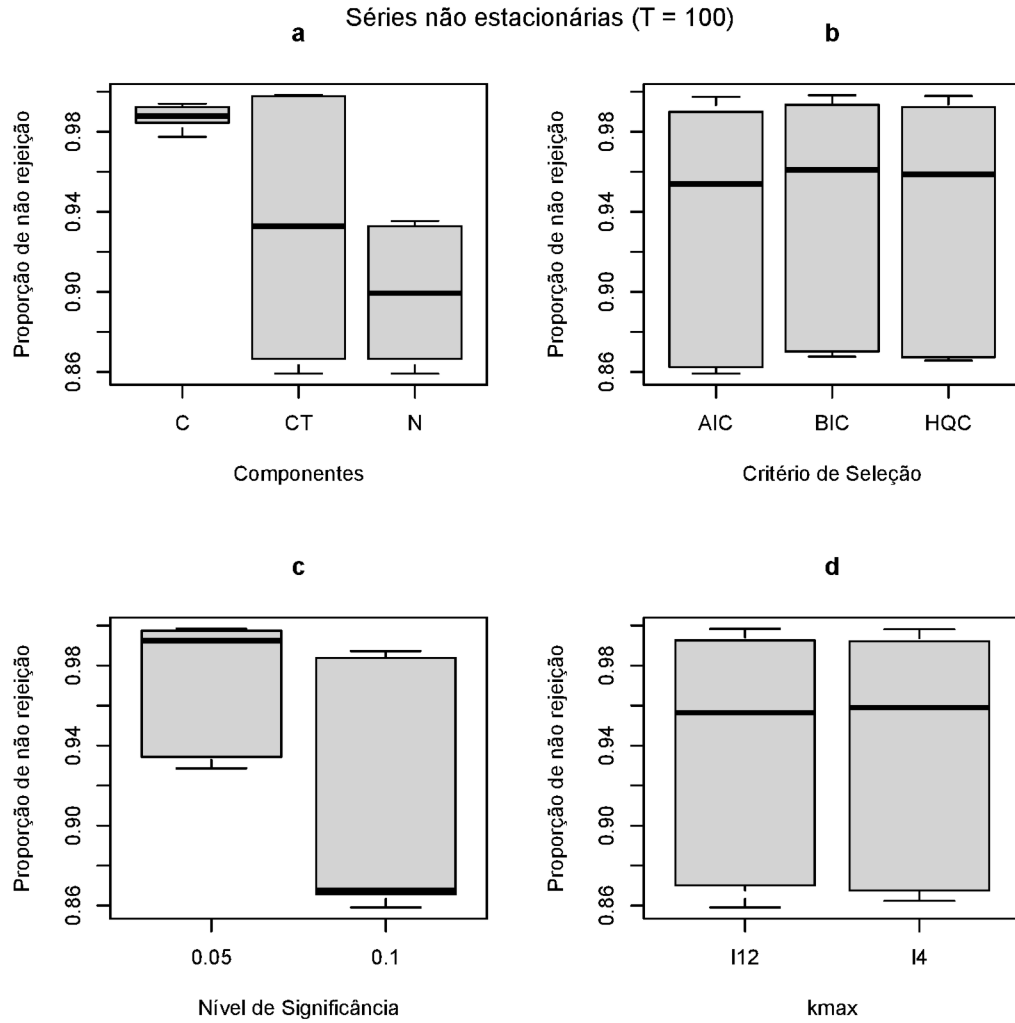


Figura 4.18: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 100$

A Figura 4.19 mostra o wormplot para a verificação dos resíduos do modelo.

Ao observar o gráfico da Figura 4.19, assim como nos casos anteriores, há evidências de que o modelo está bem ajustado, seguindo-se para a interpretação dos resultados presentes na Tabela 4.10.

Os resultados apresentados na Tabela 4.10 seguem o padrão dos achados anteriores, entretanto, para $T=100$, não há indícios de diferença significativa entre selecionar a defasagem máxima por l_4 ou l_{12} .

Além disso, a quantidade de iterações significativas diminuiu bastante. Ao analisá-las, há evidências de interação significativa entre os resultados do teste ADF com uso de diferentes critérios de informação e diferentes critérios de definição da defasagem máxima (Figura 4.20). Neste caso, o desempenho do teste quando adotado o critério AIC é superior se a defasagem máxima for selecionada por l_4 , em vez de l_{12} , o mesmo não acontecendo quando adotados HQC ou BIC.

Em relação à interação entre estrutura do modelo e nível de significância, como pode ser

Tabela 4.10: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 100$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	5,99902	0,06041	0,00000
criterioAIC	-0,36421	0,07398	0,00018
criterioBIC	0,22795	0,07398	0,00760
modeloC	1,16746	0,07398	0,00000
modeloN	-3,35393	0,07398	0,00000
nivelsig0,1	-4,14824	0,06617	0,00000
kmaxl4	0,10372	0,06617	0,13786
criterioAIC.modeloC	-0,03047	0,08104	0,71219
criterioBIC.modeloC	-0,02677	0,08104	0,74568
criterioAIC.modeloN	0,15558	0,08104	0,07411
criterioBIC.modeloN	-0,11532	0,08104	0,17520
criterioAIC.nivelsig0.1	0,11950	0,06617	0,09102
criterioBIC.nivelsig0.1	-0,08344	0,06617	0,22657
criterioAIC.kmaxl4	0,19708	0,06617	0,00938
criterioBIC.kmaxl4	-0,08893	0,06617	0,19895
modeloN.nivelsig0.1	3,46362	0,06617	0,00000
modeloC.nivelsig0.1	3,39850	0,06617	0,00000
modeloC.kmaxl4	-0,03867	0,06617	0,56766
modeloN.kmaxl4	-0,11598	0,06617	0,10006
nivelsig0.1.kmaxl4	-0,05205	0,05403	0,35059

observado na Figura 4.21, houve interação significativa apontada na Tabela 4.10. Neste caso cabe ressaltar que ao selecionar o nível de significância de 0,05 o teste ADF apresentou melhor desempenho independente da estrutura do modelo. Todavia, para este nível de significância, quando a estrutura do modelo é modificada de C para CT, é observada uma queda da proporção de não rejeição de H_0 . Já quando o modelo é alterado de CT para N, a proporção se mantém praticamente constante. Ao fazer a mesma análise considerando o nível de significância de 0,05, há um crescimento na proporção quando o modelo é alterado de C para CT e uma queda acentuada quando há alteração de CT para N.

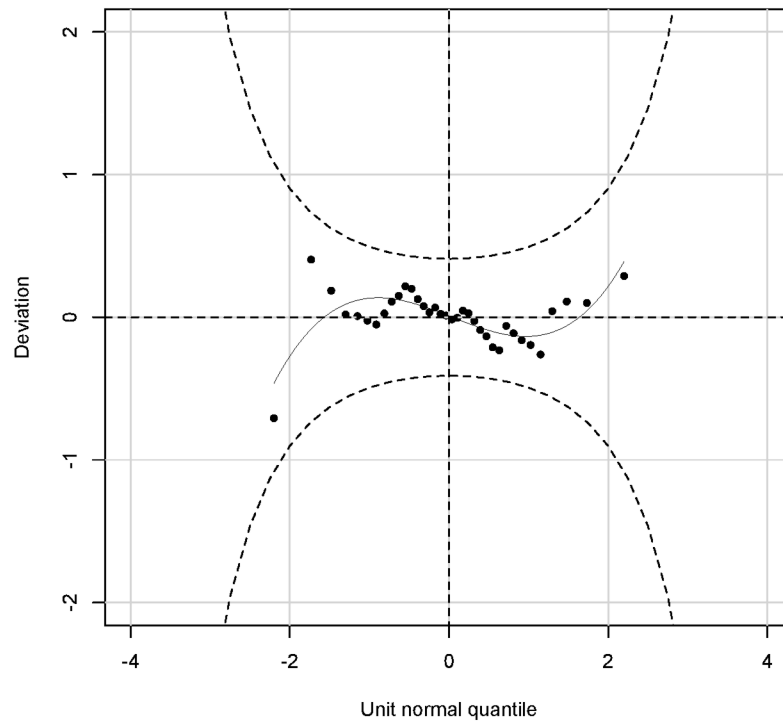


Figura 4.19: *Worm plot* gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 100$

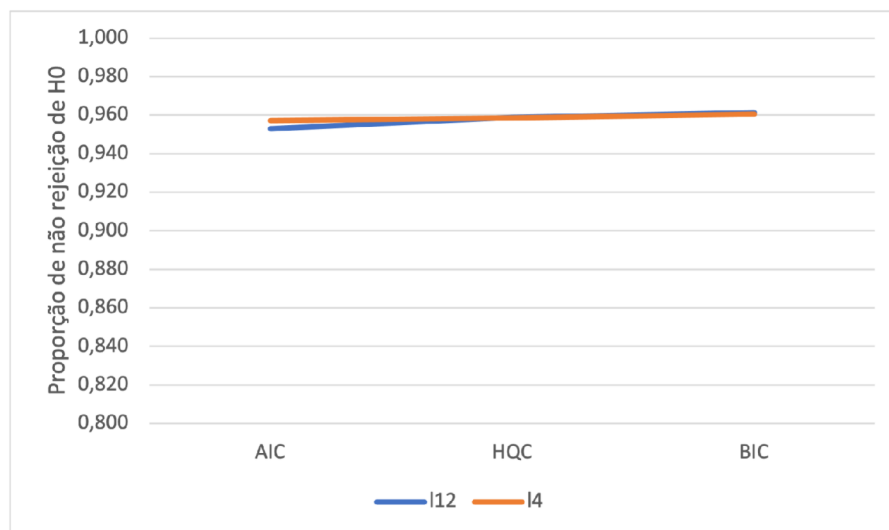


Figura 4.20: Interações entre critérios de seleção e de determinação da defasagem máxima, considerando a presença de raiz unitária e $T=100$

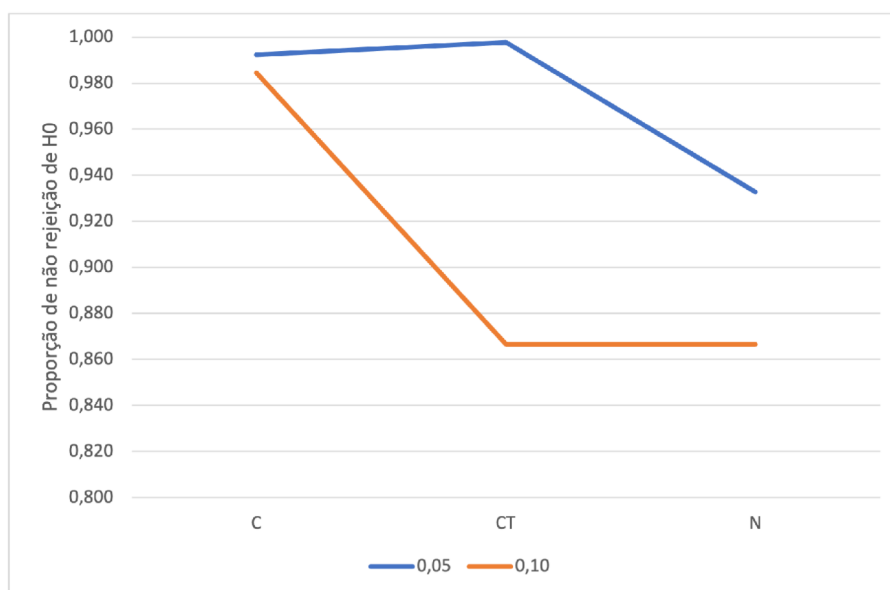


Figura 4.21: Interações entre tipo de modelo e nível de significância, considerando a presença de raiz unitária e $T=100$

iv. Considerações finais sobre as séries não estacionárias

Os resultados dos experimentos para as séries não estacionárias com os diferentes tamanhos se mostraram bastantes semelhantes em relação à influência dos fatores na rejeição da hipótese nula do teste ADF. A Tabela 4.11 apresenta os fatores associados ao melhor desempenho do teste para cada comprimento de série.

Tabela 4.11: Resumo dos resultados na presença de raiz unitária

Comprimento da série	Critério de Informação	Defasagem Máxima	Estrutura do Modelo	Nível de significância
25	BIC	l_4	CT	0,05
50	BIC	l_4	CT	0,05
100	BIC	Indiferente	C	0,05

Para todos os experimentos, exceto os de $T=100$, quando a série não é estacionária, o teste apresenta melhor desempenho quando o número de defasagens é selecionado por BIC, a defasagem máxima por l_4 e o nível de significância adotado é 0,05. Além disso, o modelo que apresentou maior proporção de não rejeição de H_0 é o CT. Já para os experimentos em que foi considerado $T=100$, o critério de seleção de defasagem máxima não apresentou relação significativa com a taxa de rejeição de H_0 e o modelo em que foram encontrados os melhores resultados é o N. Os demais achados são semelhantes aos encontrados para os outros experimentos.

De maneira geral, o melhor desempenho do teste aconteceu em séries de tamanho 50, embora ele se torne mais robusto em séries de tamanho 100. Estas séries maiores apresentaram menos interações significativas, assim como também o efeito principal de um dos fatores (critério de determinação da defasagem máxima) deixou de ser significativo.

Em relação às interações, apesar de em diversos experimentos se apresentarem significativas, somente em um caso foi encontrada alteração no resultado do teste. Para as séries em que $T=100$, se a defasagem máxima for selecionada por l_{12} , os critérios de seleção de defasagem associados a um melhor desempenho do teste ADF, são HQC e BIC. Por outro lado, ao determinar a defasagem máxima por l_4 , o AIC apresentou melhores resultados. Entretanto, cabe ressaltar que para estes experimentos, não foram encontradas evidências de diferenças significativas entre os níveis do fator defasagem máxima.

No que se refere às médias, os fatores apresentaram comportamento semelhante em relação ao comprimento das séries, tendo as menores proporções de não rejeição da hipótese nula para as séries de comprimento 25, aumentando estas taxas em $T = 50$ e apresentando uma leve queda quando $T=100$. As exceções são os fatores CT e 0,10, que apresentaram uma queda mais acentuada entre os tamanhos 50 e 100. Além destes, os fatores N, C e 0,05 apresentaram crescimento conforme aumentou o T .

O critério de informação HQC apresentou média próxima ao do AIC em $T=25$ e passou a se aproximar do BIC conforme o aumento do comprimento da série, tendo seu valor médio

próximo ao do BIC em $T=100$. Quanto ao critério de seleção da defasagem máxima, o l_{12} se aproxima do l_4 conforme aumenta o comprimento da série.

Outro ponto observado é o fraco desempenho do teste ADF em modelos que não apresentam constante e nem tendência, independentemente do tamanho da série, embora na prática esta estrutura não seja muito frequente. Além disso, o desempenho dos teste também foi ruim quando utilizado o critério de determinação da defasagem máxima foi l_{12} em séries pequenas ($T=25$).

Por fim, cabe salientar que na série de comprimento igual a 100 houve o menor número de interações significativas, cabendo investigar se para séries maiores existe menor influência da interação entre os fatores no resultado do teste.

4.2.2 Séries Estacionárias($\phi < 1$)

No caso das séries estacionárias é esperado que o teste ADF rejeite a hipótese nula da presença de raiz unitária. Deste modo, é tomada a taxa de rejeição de H_0 para a realização das análises.

i. $T=25$

Na Tabela 4.12 são apresentados a média, a mediana e o desvio padrão da taxa de rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, enquanto nos gráficos da Figura 4.22, box plots que informam a mediana, os quartis e valores máximo e mínimo, além da presença de valor fora do padrão e de assimetria.

Tabela 4.12: Média e o desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 25$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,52137	0,50175	0,28545
	BIC	0,52044	0,45360	0,28899
	HQC	0,51883	0,46855	0,28684
Defasagem Máxima	112	0,38518	0,35770	0,17491
	14	0,65524	0,73460	0,30911
Estrutura do modelo	C	0,66483	0,70445	0,25893
	N	0,62288	0,54325	0,25014
	CT	0,27293	0,28235	0,14972
Nível de Significância	0,05	0,46268	0,36185	0,28534
	0,10	0,57774	0,52445	0,27411
Φ	0,55	0,55708	0,50700	0,24634
	0,75	0,55641	0,46180	0,28184
	0,95	0,44715	0,38195	0,31452

De acordo com as informações apresentadas na Tabela 4.12, na amostra analisada, o melhor desempenho do teste ADF, por apresentar maior tendência central da taxa de rejeição da

hipótese nula, ocorre em modelos C, selecionados pelo AIC, com determinação da defasagem máxima pelo l_4 , utilização do nível de significância 0,10 e valor de ϕ 0,55, embora o valor da média deste tenha sido bem próximo ao de 0,75. Observa-se também que a menor variabilidade da taxa ocorre em modelos com tendência e constante, enquanto a maior variabilidade ocorre $\phi = 0,95$.

