

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

**Modelo de Séries Temporais Autorregressivo
Periódico - PAR**

Natália Leite Cavalaro

Trabalho de Conclusão de Curso

Natália Leite Cavalaro

Modelo de Séries Temporais Autorregressivo Periódico - PAR

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Natália Leite Cavalaro e aprovado pela banca examinadora.

São Carlos, 29 de junho de 2021.

Banca Examinadora

- Prof.^a Dra. Maria Sílvia de Assis Moura
- Prof. Dr. Francisco Antonio Rojas
- Prof. Dr. Ricardo Felipe Ferreira

Dedicatória

Dedico este trabalho a meus pais, João Reginaldo Cavalaro e Sandra Aparecida Leite Cavalaro, e a minha avó, Elvira Freri Leite, que já se foi, mas que me deu força e amor por todo esse caminho.

Agradecimentos

Primeiramente agradeço a Deus, pois sem ele, nada disso seria possível.

Aos meus pais, João Reginaldo Cavalaro e Sandra Aparecida Leite Cavalaro, por todo suporte, atenção, carinho e amor. Só tenho a agradecer por todos os momentos em que não me deixaram desistir.

Um agradecimento especial a minha avó, Elvira Freri Leite, que se foi em 2020, mas que continua presente no meu coração. Por toda força, generosidade e amor. Esta graduação não seria possível sem ela, minha fonte de inspiração e proteção.

Me sinto muito grata ao meu irmão, Lucas Leite Cavalaro, que me trouxe ao mundo da Estatística. Em muitos momentos viu força em mim, quando nem eu mesma via. Obrigada por todas as conversas, risadas, jogos e festas.

À minha orientadora, Prof.^a Dra. Maria Sílvia de Assis Moura, pela excelente orientação neste Trabalho de Graduação. Obrigada professora, por todo apoio e por me ajudar tanto nesta Graduação.

À todos os professores que fizeram parte da minha formação acadêmica. Em especial, agradeço ao Prof. Dr. Francisco Antonio Rojas e Prof. Dr. Ricardo Felipe Ferreira pelos ótimos apontamentos, sugestões e correções neste trabalho.

Aos meus amigos Thais Maira, Ana Beatriz Monteiro, Samuel Treméa, Henrique Fracjaco, Pedro Moraes Gomes, Pedro Grazziano e, principalmente a Fábio Henrique Nais, por toda a força durante esses anos e por não me deixarem desistir durante todo este processo. A quarentena nos separou, mas logo estaremos juntos comemorando todas as conquistas.

À Universidade Federal de São Carlos por me proporcionar a oportunidade de fazer esta Graduação em Estatística. Obrigada pela Bolsa Alimentação durante o tempo de Graduação. Sou grata ao Programa de Educação Tutorial - PET e a Stepwise, por me oferecer momentos de conhecimento, crescimento pessoal e profissional.

Muito obrigada a todos que fizeram parte deste momento importante da minha vida.

Resumo

Este trabalho apresenta o estudo de um tipo de modelo de séries temporais, chamado de modelo autorregressivo periódico, que surgiu a partir das pesquisas de Thomas e Fiering (1962), de acordo com Hipel e McLeod (1994). Seu uso se dá principalmente em séries temporais que apresentam um comportamento periódico na média, variância e função de autocorrelação. Serão ainda exibidos exemplos de aplicação, em que a série temporal apresenta as características ideais de uso do modelo PAR, e como fazer o procedimento de escolha, estudo, adequabilidade e previsão, desse tipo de modelo, além de ser realizada uma comparação com o modelo de séries temporais sazonais.