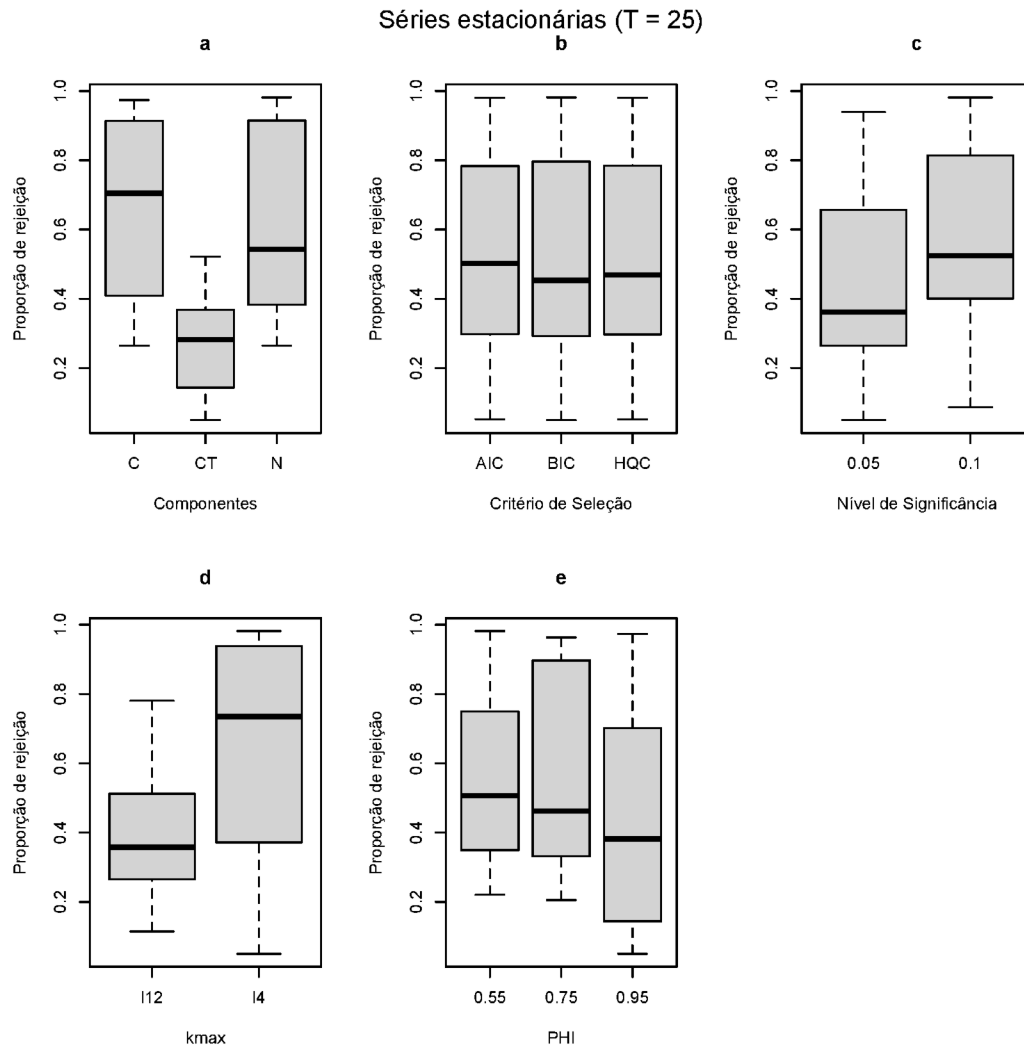


Figura 4.22: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 25$

No gráfico a da Figura 4.22 é possível observar que o modelo CT apresenta menores taxas de rejeição e menor variabilidade, enquanto no gráfico b desta mesma figura os critérios de seleção apresentam comportamento semelhante em relação a esta taxa, apesar do BIC apresentar maior variabilidade e menor valor de mediana. De acordo com o gráfico c, em relação ao nível de significância, quando considerado 0,10, a proporção de rejeição de H_0 assume maior valor. Já quanto ao critério para selecionar o número máximo de defasagens, a maior proporção de não rejeição de H_0 é encontrada no l_4 , conforme gráfico d, que também apresenta maior variabilidade. Por fim, o gráfico e mostra que a maior taxa de rejeição mediana do teste é encontrada para $\phi = 0,55$, já a maior variabilidade é encontrada para o valor de 0,75. Além

disso, em nenhum dos gráficos foi detectada a presença de outliers.

Após, o efeito de cada fator e respectivas interações são analisados por meio da modelagem GAMLSS, cujos resultados são apresentados na Tabela 4.13, seguida do wormplot para verificação das suposições teóricas, o que é feito nos resíduos do modelo, podendo ser visualizado na Figura 4.23. Nesta análise foi tomado como referência o critério de informação HQC, modelo de estrutura N, nível de significância 0.05, critério de determinação da defasagem máxima l_{12} e $\phi = 0,75$.

Tabela 4.13: Análise do efeito principal dos fatores e respectivas interações na taxa rejeição da hipótese nula em experimentos com séries estacionárias e $T = 25$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	-0,63595	0,15348	0,00009
criterioAIC	0,03622	0,17058	0,83246
criterioBIC	-0,02006	0,17058	0,90669
modeloC	-0,27702	0,17058	0,10869
modeloCT	-0,75064	0,17058	0,00004
nivelsig0,1	0,76252	0,14889	0,00000
kmaxl4	2,76920	0,14889	0,00000
phi0,55	0,95187	0,17058	0,00000
phi0,95	-0,64709	0,17058	0,00030
criterioAIC.modeloC	-0,04502	0,15793	0,77636
criterioBIC.modeloC	0,04740	0,15793	0,76494
criterioAIC.modeloCT	-0,03115	0,15793	0,84417
criterioBIC.modeloCT	0,03739	0,15793	0,81350
criterioAIC.nivelsig0.1	0,02256	0,12895	0,86158
criterioBIC.nivelsig0.1	0,00455	0,12895	0,97195
criterioAIC.kmaxl4	-0,03189	0,12895	0,80537
criterioBIC.kmaxl4	-0,01297	0,12895	0,92013
modeloC.nivelsig0.1	-0,12324	0,12895	0,34236
modeloCT.nivelsig0.1	-0,31226	0,12895	0,01794
modeloC.kmaxl4	0,36175	0,12895	0,00644
modeloCT.kmaxl4	-1,86025	0,12895	0,00000
nivelsig0.1.kmaxl4	0,24007	0,10528	0,02552
modeloC.phi0.55	-1,43546	0,15793	0,00000
modeloCT.phi0.55	-0,55530	0,15793	0,00076
modeloC.phi0.95	2,50030	0,15793	0,00000
modeloCT.phi0.95	0,02949	0,15793	0,85237
nivelsig0.1.phi0.55	0,01536	0,12895	0,90552
nivelsig0.1.phi0.95	-0,05105	0,12895	0,69331
criterioAIC.phi0.55	-0,01366	0,15793	0,93133
criterioBIC.phi0.55	-0,04173	0,15793	0,79232
criterioAIC.phi0.95	0,00997	0,15793	0,94984
criterioBIC.phi0.95	0,00711	0,15793	0,96423
kmaxl4.phi0.55	-0,61777	0,12895	0,00001
kmaxl4.phi0.95	-1,82182	0,12895	0,00000

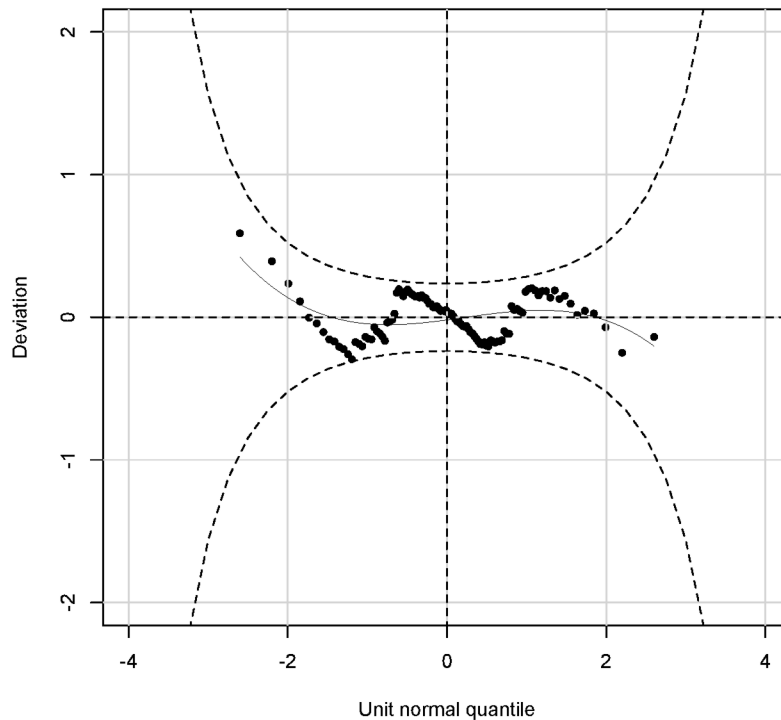


Figura 4.23: *Worm plot* gerado para o modelo GAMLSS considerando a série estacionária e $T = 25$

Ao observar o gráfico da Figura 4.23, nota-se que todos os pontos estão dentro da banda de confiança, indicando que o modelo está bem ajustado. Diante disso, parte-se para a interpretação dos resultados presentes na Tabela 4.13.

Os resultados apresentados na Tabela 4.13 indicam não haver evidências de melhor desempenho do teste ADF quando o critério de informação é alterado. Em relação à estrutura do modelo, os resultados indicam haver evidências de melhor desempenho do teste ADF em modelo com estrutura N em comparação ao CT, enquanto o C não apresentou diferença significativa quando comparado ao N. Também foram encontradas evidências de que o teste ADF apresenta melhor desempenho com nível de significância 0,10 e quando o critério de determinação da defasagem máxima é l_4 . Quanto ao valor de ϕ , os melhores resultados foram encontrados para o valor de 0,55, e os piores para o valor 0,95.

Em relação às interações, há evidências de sua presença entre o nível de significância adotado e diferentes tipos de modelos, conforme indicado na Tabela 4.13. Conforme pode ser visto na Figura 4.24, a proporção de rejeição da hipótese nula, quando o nível de 0,10 é adotado, apresenta maior redução quando o modelo é alterado de N para CT do que quando o nível de significância é 0,05. Cabe ressaltar que embora a taxa de rejeição de H_0 seja mais próxima entre os níveis 0,05 e 0,10 para o modelo CT, os valores de 0,10 sempre apresentaram valores superiores.

Também há evidência de interação entre o tipo de modelo e critério para determinação da

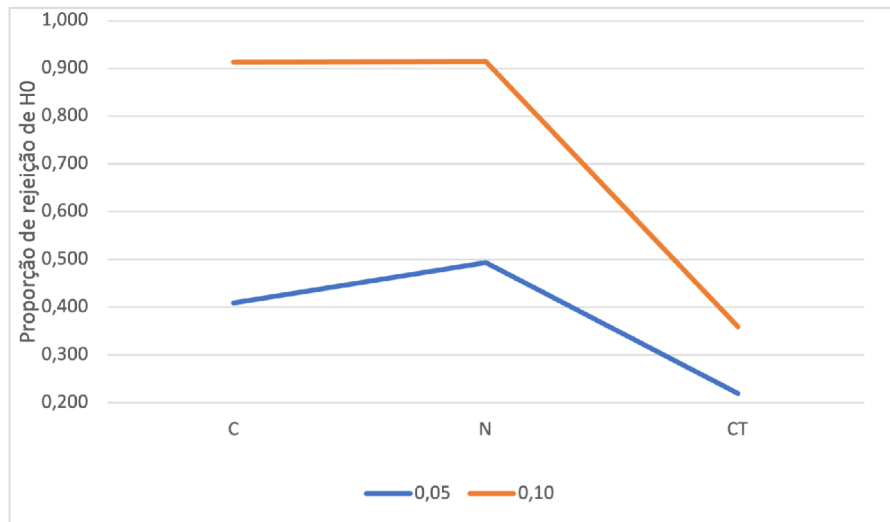


Figura 4.24: Interações entre nível de significância e tipo de modelo considerando a série estacionária e $T = 25$

defasagem máxima. Conforme pode ser visto na Figura 4.25, a proporção de rejeição da hipótese nula, quando o modelo é alterado de C para N apresenta um crescimento quando é escolhido o l_{12} , ao passo que quando é adotado o l_4 este valor se mantém praticamente constante. Além disso, embora haja um decréscimo na taxa de rejeição de H_0 quando o modelo é alterado de N para CT para ambos os critérios, este é muito mais acentuado para l_4

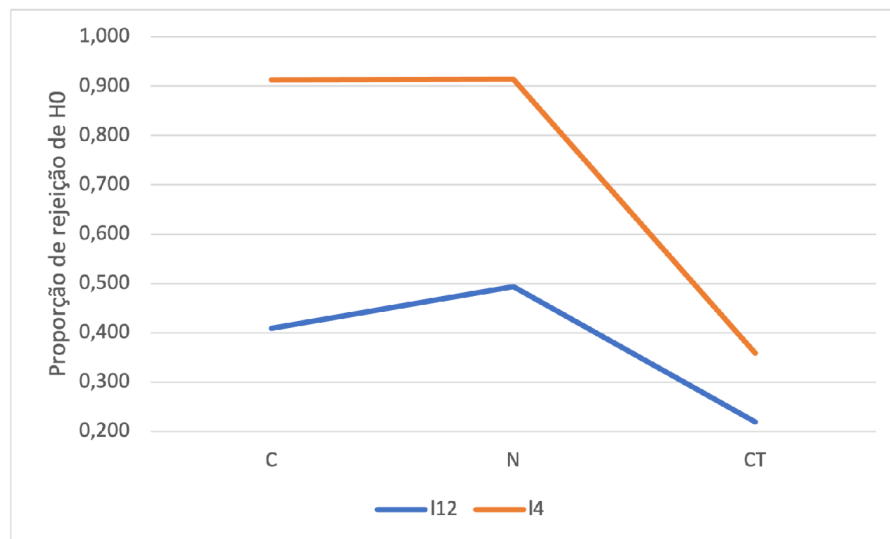


Figura 4.25: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 25$

Já em relação à interação entre k_{max} e nível de significância, esta se mostrou significativa. Conforme Figura 4.26, o crescimento da proporção de não rejeição da hipótese nula de l_{12} para l_4 é maior quando usado o nível de significância 0,05 do que quando usado 0,10, embora esta proporção seja sempre maior com uso de 0,10.

Já o tipo de estrutura do modelo também interage com valor de ϕ . De acordo com a Figura 4.27, quando o modelo é alterado de N para CT há uma redução na proporção de rejeição de

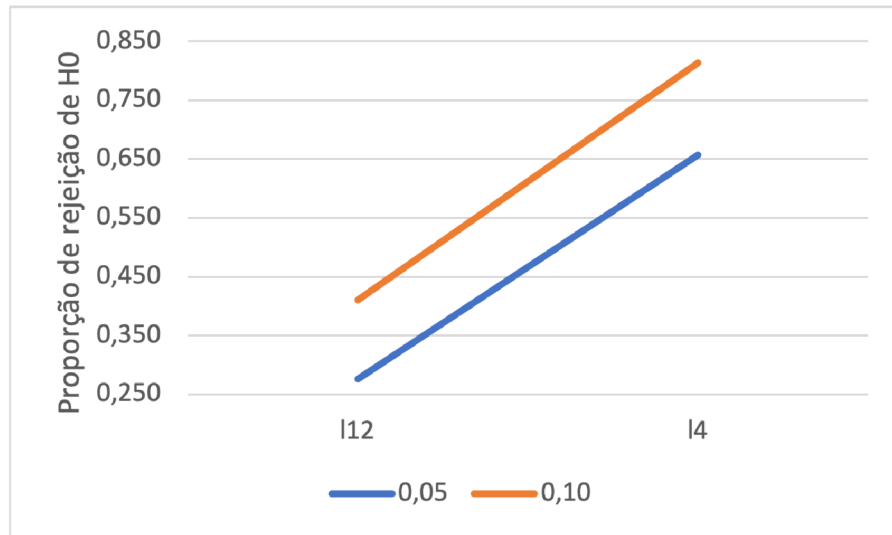


Figura 4.26: Interações entre tipo de modelo e nível de significância considerando a presença de raiz unitária e $T = 25$

H_0 , porém esta é mais acentuada para $\phi = 0,55$. Quando o modelo é alterado de C para N, há um crescimento na taxa de rejeição quando ϕ assume os valores 0,55 e 0,75, entretanto, é observada uma redução na taxa de rejeição de H_0 para 0,95. Assim, quando ϕ é 0,95, o teste obteve melhores resultados apenas para a estrutura C.

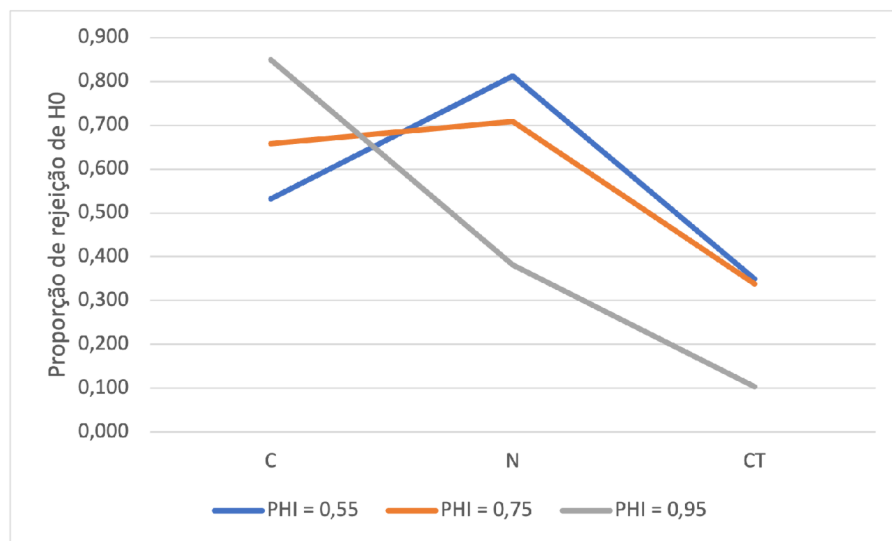


Figura 4.27: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 25$

Por fim, há indícios de interação entre os fatores k_{max} e ϕ , que podem ser observados com o auxílio da Figura 4.28. É possível observar que quando o valor de ϕ é alterado de 0,55 para 0,75 há um crescimento na taxa de rejeição de H_0 quando considerado o critério l_4 e uma redução quando considerado o l_{12} . Ao alterar o valor de ϕ de 0,75 para 0,95, os valores da taxa se aproximam, com decréscimo do valor quando utilizado l_4 e acréscimo quando adotado o l_{12} . Embora o comportamento dos critérios se apresente diferente conforme é alterada a estrutura do modelo, os melhores resultados são encontrados para l_4 em todos os casos.

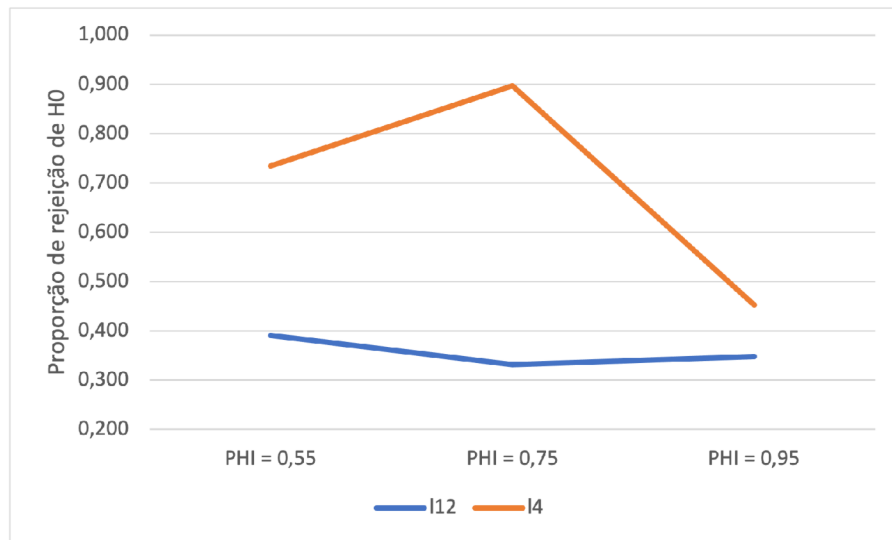


Figura 4.28: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 25$

ii. T=50

Na Tabela 4.14 são apresentados a média, a mediana e o desvio padrão da taxa de rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, enquanto nos gráficos da Figura 4.29, box plots que informam a mediana, os quartis e valores máximo e mínimo, além da presença de valor fora do padrão e de assimetria.