Palavras-chave: *modelo autorregressivo periódico, previsão, sazonalidade, série temporal.*

Sumário

1	Introdução	1
2	Conceitos iniciais	3
2.1	Séries temporais	3
2.2	Estacionariedade	4
2.3	Metodologia Box e Jenkins	4
2.3.1	Função de autocorrelação (FAC)	6
2.3.2	Função de autocorrelação parcial (FACP)	6
2.4	Modelo autorregressivo - AR	7
2.5	CrITÉrio de informação Akaike (AIC)	8
2.6	CrITÉrio de informação Bayesiano (BIC)	9
3	Modelo autorregressivo periÓdico	11
3.1	Definição do modelo	11
3.1.1	Modelo autorregressivo periÓdico integrado - PIAR	13
3.2	Exemplo de aplicação do modelo PAR	14
3.2.1	SÉrie temporal diferenciada - Z_t	19
3.3	Identificação da ordem do modelo	24
3.3.1	Função de autocorrelação (FAC)	24
3.3.2	Função de autocorrelação parcial (FACP)	25
3.3.3	Teste de hipótese	26
3.3.4	CrITÉrio de informação Akaike (AIC)	26
3.3.5	CrITÉrio de informação Bayesiano (BIC)	27
3.4	Ordem do modelo do exemplo	27
3.5	Periodicidade no modelo	29

4	Estimação dos parâmetros do modelo	31
4.1	Método de Yule-Walker	31
4.2	Método de representação matricial	32
4.2.1	Estimativas do modelo PAR do exemplo	34
4.3	Verificação de diferenciação no modelo do exemplo	35
4.3.1	Autovalores da matriz estimada	35
4.3.2	Teste de hipótese para o modelo PIAR	36
4.3.3	Estimativas do modelo PIAR	37
4.3.4	Matriz acumulada de choques aleatórios	38
5	Diagnóstico do modelo autorregressivo periódico ajustado	39
5.1	Requisitos para a adequabilidade do modelo ajustado	39
5.2	Teste de hipótese para heterocedasticidade sazonal	40
5.3	Teste de Ljung-Box	40
5.4	Teste de Jarque-Bera	41
6	Previsão pelo modelo PAR	43
6.1	Método de previsão de modelos ARIMA	43
6.2	Método de previsão do modelo PAR	43
6.3	Previsão do modelo do exemplo	44
7	Modelo de partículas inaláveis	47
7.1	Partículas inaláveis	47
7.1.1	Estimativas do modelo PAR do exemplo	51
7.1.2	Diagnóstico do modelo	53
7.2	Modelo sazonal	53
8	Conclusões	57
A	Códigos do R	59

Lista de Tabelas

3.1	Seleção da ordem do modelo periódico autorregressivo.	29
4.1	Coefficientes do modelo PAR	34
4.2	Coefficientes do modelo PIAR.	38
7.1	Coefficientes do modelo autorregressivo periódico - PAR.	52
7.2	Coefficientes do modelo autorregressivo periódico diferenciado - PIAR . . .	54
7.3	Tabela de comparação de AIC e BIC dos modelos.	55

Lista de Figuras

2.1	Fluxograma da metodologia Box e Jenkins.	5
3.1	Gráfico do logaritmo do PIB real da Alemanha (1960-1990).	14
3.2	Gráfico sazonal do PIB da Alemanha (1960-1990).	15
3.3	Gráfico sazonal polar PIB da Alemanha (1960-1990).	16
3.4	Gráfico da subsérie sazonal do PIB da Alemanha (1960-1990).	17
3.5	Gráfico de dispersão entre os trimestres do PIB da Alemanha (1960-1990), nos lags 1 até 9.	18
3.6	Decomposição multiplicativa da série PIB da Alemanha (1960-1990).	19
3.7	Gráfico do logaritmo da série Z_t do PIB real da Alemanha (1960-1990).	20
3.8	Representação periódica e sazonal da série Z_t	20
3.9	Gráfico sazonal da série temporal diferenciada do PIB da Alemanha (1960- 1990).	21
3.10	Gráfico sazonal polar do PIB da Alemanha (1960-1990).	21
3.11	Gráfico da subsérie sazonal do PIB da Alemanha (1960-1990).	22
3.12	Gráfico de dispersão entre os trimestres do PIB da Alemanha (1960-1990), nos lags 1 até 9.	23
3.13	Decomposição multiplicativa da série Z_t do PIB da Alemanha (1960-1990).	24
3.14	Gráficos da função de autocorrelação e da função de autocorrelação parcial, respectivamente.	27
6.1	Previsão e intervalos de confiança do PIB da Alemanha (1991-1996)	45
7.1	Gráfico das partículas inaláveis da Enseada de Suá (2014-2015).	48
7.2	Gráfico sazonal das partículas inaláveis da Enseada de Suá (2014-2015).	49
7.3	Boxplot das partículas inaláveis da Enseada de Suá (2014-2015).	49

7.4	Gráfico de dispersão entre os dias das partículas inaláveis da Enseada de Suá (2014-2015).	50
7.5	Decomposição multiplicativa da série temporal das partículas inaláveis da Enseada de Suá (2014-2015).	51
7.6	Gráfico do modelo sazonal das partículas inaláveis da Enseada de Suá (2014-2015)	54

Capítulo 1

Introdução

Diversas pesquisas foram realizadas, a partir do ano de 1960, tendo como foco a construção do modelo periódico autorregressivo - PAR. De acordo com Hipel e McLeod (1994), podemos citar as pesquisas de Gladyshev (1963), Jones e Brelsford (1967), Tao e Delleur (1976), entre outros. A principal característica que faz a escolha deste tipo de modelo ideal é a de conter uma estrutura sazonal na série temporal e, muitas vezes, a correlação entre observações vizinhas, x_t e x_{t-1} , dependem apenas do tempo t . Portanto, sendo necessário o estudo prévio da estrutura da série para que o modelo ajustado seja o mais adequado.

O interesse inicial para a utilização deste tipo de modelo era o de poder fazer previsões para melhorar o desempenho do sistema hidrológico aumentando os benefícios e reduzindo os custos, utilizando para isto a relação de dependência mencionada. Como exemplo, atualmente, ele foi aplicado à geração de cenários de energias e vazões no Cepel (2018), Centro de pesquisas em energia elétrica vinculado às empresas Eletrobras. conjuntamente com outros tipos de modelos para as series temporais hidrológicas geradas.

Outro uso deste tipo de modelo, que é interessante mencionar, foi o de construção de um modelo para a poluição diária da cidade de São Paulo, feita por Sarnaglia *et al.* (2015), em que o objetivo era o de avaliar a qualidade do ar e de fazer previsões de concentrações dos poluentes presentes na série. Neste estudo, será utilizada uma série de poluentes, as partículas inaláveis, em que será feita a comparação do modelo autorregressivo periódico e o modelo sazonal.

O objetivo deste trabalho é o de apresentar detalhadamente o modelo autorregressivo periódico e fazer uma comparação com o modelo sazonal, conjuntamente com exemplos, explorando, para isso, a implementação do pacote *partsm* no software R.

Este estudo está organizado da seguinte maneira, no capítulo 2, apresenta-se os conceitos iniciais de séries temporais; no capítulo 3, estão os detalhes do modelo autorregressivo periódico com o auxílio de um exemplo. Métodos de estimação dos parâmetros desse modelo estão apresentados no capítulo 4. No Capítulo 5, o diagnóstico do modelo é apresentado e no capítulo seguinte como fazer a previsão do modelo PAR integrado. Por fim, no capítulo 7, um exemplo de série temporal real é utilizado para a construção de tal modelo e uma comparação com o modelo sazonal é realizada.

Capítulo 2

Conceitos iniciais

Neste Capítulo, serão abordados alguns dos conceitos da análise de séries temporais que serão utilizados para um maior entendimento acerca do modelo PAR que será abordado futuramente.

2.1 Séries temporais

Uma série temporal, de acordo com Souza e Camargo (2004), é definida como sendo um conjunto de observações de uma variável aleatória ordenada ao longo do tempo e com dependência das observações em distintos instantes de tempo. Em probabilidade, uma variável aleatória é tida como uma função mensurável X que associa um número real a cada evento elementar do espaço amostral.

Conforme Mendenhall e Sincich (2012), as séries temporais podem ser representadas pela combinação de alguns componentes, sendo eles, tendência, ciclo, sazonalidade e ruído aleatório - ruído branco.