Tabela 4.14: Média e o desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 50$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,71707	0,75105	0,24738
	BIC	0,71654	0,76200	0,27025
	HQC	0,72000	0,76230	0,25787
Defasagem Máxima	l12	0,59193	0,58140	0,26819
	l4	0,84381	0,91025	0,16828
Estrutura do modelo	C	0,83831	0,95105	0,21024
	N	0,76876	0,87995	0,24893
	CT	0,54654	0,62160	0,21632
Nível de Significância	0,05	0,66646	0,66465	0,27612
	0,10	0,76928	0,82235	0,22568
Φ	0,55	0,80585	0,83530	0,16753
	0,75	0,71137	0,77835	0,27022
	0,95	0,63639	0,63855	0,29123

De acordo com as informações apresentadas na Tabela 4.14, na amostra analisada, o melhor desempenho do teste ADF, por apresentar maior tendência central da taxa de rejeição da hipótese nula, ocorre em modelos C, selecionados pelo HQC, com determinação da defasagem máxima pelo l_4 , utilização do nível de significância 0,10 e valor de ϕ 0,55. Observa-se também que a menor variabilidade da taxa ocorre quando o valor de ϕ é 0,55, enquanto a maior variabilidade ocorre em $\phi = 0,95$.

A Figura 4.29 apresenta valores, em geral, mais próximos de um para a proporção de rejeição da hipótese nula, quando é feita a comparação com a série de tamanho menor. Em relação ao gráfico a, há maior variabilidade em todos os fatores, ao comparar com a série de $T=25$, porém o comportamento dos fatores é semelhante ao das séries de tamanho inferior. Além disso, neste caso, a mediana do critério de informação AIC se apresentou levemente inferior a dos demais critérios, cabendo lembrar que a maior proporção média foi encontrada no critério HQC. Em relação ao k_{max} , além de terem sido encontrados resultados melhores com o uso de l_4 , houve aumento na variabilidade de l_{12} e redução na variabilidade de l_4 , realizando a comparação com a série de tamanho menor. Em nenhum dos gráficos foi encontrada a presença de valores fora do padrão. Quanto ao valor de ϕ , os achados são semelhantes aos encontrados para $T=25$, porém com os valores mais próximos do máximo.

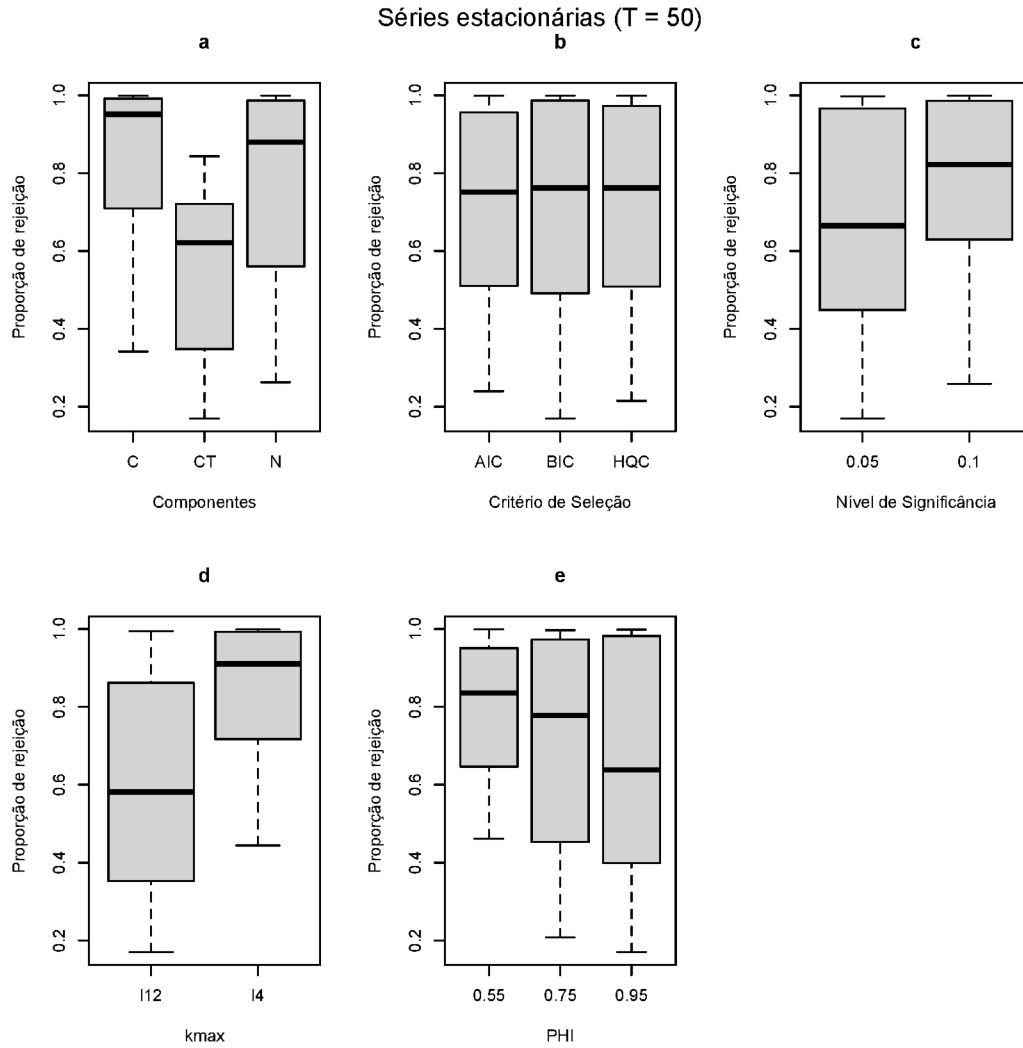


Figura 4.29: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 50$

Da mesma maneira que nas séries anteriores, o efeito de cada fator e respectivas interações são analisados por meio da modelagem GAMLSS, em que os resultados são apresentados na Tabela 4.15, seguida do wormplot que pode ser visualizado na Figura 4.30. Nesta análise foi tomado como referência o critério de informação HQC, modelo de estrutura N, nível de significância 0.05, critério de determinação da defasagem máxima l_{12} e ϕ 0,75.

Ao observar o gráfico da Figura 4.30, nota-se que todos os pontos estão dentro da banda de confiança, indicando que o modelo está bem ajustado. Diante disso, parte-se para a interpretação dos resultados presentes na Tabela 4.15.

Os resultados apresentados na Tabela 4.15 indicam não haver evidências de melhor desempenho do teste ADF quando o critério de informação é alterado, assim como ocorreu nas séries de tamanhos inferiores. A estrutura do modelo se mostrou significativa na comparação entre os modelos N e CT, sendo que o modelo N apresentou melhores resultados. Em relação ao nível de significância e ao k_{max} , ambos apresentaram diferença significativa entre seus níveis, nível de significância de 0.10 e critério de determinação da defasagem máxima l_4 . Quanto ao valor

Tabela 4.15: Análise do efeito principal dos fatores e respectivas interações na taxa rejeição da hipótese nula em experimentos com séries estacionárias e $T = 50$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	0,51085	0,25782	0,05131
criterioAIC	-0,08042	0,28655	0,77977
criterioHQC	-0,01068	0,28655	0,97038
modeloC	-0,18068	0,28655	0,53031
modeloCT	-2,15662	0,28655	0,00000
nivelsig0,1	1,11213	0,25012	0,00003
kmaxl4	3,28685	0,25012	0,00000
phi0,55	2,46157	0,28655	0,00000
phi0,95	-1,59275	0,28655	0,00000
criterioAIC.modeloC	-0,49884	0,26529	0,06405
criterioHQC.modeloC	-0,30026	0,26529	0,26142
criterioAIC.modeloCT	0,39001	0,26529	0,14583
criterioHQC.modeloCT	0,19781	0,26529	0,45828
criterioAIC.nivelsig0.1	-0,08117	0,21661	0,70894
criterioHQC.nivelsig0.1	-0,03359	0,21661	0,87719
criterioAIC.kmaxl4	-0,04621	0,21661	0,83167
criterioHQC.kmaxl4	-0,04982	0,21661	0,81872
modeloC.nivelsig0.1	-0,38633	0,21661	0,07866
modeloCT.nivelsig0.1	-0,47585	0,21661	0,03121
modeloC.kmaxl4	0,41282	0,21661	0,06061
modeloCT.kmaxl4	-0,62134	0,21661	0,00539
nivelsig0.1.kmaxl4	0,27264	0,17686	0,12750
modeloC.phi0.55	-2,26886	0,26529	0,00000
modeloCT.phi0.55	-1,27147	0,26529	0,00001
modeloC.phi0.95	5,30081	0,26529	0,00000
modeloCT.phi0.95	2,50248	0,26529	0,00000
nivelsig0.1.phi0.55	0,05063	0,21661	0,81583
nivelsig0.1.phi0.95	-0,29623	0,21661	0,17563
criterioAIC.phi0.55	0,08470	0,26529	0,75044
criterioHQC.phi0.55	-0,01054	0,26529	0,96842
criterioAIC.phi0.95	-0,26520	0,26529	0,32078
criterioHQC.phi0.95	-0,15928	0,26529	0,55010
kmaxl4.phi0.55	-1,55984	0,26529	0,00000
kmaxl4.phi0.95	-2,00686	0,26529	0,00000

de ϕ , os melhores resultados foram encontrados para o valor de 0,55, assim como nos casos das séries de comprimentos inferiores.

Em relação às interações, há evidências de sua presença entre o nível de significância adotado e diferentes tipos de modelos, conforme indicado na Tabela 4.15. Conforme pode ser visto na Figura 4.31, a proporção de rejeição da hipótese nula, quando o nível de 0,05 é adotado, apresenta maior redução quando o modelo é alterado de C para N e de N para CT do que quando o nível de significância é 0,10. Assim como nos casos anteriores, o desempenho do teste

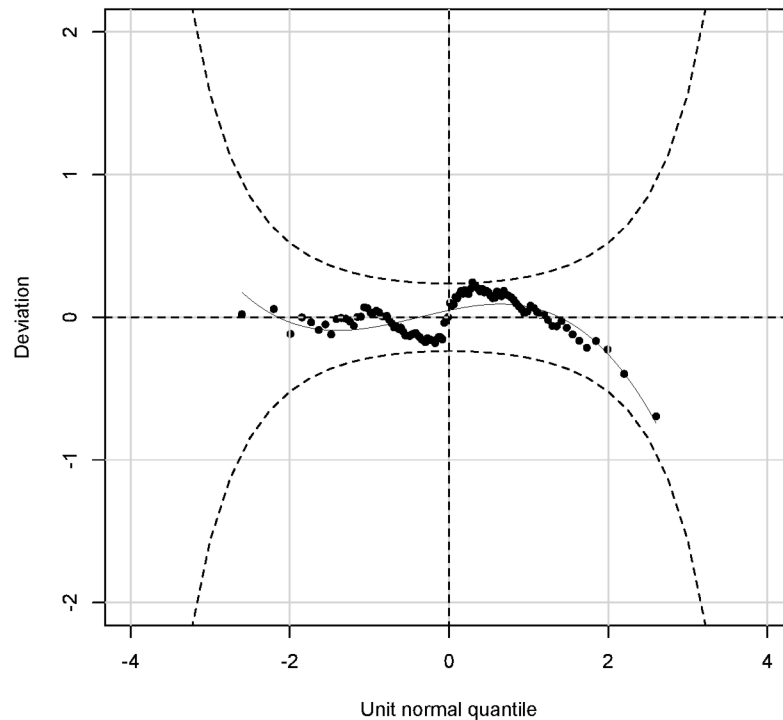


Figura 4.30: *Worm plot* gerado para o modelo GAMLSS considerando a série estacionária e $T = 50$

foi superior quando adotado o nível de significância de 0,10.

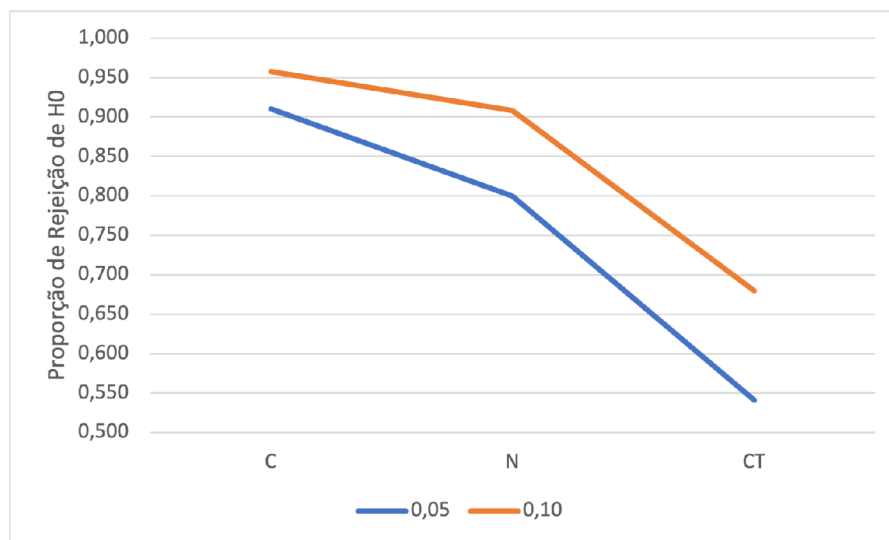


Figura 4.31: Interações entre nível de significância e tipo de modelo considerando a série estacionária e $T = 50$

Também há evidência de interação entre o tipo de modelo e critério para determinação da defasagem máxima. Conforme pode ser visto na Figura 4.32, a proporção de rejeição da hipótese nula, quando o modelo é alterado de C para N apresenta um decréscimo maior quando

é escolhido o l_4 , ao passo que quando é adotado o l_{12} este valor apresenta crescimento. Além disso, embora haja um decréscimo na taxa de rejeição de H_0 quando o modelo é alterado de N para CT para ambos os critérios, este é muito mais acentuado para l_{12} . Cabe ressaltar que o desempenho do teste foi superior para todas as estruturas de modelo, quando adotado o l_4 .

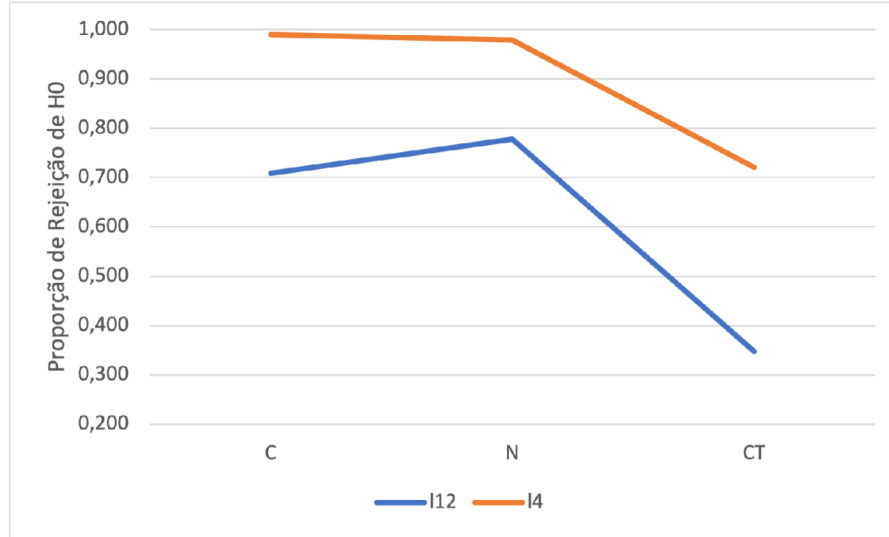


Figura 4.32: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 50$

Já o tipo de estrutura do modelo também interage com valor de ϕ . De acordo com a Figura 4.33, quando o modelo é alterado de C para N há uma redução na proporção de rejeição de H_0 quando ϕ é 0,95, que se mantém praticamente constante quando o modelo é alterado de N para CT. Quando o valor de ϕ é 0,55 ou 0,75, é encontrado comportamento semelhante. Diante disso, quando ϕ é 0,95, o teste obteve melhores resultados apenas para a estrutura C, enquanto para o valor de $\phi = 0,55$ o teste apresentou melhor desempenho em todos os modelos, quando comparado ao valor de 0,75.

Por fim, há indícios de interação entre os fatores k_{max} e ϕ , que podem ser observados com o auxílio da Figura 4.34. Neste caso, os achados são semelhantes aos encontrados na série de tamanho inferior. Quando usado l_4 , o melhor desempenho do teste ADF ocorre quando $\phi = 0,75$, enquanto que quando usado l_{12} , o melhor desempenho ocorre quando $\phi = 0,55$.

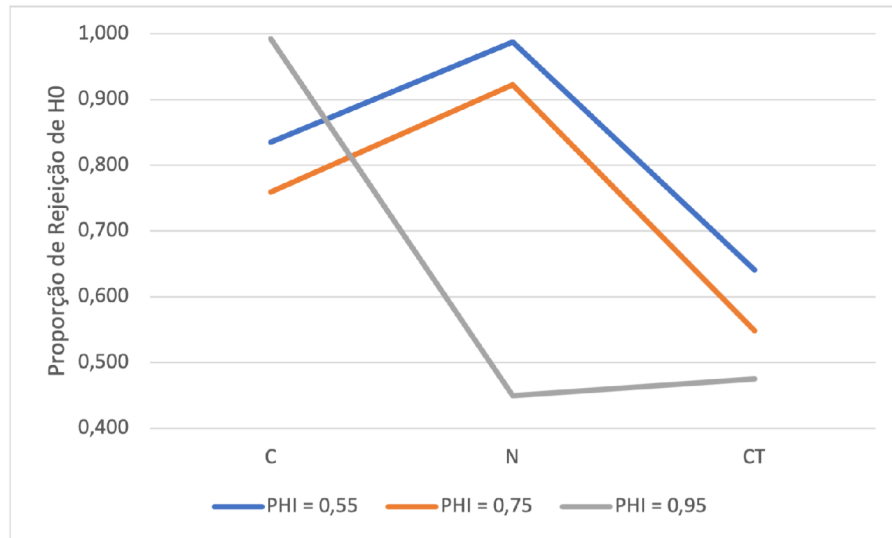


Figura 4.33: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 50$

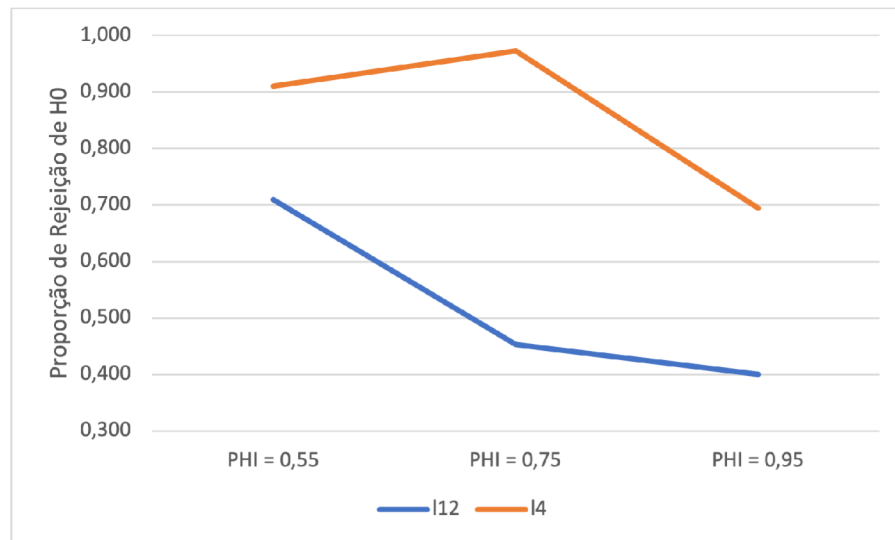


Figura 4.34: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 50$

iii. T=100

Na Tabela 4.16 são apresentados a média, a mediana e o desvio padrão da taxa de rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, enquanto nos gráficos da Figura 4.35, box plots que informam a mediana, os quartis e valores máximo e mínimo, além da presença de valor fora do padrão e de assimetria.