Morettin e Toloí (2004) destacam alguns dos objetivos principais da construção de modelos para análise de séries temporais. Dentre eles está o auxílio na tomada de decisões, visando objetivos específicos, utilizando para este fim a previsão de valores futuros da série. Além disso, pode-se utilizar a análise de séries temporais para identificar periodicidades relevantes nos dados e descrever o comportamento da série temporal em questão.

1. Tendência: é responsável pelo sentido de deslocamento da série, sendo dada por mudanças graduais na série temporal a longo prazo, podendo ainda ter formato linear, parabólico, exponencial, entre outros.

2. Sazonalidade: corresponde a repetição de um certo padrão durante um período de tempo. São as oscilações, subidas e descidas, da série temporal.
3. Ciclo: é semelhante a sazonalidade, embora aconteça em um período de tempo maior e é menos regular que a componente sazonalidade.
4. Ruído branco: é caracterizado como a variabilidade dos dados que não pode ser explicada.

2.2 Estacionariedade

As séries temporais podem ser estacionárias ou não estacionárias. Segundo Morettin e Tolo (2004), a série temporal é estacionária se ela se desenvolve de forma aleatória independente do tempo, com algumas medidas estatísticas constantes, refletindo um equilíbrio estável. Grande parte dos modelos aplicados a séries temporais consideram suficiente a estacionariedade fraca, que será explicada a seguir.

Neste sentido, de acordo com Gujarati e Porter (2011), uma série pode ser considerada fracamente estacionária quando sua média e variabilidade forem independentes do tempo, e o valor da covariância entre dois períodos depender somente da distância do intervalo das observações. Cabe ressaltar que utilizaremos neste estudo a estacionariedade fraca. Embora também possa haver a estacionariedade forte, em que além destas características, a distribuição de um vetor de variáveis aleatórias que compõem a série temporal é invariante ao longo do tempo.

Uma série temporal é classificada como não estacionária quando possui média e/ou variabilidade das observações variando conforme o tempo. De acordo com Morettin e Tolo (2004), como a maioria das análises requerem algum tipo de estacionariedade, as transformações dos dados originais são bastante utilizadas. Uma das transformações mais usada consiste em tomar diferenças sucessivas da série original, até se obter estacionariedade da série.

2.3 Metodologia Box e Jenkins

A metodologia Box e Jenkins foi proposta em 1970 para a modelagem e previsão de valores futuros de séries temporais, dado o pressuposto de estacionariedade da série

temporal em questão. A metodologia baseia-se no ajuste de modelos autorregressivos integrados de médias móveis - ARIMA, que representa a combinação de três componentes, sendo eles, componentes autorregressivos, componentes de médias móveis e componentes de integração. A estratégia de construção do modelo final é baseada em um ciclo de ajustes, até a obtenção da melhor estrutura de modelo.

A Figura 2.1 mostra o fluxograma da estratégia de construção do modelo ARIMA proposta por Box e Jenkins. A identificação do modelo e a estimação dos parâmetros são as partes mais complicadas da construção do modelo, nos quais ocorrem a determinação dos valores p, d, q (componente autorregressivo, integração e de médias móveis, respectivamente) do modelo $ARIMA(p, d, q)$. Estes passos do ajuste são feitos utilizando, em geral, a Função de autocorrelação - FAC, e a Função de autocorrelação parcial - FACP, que serão expostas mais adiante, neste texto.