Tabela 4.16: Média e o desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 100$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,92022	0,98060	0,13743
	BIC	0,93276	0,99820	0,13936
	HQC	0,93047	0,99345	0,13727
Defasagem Máxima	l_2	0,89283	0,97380	0,16182
	l_4	0,96280	0,99945	0,09551
Estrutura do modelo	C	0,97683	0,99865	0,04807
	N	0,87236	0,99550	0,19573
	CT	0,93425	0,97565	0,10461
Nível de Significância	0,05	0,90540	0,98630	0,16531
	0,10	0,95023	0,99560	0,09722
Φ	0,55	0,98764	0,99610	0,01886
	0,75	0,92673	0,98740	0,10989
	0,95	0,86907	0,98895	0,19385

De acordo com as informações apresentadas na Tabela 4.16, na amostra analisada, o melhor desempenho do teste ADF, por apresentar maior taxa de rejeição da hipótese nula, ocorre em modelos C, selecionados pelo BIC, com determinação da defasagem máxima pelo l_4 , utilização do nível de significância 0,10 e valor de ϕ 0,55. Observa-se também que a menor variabilidade da taxa ocorre quando o valor de ϕ é 0,55, enquanto a maior variabilidade ocorre para os modelos N.

A Figura 4.35 apresenta, em todos os gráficos, valores fora do padrão, o que não ocorreu nas séries de comprimentos menores. Além disso, em todos os casos o intervalo interquartil se apresentou menor do que nas séries anteriores, indicando uma maior estabilidade nos resultados, e os valores, em geral, mais próximos de um (valor ideal), indicando um desempenho melhor do teste na identificação da estacionariedade. No gráfico a, destaca-se a maior variabilidade do modelo N; no gráfico b, o AIC apresentou maior variabilidade e menor mediana, embora esta superioridade não seja muito elevada; no gráfico c, o nível 0,10 se mostrou com menor variabilidade e maior mediana, além de valores fora do padrão não tão extremos; no gráfico d, melhores resultados são encontrados para o critério l_4 que apresenta medida de tendência central mais elevada e menor variabilidade. Por fim, no gráfico e, a maior variabilidade e os menores valores de proporção são encontrados quando o valor de ϕ é 0,95.

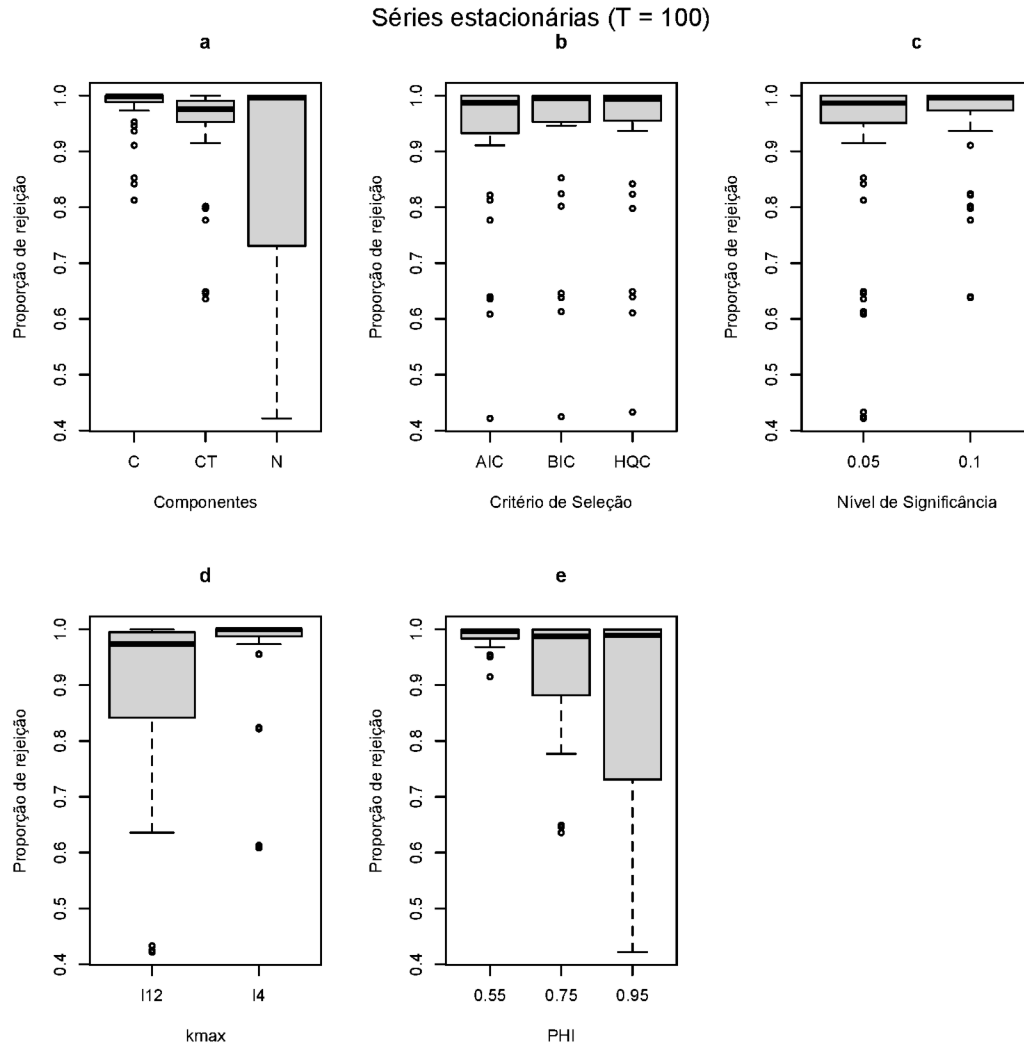


Figura 4.35: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 100$

Da mesma maneira em que foi realizado nas séries anteriores, o efeito de cada fator e respectivas interações são analisados por meio da modelagem GAMLSS, em que os resultados são apresentados na Tabela 4.17, seguida do wormplot que pode ser visualizado na Figura 4.36. Nesta análise foi tomado como referência o critério de informação HQC, modelo de estrutura N, nível de significância 0.05, critério de determinação da defasagem máxima l_{12} e ϕ 0.95.

Ao observar o gráfico da Figura 4.36, nota-se que existem poucos pontos fora da banda de confiança, o que pode ser devido a presença de valores fora do padrão, conforme visto na Figura 4.35. Como a grande maioria dos valores está dentro da banda de confiança, há indícios de que o modelo está bem ajustado. Diante disso, parte-se para a interpretação dos resultados presentes na Tabela 4.17.

Os resultados apresentados na Tabela 4.17 indicam não haver evidências de variação no desempenho do teste ADF quando o critério de informação é alterado de HQC para BIC, porém é significativa a diferença entre os critérios AIC e HQC, sendo que este último apresenta melhor desempenho. A estrutura do modelo se mostrou significativa para todos os tipos testados,

Tabela 4.17: Análise do efeito principal dos fatores e respectivas interações na taxa rejeição da hipótese nula em experimentos com séries estacionárias e $T = 100$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	-0,47169	0,42413	0,26972
criterioAIC	-1,15675	0,47139	0,01652
criterioBIC	0,57552	0,47139	0,22605
modeloC	6,70178	0,47139	0,00000
modeloCT	-4,81451	0,47139	0,00000
nivelsig0,1	0,67416	0,41146	0,10563
kmaxl4	1,62694	0,41146	0,00018
phi0,55	7,42314	0,47139	0,00000
phi0,75	-4,81111	0,47139	0,00000
criterioAIC.modeloC	-0,10254	0,43642	0,81489
criterioBIC.modeloC	0,46047	0,43642	0,29485
criterioAIC.modeloCT	0,19876	0,43642	0,65015
criterioBIC.modeloCT	0,13566	0,43642	0,75680
criterioAIC.nivelsig0.1	0,23015	0,35634	0,52038
criterioBIC.nivelsig0.1	0,25858	0,35634	0,47036
criterioAIC.kmaxl4	0,45215	0,35634	0,20851
criterioBIC.kmaxl4	-0,61243	0,35634	0,08991
modeloC.nivelsig0.1	-0,09187	0,35634	0,79728
modeloCT.nivelsig0.1	-0,17814	0,35634	0,61864
modeloC.kmaxl4	1,17441	0,35634	0,00152
modeloCT.kmaxl4	0,02559	0,35634	0,94295
nivelsig0.1.kmaxl4	0,10426	0,29095	0,72111
modeloC.phi0.55	-9,82442	0,43642	0,00000
modeloCT.phi0.55	8,74524	0,43642	0,00000
modeloC.phi0.75	-8,73011	0,43642	0,00000
modeloCT.phi0.75	-9,20873	0,43642	0,00000
nivelsig0.1.phi0.55	0,11473	0,35634	0,74839
nivelsig0.1.phi0.75	-0,28715	0,35634	0,42296
criterioAIC.phi0.55	-0,04117	0,43642	0,92510
criterioBIC.phi0.55	0,66277	0,43642	0,60111
criterioAIC.phi0.75	0,22916	0,43642	0,19857
criterioBIC.phi0.75	0,43361	0,43642	0,32372
kmaxl4.phi0.55	-0,25318	0,35634	0,47965
kmaxl4.phi0.75	-2,23147	0,35634	0,00000

havendo um melhor desempenho do teste em modelos com estrutura C. Em relação ao nível de significância e ao k_{max} , os melhores resultados foram encontrados para 0,10 e l_4 , respectivamente. Quanto ao valor de ϕ , o teste apresentou desempenho superior quando o valor de ϕ foi considerado 0,55, assim como nos casos das séries de comprimentos inferiores.

Foram encontradas evidências de interação entre o tipo de modelo e critério para determinação da defasagem máxima. Conforme pode ser visto na Figura 4.37, a proporção de rejeição da hipótese nula, quando o modelo é alterado de C para CT apresenta um decréscimo

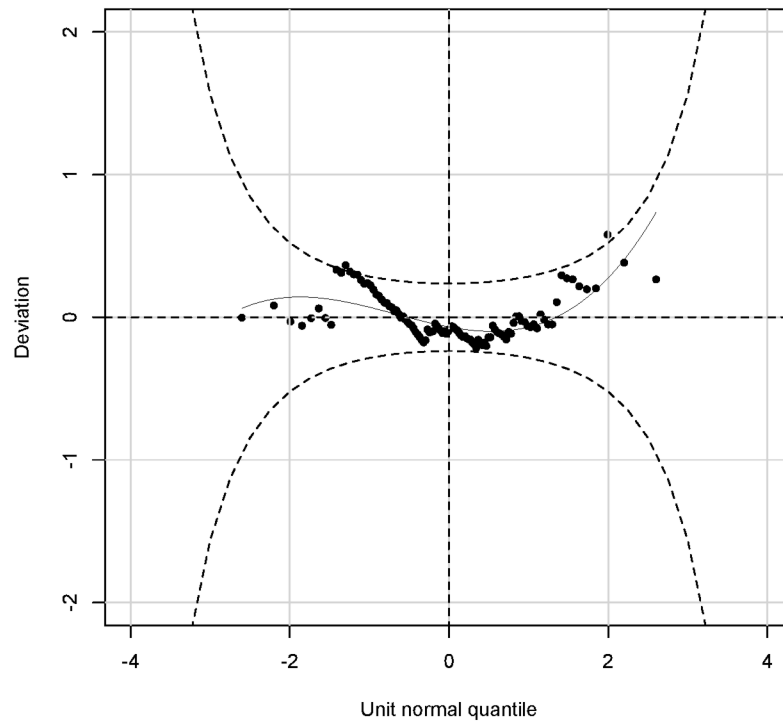


Figura 4.36: *Worm plot* gerado para o modelo GAMLSS considerando a série estacionária e $T = 100$

muito maior quando é escolhido o l_{12} do que quando é adotado o l_4 . Cabe ressaltar que o desempenho do teste foi superior para todas as estruturas de modelo quando adotado o l_4 .

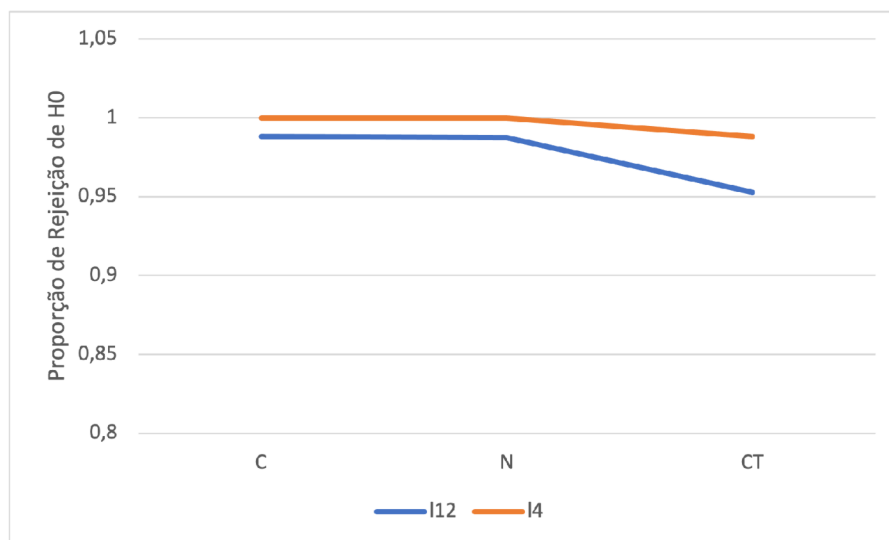


Figura 4.37: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 100$

Já o tipo de estrutura do modelo também interage com valor de ϕ . De acordo com a Figura 4.38, quando o modelo é alterado de CT para N há uma grande redução na proporção de

rejeição de H_0 quando ϕ é 0,95, ao passo que para os valores menores ϕ há um crescimento na proporção de rejeição da hipótese nula, nas mesmas condições. Quando a estrutura é alterada de C para CT, as séries com ϕ 0,75(0,55) e 0,95 apresentaram comportamento semelhante, ao passo que a série com $\phi = 0,55(0,75)$ mostrou uma queda maior na taxa de rejeição de H_0 .

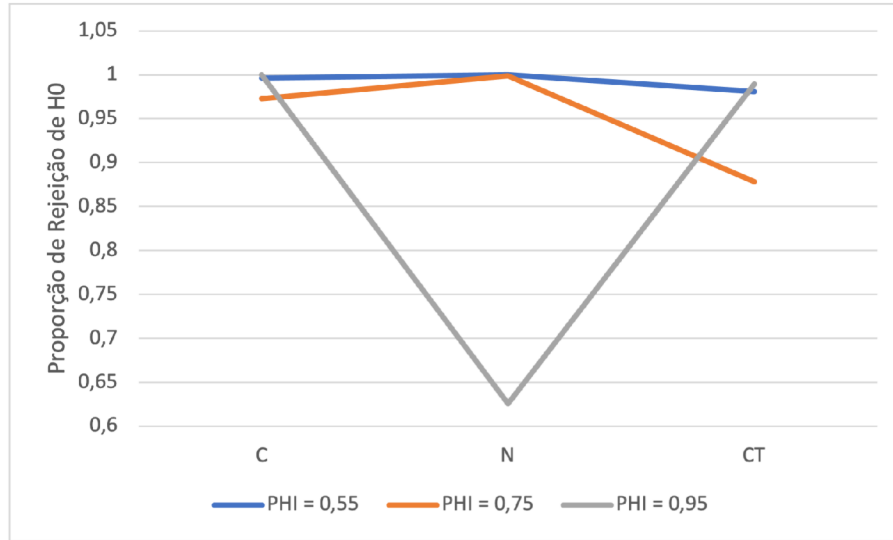


Figura 4.38: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 100$

Por fim, há indícios de interação entre os fatores k_{max} e ϕ , que podem ser observados com o auxílio da Figura 4.39. Neste caso, para l_4 , a taxa de rejeição se mantém praticamente constante, próxima a unidade. Enquanto para l_{12} quando o valor de ϕ é alterado de 0,55 para 0,75 há um decréscimo na taxa de rejeição de H_0 , que volta a subir ao trocar o valor de ϕ de 0,75 para 0,95. Além disso, assim como ocorreu nas séries de tamanhos inferiores, o critério l_4 apresentou desempenho superior, independente do valor do coeficiente ϕ .

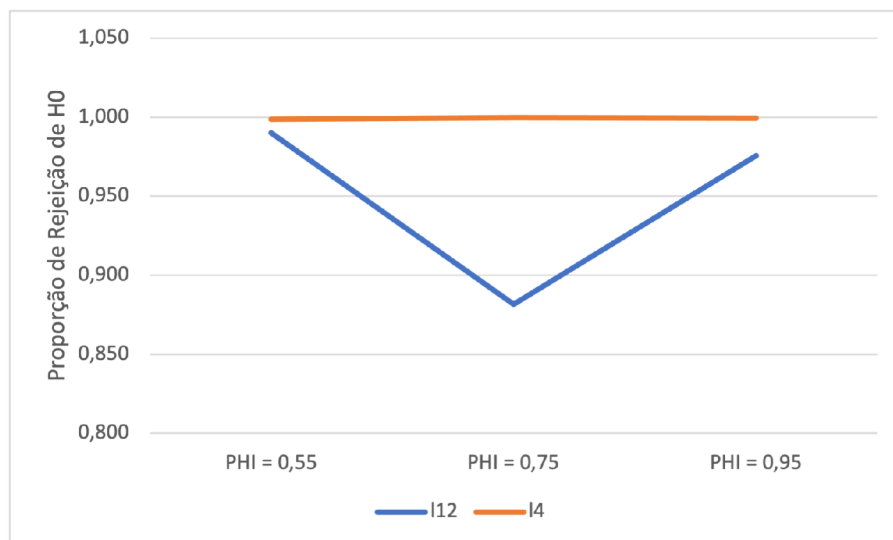


Figura 4.39: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 100$

iv. Considerações finais sobre as séries estacionárias

Os resultados dos experimentos para as séries estacionárias com os diferentes tamanhos também se mostraram bastantes semelhantes em relação à influência dos fatores na não rejeição da hipótese nula do teste ADF. A Tabela 4.18 apresenta os fatores associados ao melhor desempenho do teste para cada comprimento de série.