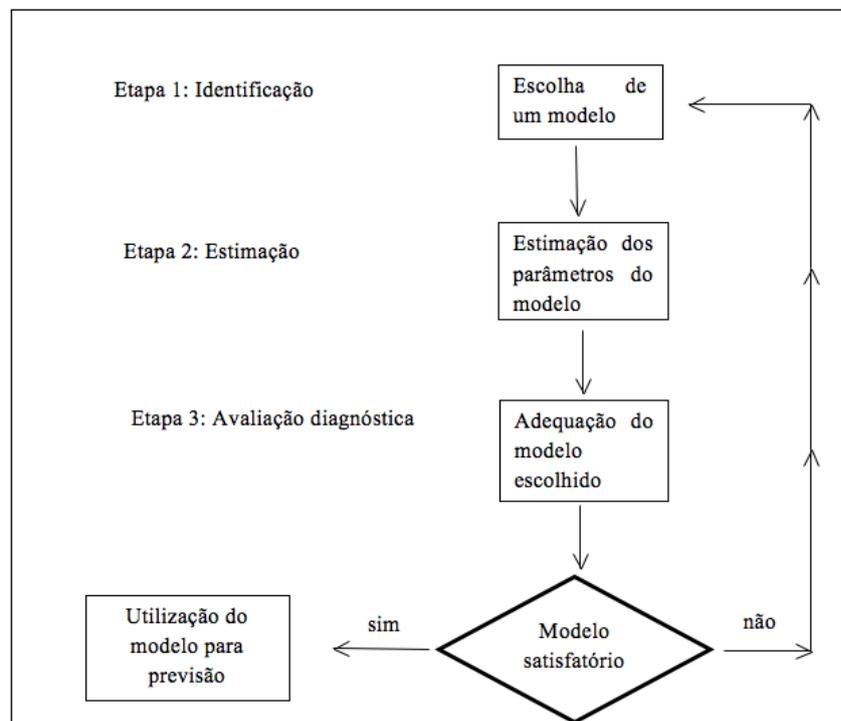


Figura 2.1: Fluxograma da metodologia Box e Jenkins.

Fonte: Adaptado de BOX et al. (1994), presente na Espacios. Vol. 35 (Nº 6) Ano 2014. Pág. 22.

Segundo Morettin e Tolo (2004), em geral os modelos obtidos com esta metodologia são parcimoniosos, pois contém um número pequeno de parâmetros e as previsões são precisas. Se o objetivo da construção do modelo for somente a previsão, será escolhido o modelo que apresente o menor erro quadrático médio de previsão, por exemplo.

2.3.1 Função de autocorrelação (FAC)

A função de autocorrelação - FAC, e a função de autocorrelação parcial - FACP são utilizadas para identificação da ordem do modelo e são estimadas a partir da amostra de valores utilizada na série temporal. Necessariamente tal amostra precisará ser estacionária, mesmo que seja fraca.

A função de autocorrelação é uma medida de correlação, ou seja, ela mede a dependência entre as observações de uma série temporal que são separadas por k unidades de tempo, ou seja, mede a correlação entre duas variáveis, X_t e X_{t+k} , com $k < t$, por exemplo.

Embora existam outras ferramentas para identificação da ordem do modelo, a FAC, ρ_K , é uma das estatísticas mais importantes para determinação do p , que pode ser definida como $\rho_k, k \in \mathbb{Z}$, tal que:

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad (2.1)$$

em que,

1. γ_k é a autocovariância, ou seja, é a covariância entre X_t e X_{t-k} e é dada por $\gamma_k = Cov[X_t, X_{t+k}] = E[(X_t - \mu_t)(X_{t+k} - \mu_{t+k})]$, sendo $\mu_t = E[X_t]$ e $\mu_{t+k} = E[X_{t+k}]$;
2. γ_0 é a variância dos dados, ou seja, $Var(X_t) = E[X_t - \mu_t]^2$.

2.3.2 Função de autocorrelação parcial (FACP)

Outra medida estatística utilizada com este mesmo objetivo é o coeficiente de autocorrelação parcial, FACP, que mede a correlação entre X_t e X_{t-k} , descontando a influência de $X_{t-1}, X_{t-2}, \dots, X_{t-k-1}$, com $t, k \in \mathbb{Z}$, com $k < t$, sendo também considerada importante para a definição da ordem do modelo que irá ser utilizado.

O coeficiente de autocorrelação parcial de ordem $k, k \in \mathbb{Z}$, denotado por ϕ_{kk} , é representado por,

$$\phi_{kk} = Cov(X_t, X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-(k-1)}). \quad (2.2)$$

A maneira mais usual para estimação dos coeficientes de autocorrelação parcial de um modelo ARIMA utiliza as equações de Yule-Walker, que na forma matricial são representadas por,

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}. \quad (2.3)$$

As equações podem ser resolvidas utilizando um procedimento recursivo, para ordens sucessivas, ou por meio de softwares estatísticos disponíveis. É possível também fazer o uso gráfico das duas medidas para facilitar a escolha do uso do modelo mais adequado.