Tabela 4.18: Resumo dos resultados para séries estacionárias

Comprimento da série	Critério de Informação	Defasagem Máxima	Estrutura do Modelo	Nível de significância	ϕ
25	Indiferente	l_4	C	0,10	0,55
50	Indiferente	l_4	C	0,10	0,55
100	BIC/HQC	l_4	C	Indiferente	0,55

Para todos os experimentos, quando a série é estacionária, o critério de informação não se mostrou como um fator significativo no resultado do teste. O melhor desempenho foi encontrado quando a defasagem máxima foi determinada por l_4 e o nível de significância adotado foi 0,10. Além disso, o modelo que apresentou maior proporção de não rejeição de H_0 é o que apresenta estrutura tipo C e os melhores resultados foram encontrados para $\phi = 0,55$, sendo este último resultado já esperado.

Em relação às interações, apesar de em diversos experimentos se apresentarem significativas, somente no caso da interação entre a estrutura do modelo e o valor de ϕ foi encontrada alteração no resultado do teste. Nestes casos, independente do comprimento da série, foram encontrados melhores resultados quando $\phi=0,95$ apenas nos modelos C. Quando o modelo não possui constante e tendência, ou quando possui ambas, e o coeficiente é 0,95, os resultados são inferiores aos demais.

No que se refere às médias, todos os fatores apresentaram comportamento semelhante em relação ao comprimento das séries, apresentando crescimento na proporção de rejeição de H_0 conforme aumenta o comprimento da série. Quanto ao critério de seleção da defasagem máxima, o l_{12} se aproxima do l_4 conforme aumenta o comprimento da série, assim como aconteceu nos casos de séries não estacionárias. O modelo CT apresentou maior crescimento na proporção de não rejeição de H_0 ao aumentar o comprimento da série, alcançando desempenho melhor que o modelo N para as séries de T=100. No que tange ao nível de significância, é possível perceber que os resultados de quando é adotado 0,05 se aproxima dos resultados de quando se escolhe 0,10 conforme aumenta o tamanho da série.

Para séries de tamanho 25, a taxa média de acerto no teste para $\phi = 0,95$ foi de aproximadamente 0,44, em comparação ao valor aproximado de 0,55 para os outros valores de ϕ . Quando o comprimento da série foi igual a 50, o valor da taxa de acerto para $\phi = 0,95$ subiu para 0,63 em comparação ao 0,71 para $\phi = 0,75$ e 0,80 para $\phi = 0,55$. Já para T=100, foram obtidas as médias aproximadas de 0,86, 0,92 e 0,98 para os valores de $\phi = 0,95$, 0,75 e 0,55, respectivamente. Aqui, caberia investigar se o comprimento da série seria um fator tão relevante quanto

o valor de ϕ na tomada de decisão do teste ADF, uma vez que quando $T=100$ os resultados foram melhores para $\phi = 0,95$, do que para $\phi = 0,55$, quando $T=25$ e $T=50$.

Por fim, cabe salientar que, assim como ocorreu nas séries não estacionárias, na série de comprimento igual a 100 houve o menor número de interações significativas, reforçando que cabe investigar se para séries maiores existe menor influência da interação entre os fatores no resultado do teste.

4.3 Análise dos resultados das simulações complementares

i. Séries não estacionárias

Assim como nos casos anteriores, na Tabela 4.19 são apresentadas medidas descritivas da taxa de não rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento que utilizou séries temporais com $T = 40$, bem como os gráficos de caixa na Figura 4.40.

Tabela 4.19: Média, mediana e o desvio padrão da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 40$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,93977	0,96300	0,05739
	BIC	0,95129	0,97960	0,05313
	HQC	0,94379	0,96810	0,05521
Defasagem Máxima	l12	0,93716	0,96300	0,05273
	l4	0,95274	0,98265	0,05540
Estrutura do modelo	C	0,97253	0,97510	0,01378
	N	0,87774	0,87535	0,03799
	CT	0,98458	0,98885	0,01165
Nível de Significância	0,05	0,96037	0,97865	0,03566
	0,10	0,92953	0,96395	0,06483

As medidas apresentadas na Tabela 4.19 levam às mesmas conclusões encontradas para o caso de $T=25$ e $T = 50$, indicando o mesmo padrão de comportamento na amostra analisada em relação aos níveis dos fatores que ocorrem com maior proporção de não rejeição da hipótese nula, assim como aqueles em que ocorreram a maior e a menor variabilidade.

Em relação ao gráfico a da Figura 4.40, os resultados encontrados também são parecidos com os que foram para o caso em que $T=25$, o que ocorre também para $T = 50$ de maneira mais acentuada. Entretanto, cabe salientar, que na amostra de séries com $T=40$, houve menor dispersão para as proporções obtidas em modelos tipo C e CT e maior diferença dos resultados encontrados para estes tipos de modelo em relação ao modelo N, conforme pode ser visualizado no gráfico a desta figura. Nos gráficos b, c e d pode ser observada assimetria bem acentuada com maior concentração dos valores maiores, o que aconteceu apenas para o critério de defasagem máxima em séries com $T = 25$, embora estas proporções para os dois critérios de determinação da defasagem máxima tenham ficado bem mais próximas nas séries com $T = 40$ (gráfico d). De maneira geral, o comportamento dos resultados na amostra de séries com $T = 40$ apresenta-se em uma situação intermediária entre o comportamento dos resultados obtidos na amostra de séries com $T = 25$ e na amostra de séries com $T = 50$, sendo mais próxima desta última.

Em seguida foi gerado o modelo através da modelagem GAMLSS, em que a resposta obtida pode ser visualizada na Tabela 4.20, com o respectivo wormplot para a verificação dos resíduos

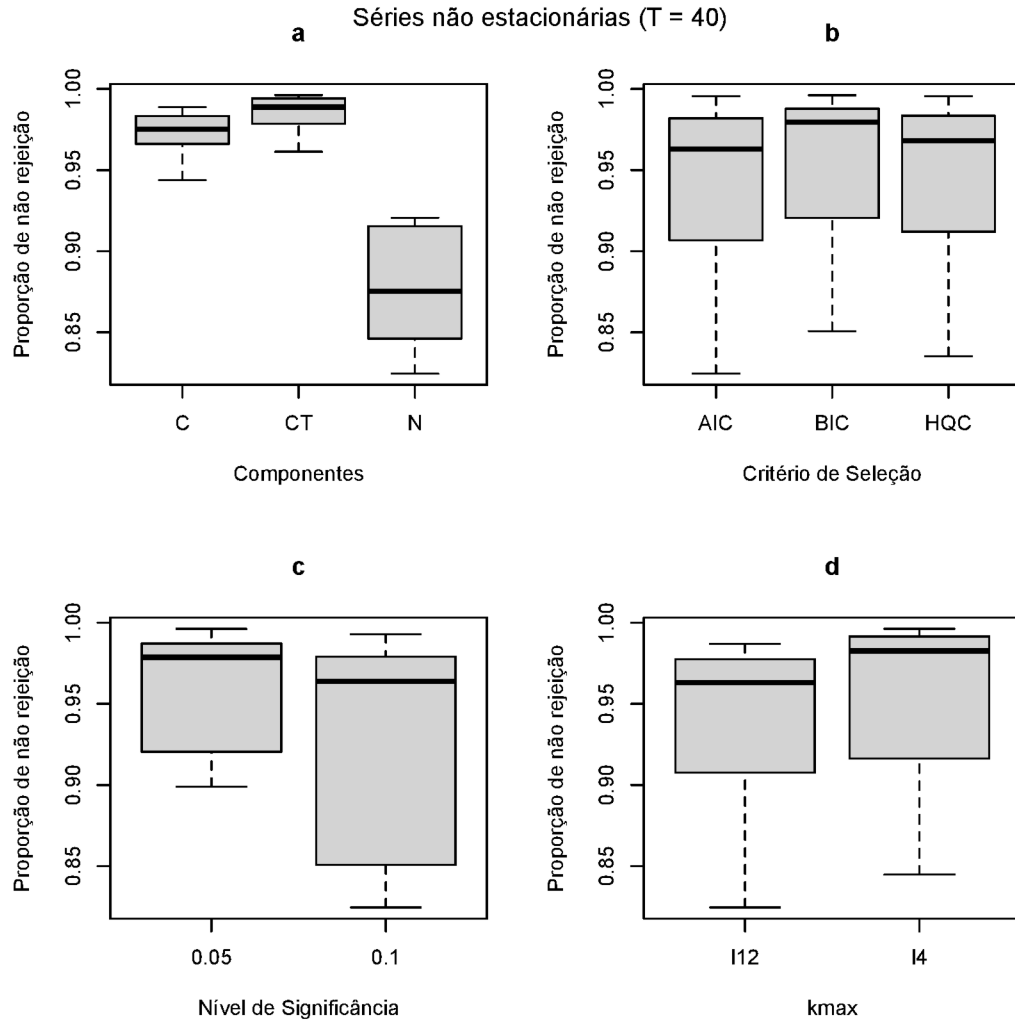


Figura 4.40: Box plots da taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 40$

do modelo, que pode ser visualizado na Figura 4.41. Nesta análise é tomado como referência os mesmos níveis dos fatores considerados nos experimentos para $T = 25$ e para $T = 50$, ou seja: critério de informação HQC, modelo de estrutura tipo C, nível de significância 0.05 e critério de determinação da defasagem máxima l_{12} .

Ao observar o gráfico da Figura 4.41, verifica-se que não há pontos fora da banda de confiança, assim como ocorreu em $T=25$ e em $T = 50$. Diante disso, parte-se para a interpretação dos resultados presentes na Tabela 4.20.

Ao analisar o efeito principal dos fatores, observa-se que os resultados apresentados na Tabela 4.20 se assemelham aos encontrados para o caso de $T=25$, ou seja, todos os fatores interferem na proporção de não rejeição da hipótese nula. Exceto que para o experimento atual, a comparação entre HQC e AIC se apresentou significativa, o que não aconteceu para $T = 25$, com evidências de que o HQC apresentou desempenho superior ao AIC. No caso em que $T = 50$, a diferença entre HQC e AIC não se apresentou significativa.

Já em relação às interações, a exemplo do que ocorreu para $T = 25$ e $T = 50$, há algumas

Tabela 4.20: Análise do efeito principal dos fatores e respectivas interações na taxa de não rejeição da hipótese nula em experimentos com séries não estacionárias e $T = 40$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	3,48712	0,02597	0,00000
criterioAIC	-0,14907	0,03180	0,00029
criterioBIC	0,39294	0,03180	0,00000
modeloN	-1,24518	0,03180	0,00000
modeloCT	0,41226	0,03180	0,00000
nivelsig0,1	-0,49948	0,02845	0,00000
kmaxl4	0,83732	0,02845	0,00000
criterioAIC.modeloN	0,05976	0,03484	0,10688
criterioBIC.modeloN	-0,18650	0,03484	0,00008
criterioAIC.modeloCT	0,02581	0,03484	0,47015
criterioBIC.modeloCT	0,04980	0,03484	0,17337
criterioAIC.nivelsig0.1	-0,01605	0,02845	0,58098
criterioBIC.nivelsig0.1	0,00458	0,02845	0,87428
criterioAIC.kmaxl4	0,08468	0,02845	0,00941
criterioBIC.kmaxl4	-0,23208	0,02845	0,00000
modeloN.nivelsig0.1	-0,11545	0,02845	0,00103
modeloCT.nivelsig0.1	-0,02193	0,02845	0,45275
modeloN.kmaxl4	-0,63985	0,02845	0,00000
modeloCT.kmaxl4	0,65489	0,02845	0,00000
nivelsig0.1.kmaxl4	-0,11639	0,02323	0,00015

significativas. Tal fato ocorreu, por exemplo, entre critério de informação e o tipo de estrutura modelo. Com o auxílio da Figura 4.42, é possível observar que embora o critério BIC apresente melhor desempenho em todos os casos, há um crescimento maior dessa proporção quando o modelo possui a estrutura tipo N.

A Tabela 4.20 também apresentou indícios de que há interação significativa entre critério de informação e de seleção da defasagem máxima. Neste caso, é possível observar na Figura 4.43 que, ao utilizar l_{12} há um crescimento na proporção de não rejeição de H_0 quando o critério de informação é alterado de AIC para HQC, acontecendo o mesmo quando há a troca de HQC para BIC, o mesmo não ocorrendo para l_4 , em que a troca entre os critérios não implica em grandes alterações na proporção de rejeição de H_0 . Entretanto, salienta-se que o uso de l_4 sempre apresenta proporção de não rejeição de H_0 maior do que quando usado o l_{12} , semelhantemente ao que ocorreu para $T = 25$ e $T = 50$.

Já a interação entre o tipo de estrutura do modelo e o nível de significância, apresentou evidências de ser significativa. Na Figura 4.44, é possível observar que para o nível de significância de 0,10, há um crescimento maior da proporção de não rejeição de H_0 quando o modelo é alterado de N para C, cabendo ressaltar que para 0,05 as taxas de não rejeição se apresentaram superiores para todas as estruturas de séries. Este comportamento não foi similar ao encontrado nas séries com $T = 25$, onde as evidências de interação foram identificadas no

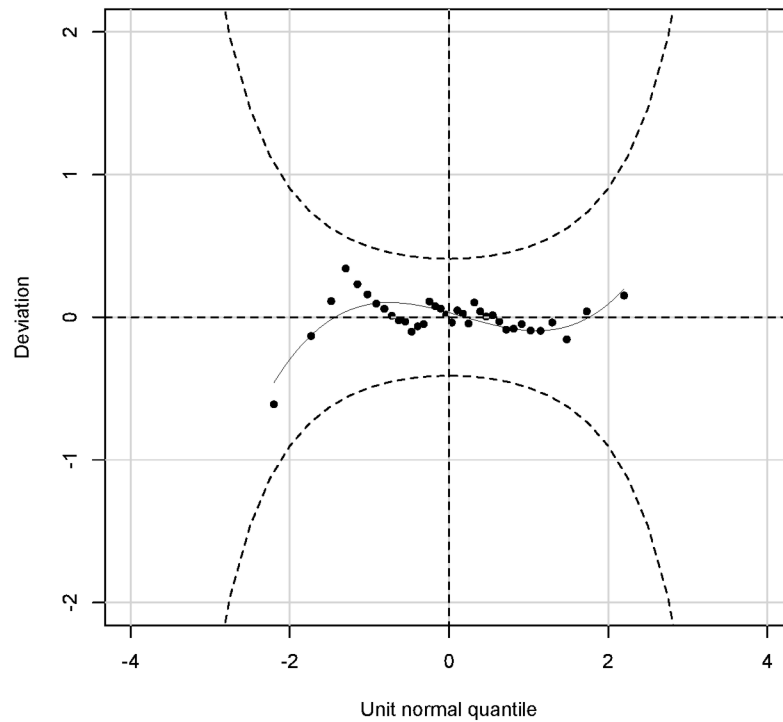


Figura 4.41: *Worm plot* gerado para o modelo GAMLSS considerando a presença de raiz unitária e $T = 40$

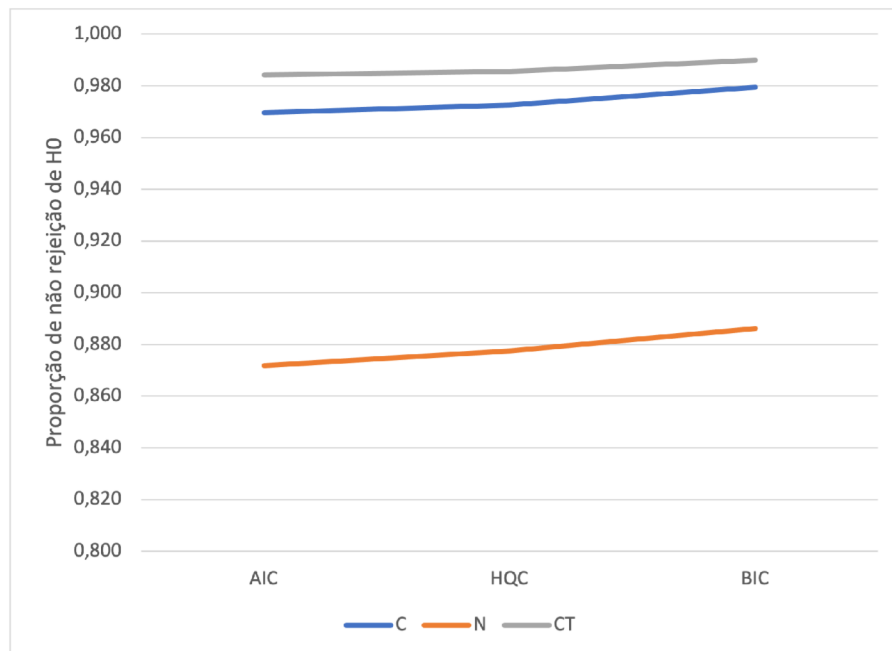


Figura 4.42: Interações entre critério de informação e tipo de modelo, considerando a presença de raiz unitária e $T=40$

modelo com estrutura do tipo CT.

Por fim, ainda observando a Tabela 4.20, é possível verificar indícios de interação entre

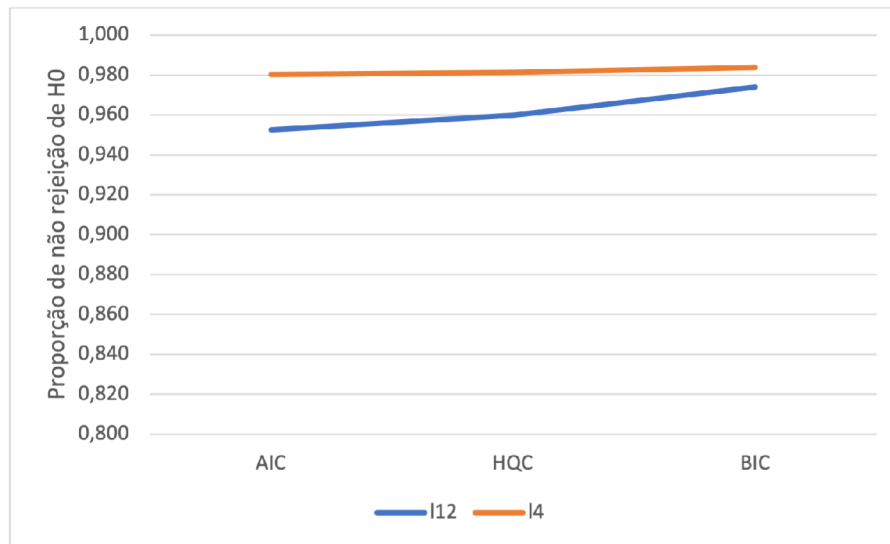


Figura 4.43: Interações entre critérios de seleção e de determinação da defasagem máxima, considerando a presença de raiz unitária e $T=40$

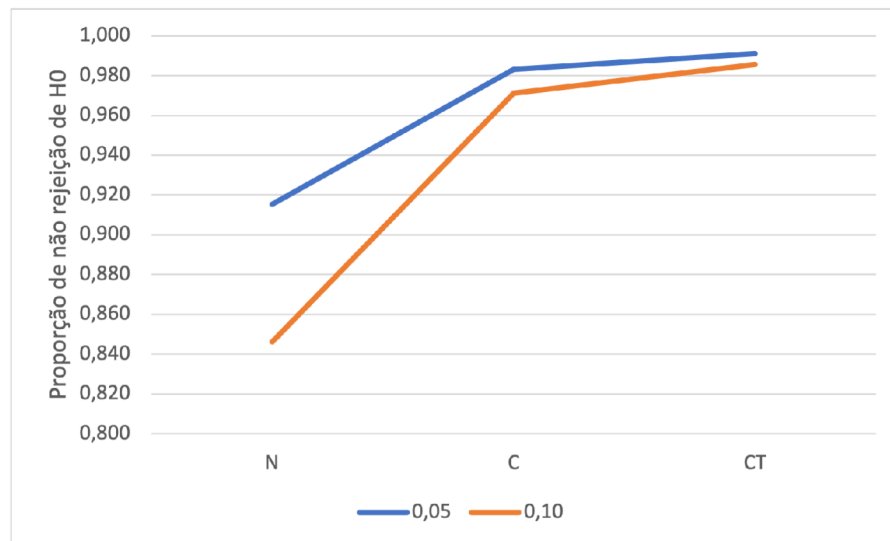


Figura 4.44: Interações entre tipo de modelo e nível de significância, considerando a presença de raiz unitária e $T=40$

a estrutura do modelo e o critério de seleção da defasagem máxima (Figura 4.45), além da interação entre o nível de significância e o critério de seleção da defasagem máxima, que pode ser observada na Figura 4.46. Nestes casos, os achados são os mesmos que para os casos em que $T=25$ e $T = 50$.

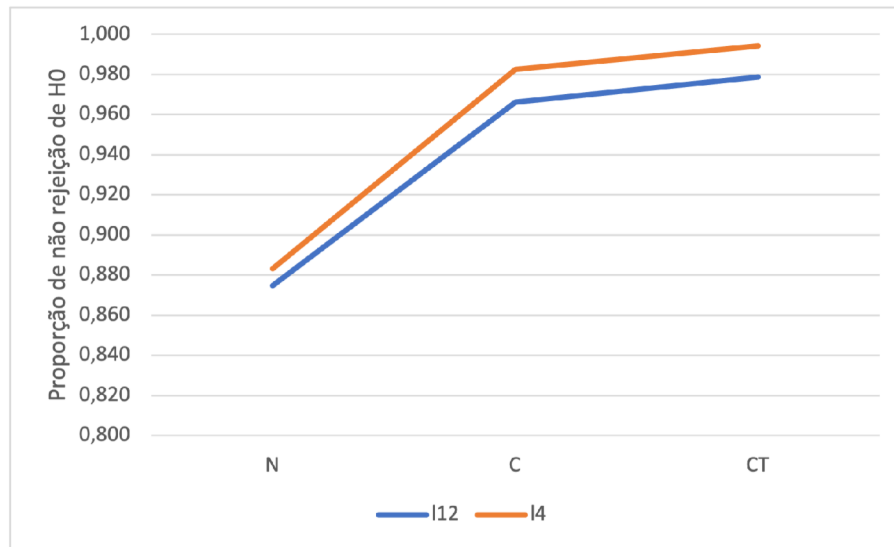


Figura 4.45: Interações entre o tipo de modelo e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=40$

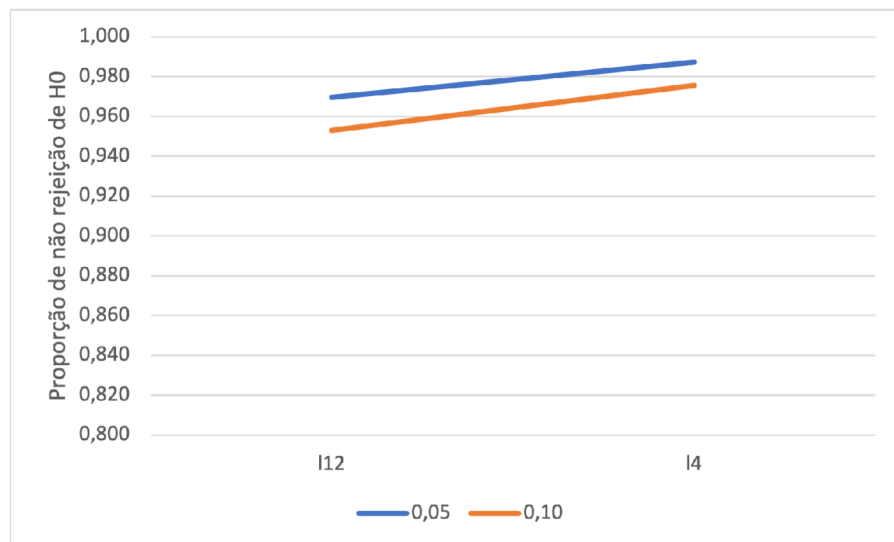


Figura 4.46: Interações entre o nível de significância e o critério de seleção da defasagem máxima, considerando a presença de raiz unitária e $T=40$

ii. Séries estacionárias

Na Tabela 4.21 são apresentados a mediana, a média e o desvio padrão da taxa de rejeição da hipótese nula para cada um dos níveis dos fatores considerados no experimento, enquanto nos gráficos da Figura 4.47, box plots que informam a mediana, os quartis e valores máximo e mínimo, além da presença de valor fora do padrão e de assimetria.

Tabela 4.21: Média, Mediana e desvio padrão da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 40$

Fatores	Níveis	Média	Mediana	Desvio padrão
Critério de Informação	AIC	0,62867	0,6175	0,27204
	BIC	0,62566	0,62380	0,29491
	HQC	0,63048	0,62295	0,28033
Defasagem Máxima	l12	0,51109	0,47395	0,26593
	l4	0,74546	0,79790	0,24395
Estrutura do modelo	C	0,77561	0,89525	0,24128
	N	0,71543	0,77085	0,25023
	CT	0,39378	0,36245	0,17902
Nível de Significância	0,05	0,51400	0,57421	0,29261
	0,10	0,68234	0,67255	0,25813
Φ	0,55	0,69930	0,68955	0,20970
	0,75	0,63877	0,63055	0,27730
	0,95	0,54674	0,41135	0,32645

Ao analisar a Tabela 4.21, na amostra analisada, os resultados diferem dos encontrados para $T=25$ e para $T = 50$ no critério de informação, sendo os melhores modelos selecionados pelo BIC. Para os demais fatores, os resultados foram semelhantes aos outros casos. Em relação aos fatores de maior e menor variabilidade, os achados se assemelham aos do caso em que $T=25$.

No gráficos da Figura 4.47 nota-se que o comportamento dos fatores é semelhante ao encontrado para as séries de comprimento 25, porém com os valores de mediana, em geral, superiores, o mesmo acontecendo para séries com $T = 50$, mas assumindo valores inferiores.

Após, o efeito de cada fator e respectivas interações são analisados por meio da modelagem GAMLSS, cujos resultados são apresentados na Tabela 4.22, seguida do wormplot para verificação das suposições teóricas, o que é feito nos resíduos do modelo, podendo ser visualizado na Figura 4.48. Nesta análise foi tomado como referência o critério de informação HQC, modelo de estrutura N, nível de significância 0.05, critério de determinação da defasagem máxima l_{12} e ϕ 0,75.

Ao observar o gráfico da Figura 4.48, nota-se que todos os pontos estão dentro da banda de confiança, indicando que o modelo está bem ajustado. Diante disso, parte-se para a interpretação dos resultados presentes na Tabela 4.22.

Em relação aos efeitos principais, os resultados, apresentados na Tabela 4.22, indicam acha-

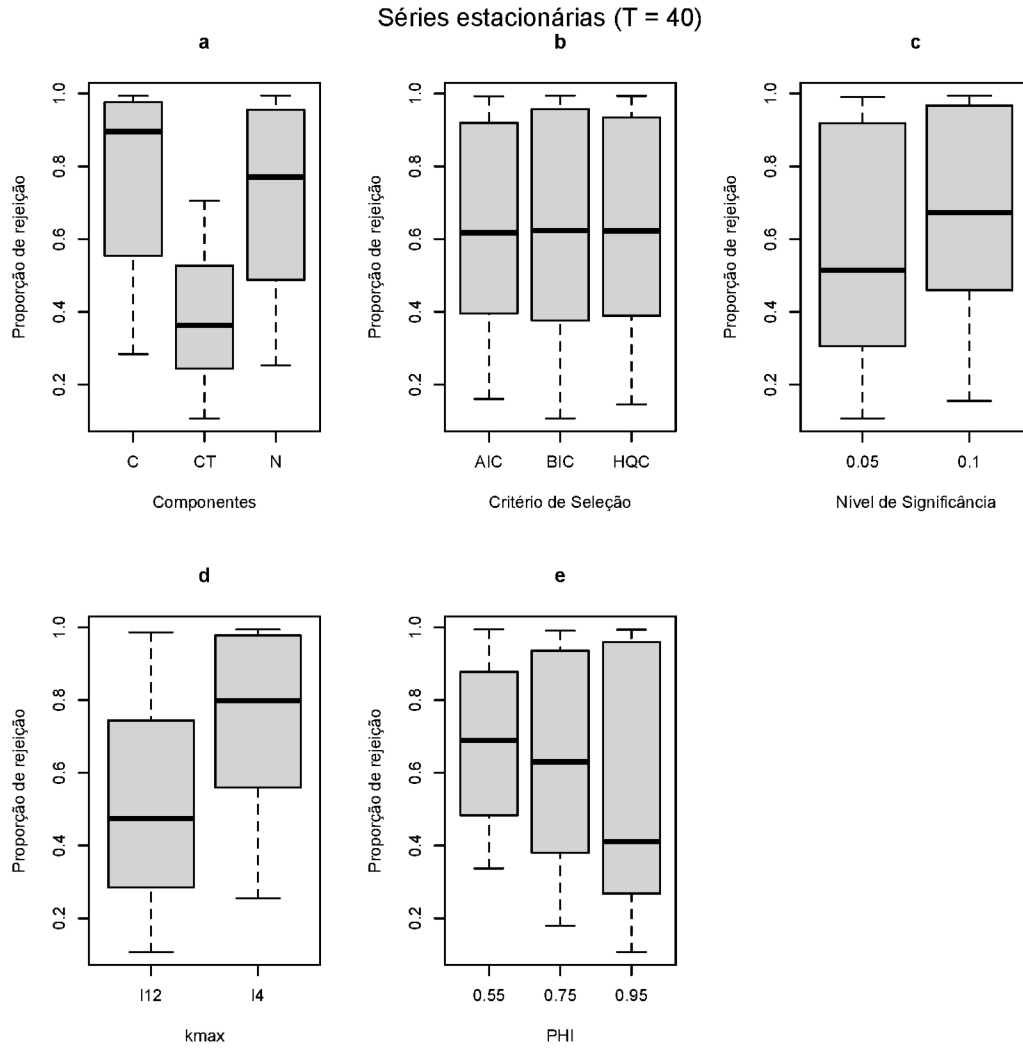


Figura 4.47: Box plots da taxa de rejeição da hipótese nula em experimentos com séries estacionárias e $T = 40$

dos semelhantes aos encontrados para as séries de comprimento 25 e 50.

Em se tratando das interações, há evidências de sua presença entre o nível de significância e diferentes tipos de modelos, conforme indicado na Tabela 4.22. De acordo com a Figura 4.49, a proporção de rejeição da hipótese nula, quando o nível de 0,10 é adotado, apresenta menor redução quando o modelo é alterado de C para N do que quando o nível de significância é 0,05. Cabe ressaltar que quando considerado nível de significância de 0,10, a taxa de rejeição de H_0 apresentou valores superiores para todas as estruturas de modelos.

Também há evidência de interação entre o tipo de modelo e critério para determinação da defasagem máxima. Conforme pode ser visto na Figura 4.50, a proporção de rejeição da hipótese nula, quando o modelo é alterado de C para N apresenta uma redução quando é escolhido o l_4 , ao passo que quando é adotado o l_{12} este valor se aumenta. Além disso, embora haja um decréscimo na taxa de rejeição de H_0 quando o modelo é alterado de N para CT para ambos os critérios, este é menos acentuado para l_{12} .

O tipo de estrutura do modelo interage com valor de ϕ . De acordo com a Figura 4.51,

Tabela 4.22: Análise do efeito principal dos fatores e respectivas interações na taxa rejeição da hipótese nula em experimentos com séries estacionárias e $T = 40$

Fator/Nível	Estimativa	Erro padrão	p-valor
Intercepto	-0,07795	0,18828	0,68007
criterioAIC	-0,02124	0,20926	0,91942
criterioBIC	-0,04046	0,20926	0,84722
modeloC	-0,26953	0,20926	0,20181
modeloCT	-1,49447	0,20926	0,00000
nivelsig0,1	0,98617	0,18265	0,00000
kmaxl4	2,82136	2,82136	0,00000
phi0,55	1,94277	0,20926	0,00000
phi0,95	-1,12219	0,20926	0,00000
criterioAIC.modeloC	-0,13437	0,19373	0,49015
criterioBIC.modeloC	0,24271	0,19373	0,21429
criterioAIC.modeloCT	0,10400	0,19373	0,59302
criterioBIC.modeloCT	-0,18427	0,19373	0,34467
criterioAIC.nivelsig0.1	-0,00747	0,15818	0,96245
criterioBIC.nivelsig0.1	0,05155	0,15818	0,74543
criterioAIC.kmaxl4	-0,01556	0,15818	0,92191
criterioBIC.kmaxl4	0,07482	0,15818	0,63761
modeloC.nivelsig0.1	-0,37295	0,15818	0,02107
modeloCT.nivelsig0.1	-0,46996	0,15818	0,00402
modeloC.kmaxl4	0,34623	0,15818	0,03181
modeloCT.kmaxl4	-0,79833	0,15818	0,00000
nivelsig0.1.kmaxl4	0,20997	0,12916	0,10832
modeloC.phi0.55	-2,03594	0,19373	0,00000
modeloCT.phi0.55	-0,98099	0,19373	0,00000
modeloC.phi0.95	4,37957	0,19373	0,00000
modeloCT.phi0.95	1,28462	0,19373	0,00000
nivelsig0.1.phi0.55	0,04202	0,15818	0,79124
nivelsig0.1.phi0.95	-0,25999	0,15818	0,10455
criterioAIC.phi0.55	-0,02673	0,19373	0,89063
criterioBIC.phi0.55	0,00665	0,19373	0,97272
criterioAIC.phi0.95	-0,06911	0,19373	0,72234
criterioBIC.phi0.95	0,11002	0,19373	0,57186
kmaxl4.phi0.55	-1,42601	0,15818	0,00000
kmaxl4.phi0.95	-1,92276	0,15818	0,00000

quando o modelo é alterado de N para CT há uma redução na proporção de rejeição de H_0 , porém esta é mais acentuada para $\phi = 0,55$. Quando o modelo é alterado de C para N, há um crescimento na taxa de rejeição quando ϕ é considerado 0,55 e 0,75. Entretanto, para 0,95 é observada uma redução na taxa de rejeição de H_0 . Assim, quando ϕ é 0,95, o teste obteve melhores resultados apenas para a estrutura C.

Por fim, há indícios de interação entre os fatores k_{max} e ϕ , que podem ser observados com o auxílio da Figura 4.52. É possível observar que os achados neste caso se assemelham aos

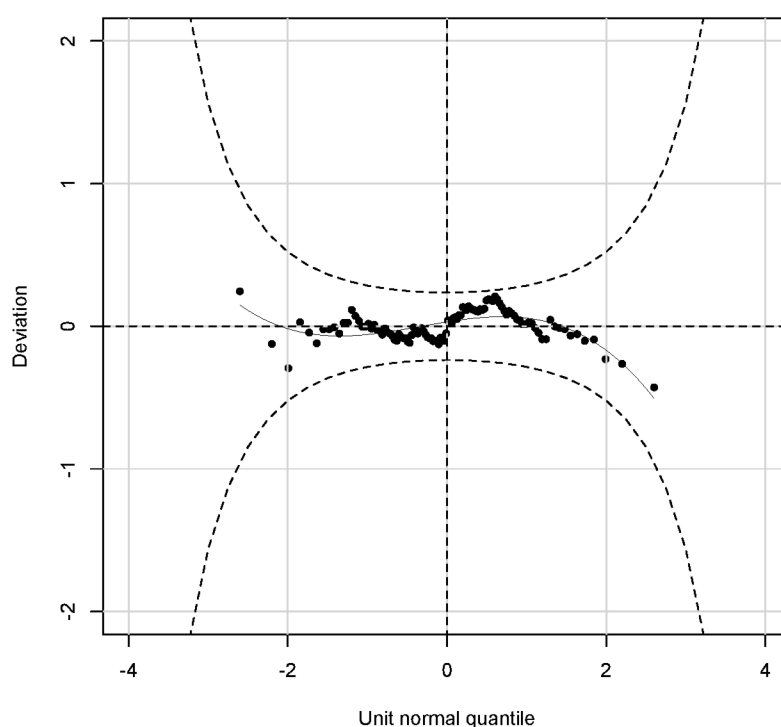


Figura 4.48: *Worm plot* gerado para o modelo GAMLSS considerando a série estacionária e $T = 40$

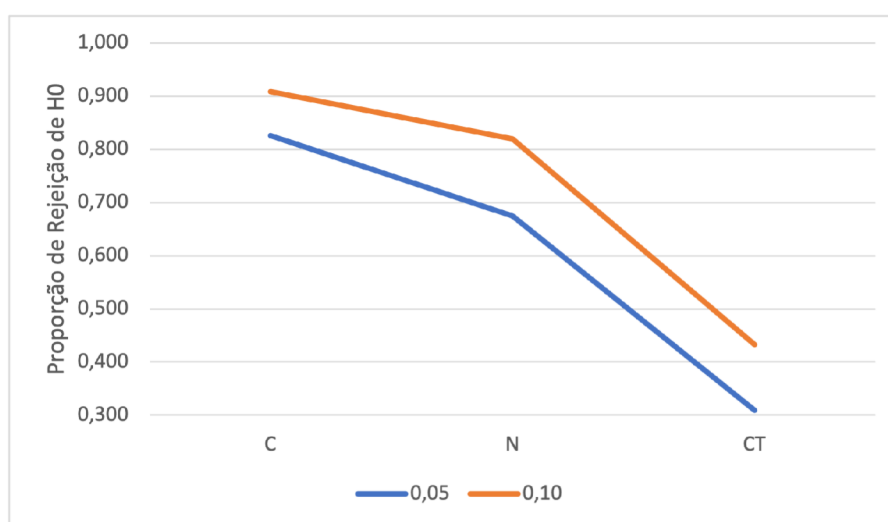


Figura 4.49: Interações entre nível de significância e tipo de modelo considerando a série estacionária e $T = 40$

encontrados nas séries de comprimento 25.

Diante do exposto, é possível considerar que a análise dos resultados obtidos nas simulações complementares, com $T = 40$, apresentou resultados coerentes com os resultados das análises encontrados para $T = 25$ e $T = 50$.

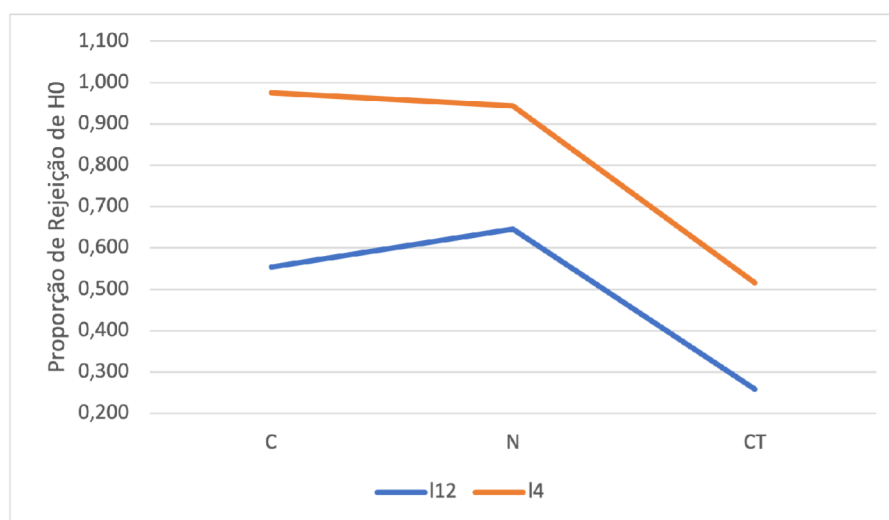


Figura 4.50: Interações entre tipos de modelo e de determinação da defasagem máxima considerando a série estacionária e $T = 40$

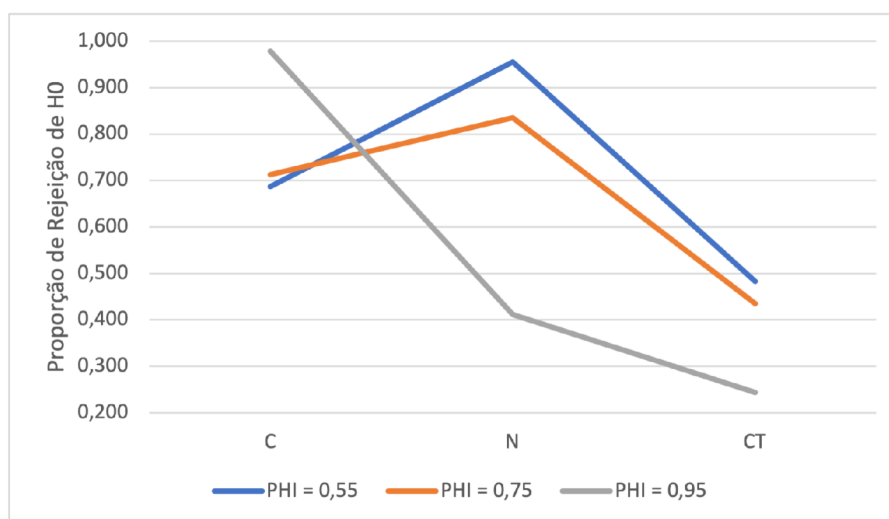


Figura 4.51: Interações entre o tipo de modelo e valor de ϕ considerando a série estacionária e $T = 40$

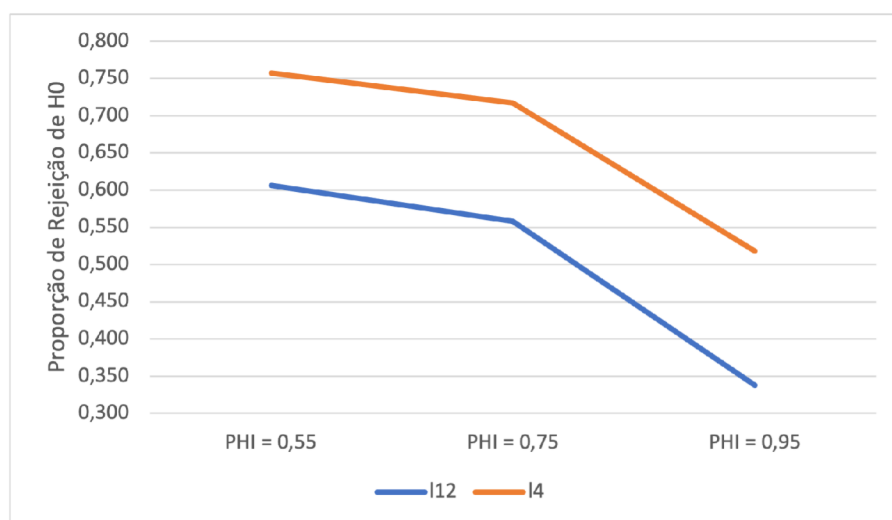


Figura 4.52: Interações entre o nível de significância e o critério de seleção da defasagem máxima considerando a presença de raiz unitária e $T = 40$

4.4 A estratégia

Foram realizadas simulações para séries não estacionárias considerando diferentes cenários de seleção de defasagens, diferentes níveis de significância, diferentes estruturas e comprimentos. Para todos os casos, os melhores resultados foram encontrados quando a defasagem máxima foi determinada por l_4 . Em relação ao critério de informação, o melhor desempenho foi encontrado para o BIC. Já o nível de significância onde foi encontrada maior taxa de acerto para o teste foi 0,05. Considerando todos os fatores e níveis individualmente, a proporção média com a menor taxa de acerto se manteve superior a 0,88.

Em seguida, foram testadas as séries estacionárias, levando em conta os mesmos fatores e níveis considerados para o caso em que foi considerada a raiz unitária. Além disso, foi adicionado o fator ϕ , representando o coeficiente do componente autorregressivo. Neste caso, os resultados foram semelhantes aos encontrados para o caso anterior, porém com os critérios de informação sem apresentar diferença significativa entre eles. Em relação ao nível de significância, o melhor resultado ocorreu ao adotá-lo com o valor de 0,10. Além disso, em relação ao ϕ , melhores resultados foram encontrados para o valor de 0,55. Para as séries estacionárias foram encontrados valores médios de acerto mais baixos do que para o caso anterior, sendo o menor deles com valor médio pouco superior a 0,27.

Diante disso, para séries de comprimento entre 25 e 100, recomenda-se a utilização do critério l_4 para a determinação do número máximo de defasagens e que a escolha do número ideal seja feita pelo BIC. Este achado concorda com o trabalho de Hall (1994), onde o autor recomenda a utilização dos critérios HQC ou BIC, porém para séries de comprimento 100 e 250. Além disso, é indicado que sejam considerados os níveis de significância de 0,05 e 0,10 na aplicação do teste, considerando o primeiro valor para resultados que mostrem a presença de raiz unitária e o segundo, para a sua ausência, sendo a série estacionária. Caso seja encontrado um valor-p entre os dois níveis de significância adotados, sugere-se que seja feita também a aplicação do teste KPSS, pois possui como hipótese nula a estacionariedade e complementa a análise realizada.

5 ESTUDO DE CASO

Para a realização de um estudo de caso, a fim de ilustrar a aplicação da estratégia desenvolvida, a série do consumo anual residencial de eletricidade foi retirada do site do Ipeadata, e tem como fonte dos dados a Eletrobras (2022). A série apresenta dados anuais de 1952 a 2022, contando com 72 observações. Para avaliar a metodologia proposta, será feita a aplicação do teste nesta série com o auxílio da função *ur.df* - mais detalhes do procedimento realizado pela função na execução do teste podem ser encontrados em Enders (2015), considerando a seleção das defasagens com a configuração que alcançou a maior taxa de acerto do teste. Deste modo, a defasagem máxima será determinada pelo critério l_4 , a seleção de defasagens será por BIC e serão considerados os níveis de significância de 0,05 e 0,10.

A Tabela 5.1 mostra as principais medidas da série a ser analisada e a Figura 5.1 apresenta o respectivo boxplot.

Tabela 5.1: Medidas resumo do consumo de energia elétrica no Brasil, 1952-2022 (em nível)

Medida	Resultados
Valor mínimo (GWh)	8.513,00
Valor máximo (GWh)	508.708,00
Média (GWh)	202.479,00
Mediana (GWh)	179.572,00
Desvio padrão (GWh)	167.124,80
Coeficiente de assimetria	0,407
Coeficiente de curtose	-1,279

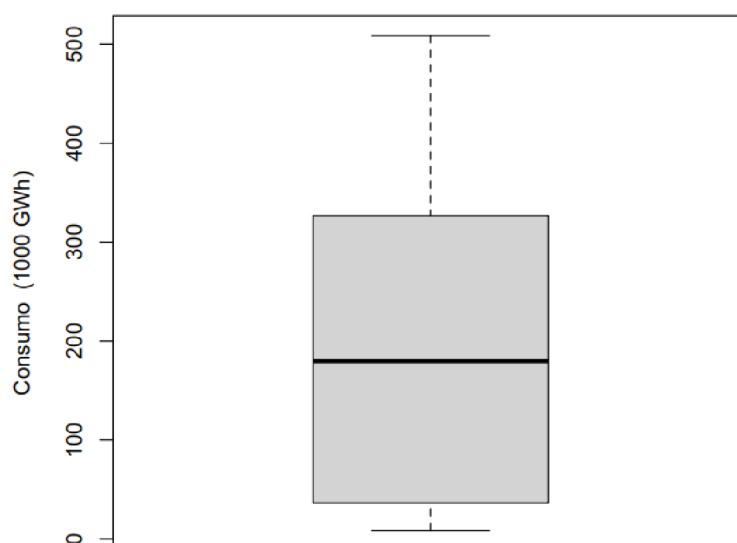


Figura 5.1: Box plot da série de consumo de eletricidade

Através das medidas presentes na Tabela 5.1 e na interpretação do gráfico da Figura 5.1, é possível notar que a série apresenta uma elevada amplitude, uma pequena diferença entre os

valores de média e mediana, simetria, principalmente para os valores entre o primeiro e terceiro quartil, baixo valor positivo para o coeficiente de assimetria, indicando maior concentração dos valores inferiores, e distribuição platicúrtica, indicando que a distribuição dos valores apresenta caudas leves.

A Figura 5.2 apresenta a evolução da série de consumo de eletricidade ao longo do tempo.

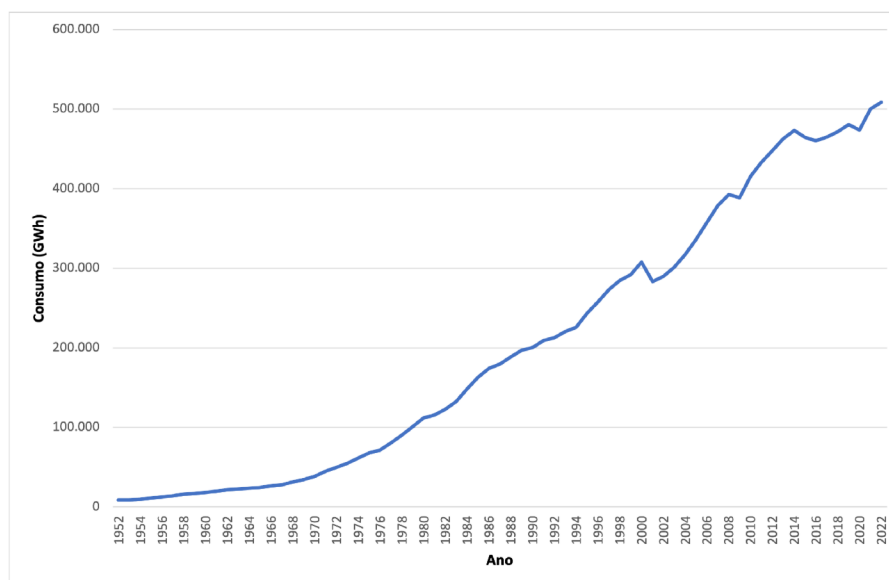


Figura 5.2: Consumo de energia elétrica no Brasil, 1952-2022

Ao observar a Figura 5.2, é possível notar que o consumo de eletricidade vem crescendo ao longo dos anos, cabendo ressaltar que este crescimento é mais acentuado após os vinte anos iniciais de observação. Além disso, existem pequenas quedas de consumo em alguns anos específicos. Aparentemente existem tendências diferentes de comportamento ao longo do tempo. Estas propriedades dos dados fazem com que haja suspeita de que a série seja não estacionária.

A aplicação do teste é feita considerando o procedimento descrito na Figura 2.1. Com o número máximo de defasagens definido, é testada a hipótese nula de que a série não possui tendência ($\beta_2 = 0$) e possui raiz unitária ($\alpha = 0$), conforme as equações 2.5 e 2.11. Caso a hipótese nula não seja rejeitada, a série possui raiz unitária e não possui tendência. Caso seja rejeitada, pelo menos uma das condições anteriores é falsa. Também é testada a hipótese de que a série não possui tendência nem constante e possui raiz unitária, $\alpha = \beta_1 = \beta_2 = 0$. Caso essa hipótese seja rejeitada, há indícios de que pelo menos uma das três condições anteriores é falsa. Em seguida é testado somente se $\alpha = 0$. Caso haja indícios de que a série não possui tendência, é estimado um modelo somente com constante. Neste caso, a hipótese nula é de que $\beta_1 = 0$ e $\alpha = 0$. Caso não seja possível rejeitar H_0 , há indícios de que a série possui raiz unitária e não possui constante. Caso seja rejeitada H_0 , pelo menos uma das afirmações anteriores é falsa. Em seguida é testado se $\alpha = 0$. Caso haja indícios de que a série não possui constante, é estimado um modelo sem constante e sem tendência.

Os resultados são apresentados na Tabela 5.2.

Considerando o nível de significância de 5%, o teste apresentou indícios de que não se pode

Tabela 5.2: Resultados do teste ADF para a série de consumo de eletricidade em nível

Modelo	H_0	Valor Calculado	Valor Crítico (5%)	Valor Crítico (10%)
CT	$\alpha = \beta_2 = 0$	5,732	6,490	5,470
	$\alpha = 0$	-2,521	-3,450	-3,150
	$\alpha = \beta_1 = \beta_2 = 0$	11,095	4,880	4,160
C	$\alpha = 0$ e $\beta_1 = 0$	11,083	4,710	3,860
	$\alpha = 0$	1,624	-2,890	-2,580

rejeitar a hipótese de que a tendência é igual a zero ($0,05 < \text{valor} - p < 0,10$). Quando realizado o teste considerando a presença de constante, há evidências de que a série apresenta constante ($\text{valor} - p < 0,05$) e possui raiz unitária ($\text{valor} - p > 0,10$).

Em seguida, diferencia-se a série e são feitas as mesmas análises para a série em primeira diferença. A Tabela 5.3 mostra as principais medidas da série em primeira diferença e a Figura 5.3 apresenta o boxplot.

Tabela 5.3: Medidas resumo do consumo de energia elétrica no Brasil, 1952-2022 (em primeira diferença)

Medida	Resultados
Valor mínimo (GWh)	-24.271,00
Valor máximo (GWh)	27.073,00
Média (GWh)	7.146,0
Mediana (GWh)	6.651,00
Desvio padrão (GWh)	8.125,78
Coeficiente de assimetria	-0,397
Coeficiente de curtose	2,2038

Com base nas medidas presentes na Tabela 5.3 e na interpretação do gráfico da Figura 5.3 é possível notar que o maior e o menor valor apresentam um valor próximo para a distância da origem, há uma pequena diferença entre os valores de média e mediana, existe simetria para os valores entre o primeiro e terceiro quartil, estão presentes valores fora do padrão, mostra baixo valor negativo para o coeficiente de assimetria e que a série apresenta uma distribuição leptocúrtica.

A Figura 5.4 apresenta a evolução da série de consumo de eletricidade, em primeira diferença, ao longo do tempo.

A Figura 5.4 mostra a diferença de consumo entre anos consecutivos durante o período observado. Em 2001 houve a maior redução do consumo durante todo o período observado. Destacam-se como maiores aumentos no consumo, os anos de 2011 e 2021. Apesar destes valores que fogem ao padrão, a série aparenta ter um comportamento constante ao longo do tempo, levando a suspeitar que ela seja estacionária.

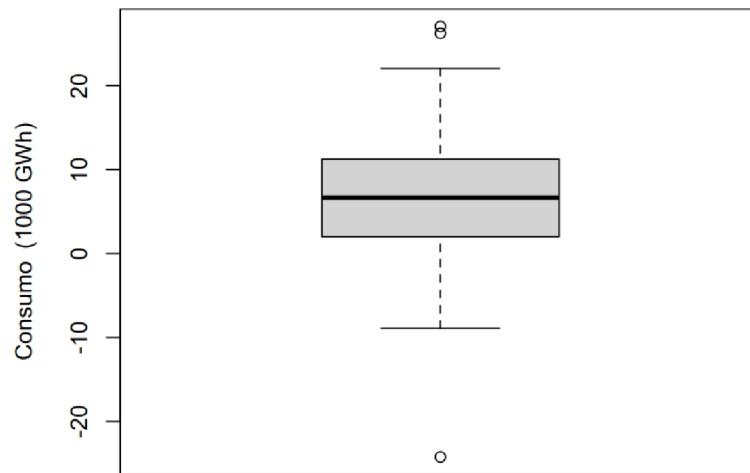


Figura 5.3: Box plot da série de consumo de eletricidade

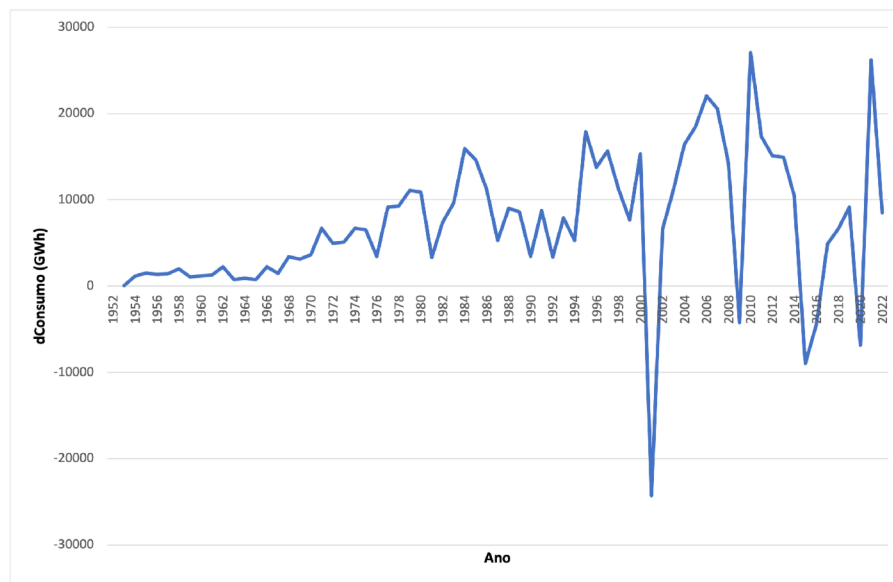


Figura 5.4: Consumo de energia elétrica no Brasil, 1952-2022 (em primeira diferença)

O procedimento para o teste é o mesmo realizado para a série em nível e os resultados são apresentados na Tabela 5.4.

Tabela 5.4: Resultados do teste ADF para a série de consumo de eletricidade em primeira diferença

Modelo	H_0	Valor Calculado	Valor Crítico (5%)	Valor Crítico (10%)
CT	$\alpha = 0$ e $\beta_2 = 0$	9,950	6,490	5,470
	$\alpha = 0$	-4,460	-3,450	-3,150
	$\alpha = \beta_1 = \beta_2 = 0$	6,666	4,880	4,160

Neste caso, há indícios de que a série não possui raiz unitária ($valor - p < 0,05$) e possui tendência ($p - valor < 0,05$). Assim, é possível concluir que há evidências de que a série possui ordem de integração igual a um, denotada por $I(1)$.

6 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi definir uma estratégia a ser utilizada para a determinação do número máximo e ótimo de defasagens na identificação da presença de raiz unitária pelo teste ADF, com base em técnicas descritas na literatura. Para atingi-lo foram utilizados conceitos de mais de uma área do conhecimento, o que caracteriza este estudo como interdisciplinar. Além disso, esta tese apresenta uma contribuição importante para pesquisadores e analistas de dados que desejem modelar séries temporais utilizando modelos que exijam estacionariedade.

Por meio de uma revisão sistemática, acompanhada de uma análise bibliométrica, foi possível identificar lacunas sobre o teste ADF, sendo confirmado que realmente não há um consenso sobre a maneira de determinar o número de defasagens a ser aplicado no teste. Fatores como a definição da defasagem máxima e o critério de informação utilizado para determinar o número de defasagens a ser aplicado no teste, podem influenciar no resultado. Além disso, é importante considerar o comprimento da série e se ela apresenta ou não termos determinísticos. Apesar de um grande número de novos testes de raiz unitária estarem surgindo nos últimos tempos, raros deles são capazes de superar o ADF, e isso só acontece em situações específicas.

Uma simulação preliminar, considerando diferentes condições experimentais, conseguiu comprovar a influência do número de defasagens na proporção de rejeição da hipótese nula do teste ADF. Foram encontradas evidências de que o número de defasagens interfere no resultado do teste apenas quando as séries são estacionárias, diminuindo o poder do teste conforme aumenta o número de defasagens. Esse achado está de acordo com o trabalho de Tam (2013) que conclui que o teste ADF é altamente sensível ao método de seleção de defasagens.

Na sequência, para analisar o resultado do teste ADF em diferentes condições, foram feitas simulações considerando séries estacionárias e não estacionárias, com diferentes estruturas (presença ou não de tendência e constante), sendo o teste ADF aplicado com diferentes procedimentos para identificar as defasagens máxima e ótima. Também foram utilizados diferentes níveis de significância para a tomada de decisão.

A análise dos resultados permitiu identificar um dos procedimentos para determinação da defasagem máxima interferindo positivamente no desempenho do teste ADF, o mesmo acontecendo com um dos critérios de informação, procedimento utilizado para identificar a quantidade de defasagens ótima. Estas informações foram usadas para elaborar a estratégia, que também levou em consideração o nível de significância. Além disso, estratégia também considerou que em alguns casos existe a necessidade de confirmar os achados com o teste KPSS.

A estrutura do modelo não foi levada em consideração, mas o pior desempenho do teste foi em modelos sem constante e sem tendência, estruturas que dificilmente ocorrem na prática, enquanto a melhor, em modelos só com constante. Em séries de menor amplitude, modelos com tendência e constante apresentaram bons resultados. Convém salientar que a análise também identificou um desempenho superior do teste em séries com a presença de raiz unitária, concordando com Arltová e Fedorová (2016), onde é mostrado que o ADF não apresenta problema

de tamanho do teste em modelos $AR(1)$.

No estudo de caso foi aplicada a estratégia proposta, onde foi possível verificar que a série de consumo residencial de eletricidade no Brasil é estacionária em primeira diferença, concordando com a maioria dos estudos desenvolvidos sobre modelagem do consumo/demanda de energia elétrica por meio de modelos estatísticos, como por exemplo em Schmidt e Lima (2004) e Viana e Silva (2014). Assim, a estratégia proposta pode ser utilizada na identificação da presença de raiz unitária em séries temporais formadas por dados reais.

A detecção, ou não, da presença de raiz unitária corretamente é indispensável para o bom desempenho de diversas técnicas estatísticas e econométricas de modelagem de séries temporais. Deste modo, aprimorar esta etapa do processo é de suma importância para otimizar o desempenho das técnicas de modelagem de séries temporais que só podem ser aplicadas em séries estacionárias.

A metodologia utilizada para gerar dados e analisá-los se mostrou pertinente e permitiu a elaboração de uma estratégia de definição, tanto da quantidade de defasagens máxima quanto da quantidade de defasagens ótima, entretanto está restrita ao caso de séries temporais sem sazonalidade e de comprimento entre 25 e 100.

Para trabalhos futuros, ficam como sugestões: replicar o estudo para séries de comprimento maior que 100; utilizar séries com sazonalidades de diferentes amplitudes (semanais, mensais, trimestrais); analisar o desempenho da estratégia em outras séries com dados reais; analisar o impacto de quebras estruturais nos resultados do teste; analisar mais detalhadamente o impacto da na estrutura do modelo, mais especificamente, da presença de constante e tendência.

REFERÊNCIAS

- AL-MUSAYLH, M. S. et al. *Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Advanced Engineering Informatics*, v. 35, p. 1-16, 2018.
- ANEEL, *Atlas de energia elétrica do Brasil. Agência Nacional de Energia Elétrica - Brasília, 2002*. Disponível em: <https://www.aneel.gov.br>; acesso em: abr. 2021
- ANEEL, *Atlas de energia elétrica do Brasil. Agência Nacional de Energia Elétrica. 3 ed. - Brasília, 2008*. Disponível em: <https://www.aneel.gov.br>; acesso em: abr. 2021
- AKAIKE, H. *Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika*, v. 60, n. 2, p. 255-265, 1973.
- ARLTOVÁ, M.; FEDOROVÁ, D. *Selection of unit root test on the basis of length of the time series and value of $ar(1)$ parameter. Statistika*, v. 96, n. 3, p. 3, 2016.
- AYLAR, E.; SMEEKES, S.; WESTERLUND, J. *Lag truncation and the local asymptotic distribution of the ADF test for a unit root. Statistical Papers*, v. 60, n. 6, p. 2109-2118, 2019.
- BAYER, F. M., *Wavelets e modelos tradicionais de previsao: Um estudo comparativo. Rev. Bras. Biom*, 28(2):40-61, 2010
- BUENO, R. L. S. *Econometria de séries temporais. Cengage Learning, 2008*.
- BUUREN, S. V. "Worm plot to diagnose fit in quantile regression." *Statistical Modelling* 7, no. 4. 2007.
- BUUREN, S. v.; FREDRIKS, M. *Worm plot: a simple diagnostic device for modelling growth reference curves. Statistics in medicine, Wiley Online Library*, v. 20, n. 8, p. 1259-1277, 2001.
- BURRIDGE, P.; HRISTOVA, D. *Consistent estimation and order selection for nonstationary autoregressive processes with stable innovations. Journal of Time Series Analysis*, v. 29, n. 4, p. 695-718, 2008.
- CAVALIERE, G. et al. *Lag length selection for unit root tests in the presence of nonstationary volatility. Econometric Reviews*, v. 34, n. 4, p. 512-536, 2015.
- CAVALIERE, G.; GEORGIEV, I.; TAYLOR, R. *Unit root inference for non-stationary linear processes driven by infinite variance innovations. 2016*.
- CANER, M.; KNIGHT, K. *An alternative to unit root tests: Bridge estimators differentiate between nonstationary versus stationary models and select optimal lag. Journal of Statistical Planning and Inference*, v. 143, n. 4, p. 691-715, 2013.

- CHEUNG, .Y.; LAI, K. S. *Lag order and critical values of the augmented Dickey–Fuller test. Journal of Business & Economic Statistics*, v. 13, n. 3, p. 277-280, 1995.
- CHEUNG, Y.; LAI, K. S. *Power of the augmented dickey-fuller test with information-based lag selection. Journal of Statistical Computation and Simulation*, v. 60, n. 1, p. 57-65, 1998.
- CHOI, I. *Almost all about unit roots: Foundations, developments, and applications. Cambridge University Press*, 2015.
- CHRISTOPOULOS, D. K. *Does a non-linear mean reverting process characterize real GDP movements?. Empirical Economics*, v. 31, n. 3, p. 601-611, 2006.
- DE ASSIS CABRAL, J.; LEGEY, L.F.; DE FREITAS CABRAL, M. V. *Electricity consumption forecasting in Brazil: A spatial econometrics approach. Energy*, v. 126, p. 124-131, 2017.
- DE OLIVEIRA, V. H.; TAVARES, R.S.; TAVARES, L. A. *Demanda residencial por energia elétrica no Brasil (2004-2015). Revista Catarinense de Economia*, v. 2, n. 1, p. 142-162, 2018.
- DE MENDIBURU, F. *Agricolae. Statistical Procedures for Agricultural Research. R package version, p. 1.3-2., 2020*
- DENG, L. *Augmented Dickey-Fuller Test and the Lag Length Selection Problem. In: Applied Mechanics and Materials. Trans Tech Publications Ltd*, p. 3019-3022, 2012.
- DICKEY, D. A.; FULLER, W. A., *Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association*, v. 74, n. 366a, p. 427-431, 1979.
- DUNN, P. K., and SMYTH, G. K. *Randomized quantile residuals. Journal of Computational and graphical statistics* 5, no. 3, 236-244, 1996
- ELETROBRÁS *Consumo residencial de energia elétrica (GWh) Disponível em: <http://www.eletrobras.gov.br/>; acesso em: jan. 2023*
- ELLIOTT, G.; ROTHENBERG, T. J.; STOCK, J. H. *Efficient tests for an autoregressive unit root. National Bureau of Economic Research*, 1992.
- ENDERS W. *Applied econometric time series fourth edition. New York (US): University of Alabama*. 2015.
- ENGLE, R. F.; GRANGER, C. W. J. *Co-integration and error correction: representation, estimation, and testing. Econometrica: journal of the Econometric Society*, p. 251-276, 1987.
- EPE. *Atlas da eficiência energética no Brasil em 2019 - Empresa de Pesquisa Energética - Rio de Janeiro, 2019. Disponível em: <http://epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/atlas-da-eficiencia-energeticabrasil-2019>.*

- EPE. *Balanço Energético Nacional 2020: Ano base 2019 - Empresa de Pesquisa Energética – Rio de Janeiro, 2020*. Disponível em <http://www.epe.gov.br/default.aspx>; acesso em: out. 2020
- FERNANDES, S. M.; BORNIA, A.C.; NAKAMURA, L.R. *The influence of boards of directors on environmental disclosure. Management Decision, 2019.*
- FERRER-PÉREZ, H.; AYUDA, M.I.; AZNAR, A. *The sensitivity of unit root tests to the initial condition and to the lag length selection: A Monte Carlo Simulation Study. Communications in Statistics-Simulation and Computation, p. 1-11, 2019.*
- FOSSATI, S. Covariate unit root tests with good size and power. *Computational Statistics & Data Analysis, v. 56, n. 11, p. 3070-3079, 2012.*
- GRANGER, C. W.; NEWBOLD, P. *Spurious regressions in econometrics. Journal of econometrics, 2(2), 111-120, 1974.*
- GUJARATI, D. N.; PORTER, D. C.(2011)., *Econometria Básica. Porto Alegre: Editora AMGH*
- HASTIE, T. J.; R. J. TIBSHIRANI. *Generalized Additive Models. Chapman e Hall, 1990.*
- HARRIS, R. I. *Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. Economics letters, 38(4), 381-386, 1992.*
- HALL, A. *Testing for a unit root in time series with pretest data-based model selection. Journal of Business & Economic Statistics, v. 12, n. 4, p. 461-470, 1994.*
- HANNAN, E.J.; QUINN, B. G. *The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological), v. 41, n. 2, p. 190-195, 1979.*
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice. OTexts, 2018.*
- IRFFI, G. et al. *Previsão da demanda por energia elétrica para classes de consumo na região Nordeste, usando OLS dinâmico e mudança de regime. Econ. Aplicada [online]. 2009*
- HYNDMAN, R. J. et al. *Automatic time series forecasting: the forecast package for R. Journal of statistical software, v. 27, n. 3, p. 1-22, 2008.*
- JAMIL, R. *Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. Renewable Energy, v. 154, p. 1-10, 2020.*
- KIVIET, J. F.; DUFOUR, J. M. *Exact tests in single equation autoregressive distributed lag models. Journal of Econometrics, v. 80, n. 2, p. 325-353, 1997.*

- KWIATKOWSKI, D. et al. *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*. *Journal of econometrics*, v. 54, n. 1-3, p. 159-178, 1992.
- LOPEZ, C.; MURRAY, C. J.; PAPELL, D. H. *State of the art unit root tests and purchasing power parity*. *Journal of Money, Credit and Banking*, p. 361-369, 2005.
- MONTAÑÉS, A; REYES, M. *Structural breaks, unit roots and methods for removing the autocorrelation pattern*. *Statistics & probability letters*, v. 48, n. 4, p. 401-409, 2000.
- MORETTIN, P.; TOLOI, C. *Análise de séries temporais. Vol 1. 3 ed. São Paulo: Editora Blucher, 2018.*
- MURSHED, M. *Modeling primary energy and electricity demands in Bangladesh: An Autoregressive distributed lag approach*. *Sustainable Production and Consumption*, v. 27, p. 698-712, 2021.
- NELDER, J. A.; WEDDERBURN, R. W. M. *Generalized linear models*. *Journal of the Royal Statistical Society, London*, v. 135, n. 3, p. 370-384, 1972.
- OTERO, J.; SMITH, J. *Testing for cointegration: power versus frequency of observation—further Monte Carlo results*. *Economics Letters*, v. 67, n. 1, p. 5-9, 2000.
- ÖZEL, Ö.; ILALAN, D. *An alternative mean reversion test for interest rates*. *Central Bank Review*, v. 18, n. 1, p. 35-39, 2018.
- PAPARODITIS, E.; POLITIS, D.N. *The asymptotic size and power of the augmented Dickey-Fuller test for a unit root*. 2016.
- PHILLIPS, P. C. B. ; PERRON, P. *Testing for a unit root in time series regression*. *Biometrika*, v. 75, n. 2, p. 335-346, 1988.
- PFAFF, B. *Non-stationary time series. Analysis of integrated and cointegrated time series with R*, p. 53-71, 2008.
- QUINECHE, R.; RODRÍGUEZ, G. *Selecting the lag length for the MGLS unit root tests with structural change: A warning note for Practitioners based on simulations*. *Econometrics*, v. 5, n. 2, p. 17, 2017.
- R Core Team *R: a language and environment for statistical computing*, 2020.
- RAMIRES, T.G et al. *Validation of stepwise-based procedure in GAMLSS*. *Journal of Data Science*, 19(1), pp.96-110, 2021.
- RIGBY R.A., STASINOPOULOS D.M. *Generalized additive models for location, scale and shape*. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2005 Jun;54(3):507-54.

- RIGBY, R. A. et al. *Distributions for modeling location, scale, and shape: Using GAMLSS in R*, Chapman and Hall/CRC. 2019
- ROQUIM, F. V. *GAMLSS na experimentação agropecuária: um estudo em doenças parasitárias de bovinos de leite. Dissertação de mestrado. Universidade Federal de Lavras - Lavras/MG. 2018*
- SAID, E.; DICKEY, D. A., *Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika*, v. 71, n. 3, p. 599-607, 1984.
- SILVEIRA, A. G., AMARAL, M. C., and DE MATTOS, V. L. D., *Relação de longo prazo entre a renda média domiciliar da população e a demanda residencial por energia elétrica. Revista Mundi Engenharia, Tecnologia e Gestão (ISSN: 2525-4782), 2019.*
- SCHMIDT, C. A. J.; LIMA, M. A. *A demanda por energia elétrica no Brasil. Revista Brasileira de Economia*, v.58, n.1, p.67-98, 2004.
- SCHWARZ, G. et al. *Estimating the dimension of a model. Annals of statistics*, v. 6, n. 2, p. 461-464, 1978.
- SHUMWAY, R. H.; STOFFER, D. S. . *Time series analysis and its applications (Vol. 3). New York: springer, 2000*
- SCHWERT, G. W. *Tests for unit roots: A Monte Carlo investigation. Journal of Business & Economic Statistics*, v. 7, n. 2, 1989.
- STASINOPOULOS, M. D.; RIGBY, R. A.; HELLER, G. Z.; VOUDOURIS, V.; DE BASTIANI, F. *Flexible Regression and Smoothing: Using GAMLSS in R. Chapman and Hall/CRC The R Series, Boca Raton, FL, 2017.*
- TAM, P. S. *Finite-sample distribution of the augmented Dickey-Fuller test with lag optimization. Applied Economics*, v. 45, n. 24, p. 3495-3511, 2013.
- TAYLOR, A. R. *The Finite Sample Effects of Deterministic Variables on Conventional Methods of Lag-selection in Unit Root Tests. Oxford bulletin of economics and statistics*, v. 62, n. 2, p. 293-304, 2000.
- VIANA, G. I. M. N.; SILVA, A. L. M. (2014), *Um modelo para projeções para demanda por energia elétrica, 2009-2017 para o setor residencial no Brasil. Revista Brasileira de Energia*, v.20. Disponível em <http://new.sbpe.org.br/artigo/327/>; acesso em: abr. 2020
- VOUGAS, D. V. *Application of the Dickey-Fuller test to the Nelson and Plosser (1982) Data. Applied Economics Letters*, v. 9, n. 8, p. 511-514, 2002.
- WEBER, C. E. *Alternative lag length selection criteria and the split-trend stationarity hypothesis. Applied Economics*, v. 33, n. 2, p. 237-247, 2001.

- WU, S.. *Lag length selection in DF-GLS unit root tests. Communications in Statistics-Simulation and Computation*, v. 39, n. 8, p. 1590-1604, 2010.
- ZAR, J. H. *Biostatistical analysis. Pearson Education India*, 1999.
- ZEILEIS, A.; HOTHORN, T. *Diagnostic checking in regression relationships. R News 2: 7–10*, 2002.
- ZUO, X.. *Several Important Unit Root Tests. In: 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP). IEEE, 2019. p. 10-14.*