2.4 Modelo autorregressivo - AR

No modelo autorregressivo - AR(p), o valor presente da série temporal, X_t , é definido como uma combinação de valores passados da própria série temporal mais um ruído aleatório ϵ_t , em que t é o índice de tempo. A ordem p deste modelo corresponde ao número de termos anteriores ao X_t que são considerados no cálculo do próprio X_t . Este tipo de modelo se baseia na suposição de estacionariedade da série em questão, como já foi visto anteriormente.

O modelo autorregressivo de ordem p , AR(p), é expresso por,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t, \quad (2.4)$$

em que,

1. $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ são os p termos anteriores a X_t nos tempos $t-1, t-2, \dots, t-p$;
2. $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros reais;
3. $\epsilon_t \sim RB(0, \sigma^2)$, isto é, ϵ_t segue uma distribuição ruído branco (RB), com média zero e variância, σ^2 , independente do tempo t .

A função de autocorrelação, FAC, de um AR(p) é dada por:

$$\rho_t = \phi_1\rho_{t-1} + \phi_2\rho_{t-2} + \dots + \phi_p\rho_{t-p}, \quad (2.5)$$

em que,

1. $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros reais;
2. $\rho_{t-1}, \rho_{t-2}, \dots, \rho_{t-p}$ são as correlações da variável X_t no tempo t com a mesma variável nos tempos $t - 1, t - 2, \dots, t - p$, respectivamente.

O gráfico de uma função de autocorrelação de uma série $AR(p)$ pode decrescer exponencialmente ou ter o formato de senoides amortecidas. É interessante notar que, para a função de autocorrelação parcial de um modelo $AR(p)$, os p coeficientes diferentes de zero indicam a ordem do modelo autorregressivo correspondente.

2.5 Critério de informação Akaike (AIC)

A partir do ano de 1970, vários critérios de identificação de ordem dos modelos foram propostos, tendo como base uma função penalizadora. Um deles foi proposto por Akaike (1974) para os modelos de séries temporais. Para o caso de um modelo $AR(p)$, por exemplo, o critério AIC, segundo Morettin e Tolo (2004), passou por uma atualização por Akaike (1979), e apresenta-se como,

$$AIC_\alpha(p) = N \ln(\hat{\sigma}_p^2) + \alpha P, \quad (2.6)$$

em que, além dos parâmetros anteriores, temos a adição de α , que é uma constante na equação, e P é um valor tal que $0 < p < P$, ou seja, valor superior para o cálculo da ordem p . Dependendo do valor de P , muitos modelos terão que ser ajustados, até que se obtenha o mínimo valor de AIC.

Segundo Morettin e Tolo (2004), Tong (1977) apresenta um resultado interessante sobre a constante α . Assintoticamente, a probabilidade da ordem correta ser selecionada, quando minimiza-se $AIC_\alpha(p)$, aumenta quando o termo α cresce.

2.6 Critério de informação Bayesiano (BIC)

Outro critério proposto para a escolha do modelo mais adequado para uma série temporal foi o critério de informação Bayesiano sugerido conjuntamente por Akaike *et al.* (1977), Rissanen (1978) e Schwarz (1978), e é dado por,

$$BIC(p) = \ln(\hat{\sigma}_p^2) + p \frac{\ln(N)}{N}, \quad (2.7)$$

em que $\hat{\sigma}_p^2$ é a estimativa de máxima verossimilhança da variância residual do modelo $AR(p)$, e os demais parâmetros são os mesmos que os descritos no critério AIC.

Conforme visto em Morettin e Tolo (2004), Hannan (1980) mostra em seu artigo que sob determinadas condições, as estimativas da ordem do modelo ARMA obtidas por este critério são fortemente consistentes, isto é, as estimativas se aproximam do valor verdadeiro quando o tamanho da amostra da série aumenta.

Capítulo 3

Modelo autorregressivo periódico

Neste capítulo, são abordados na Seção 3.1 a estrutura e componentes do modelo periódico autorregressivo e da sua forma integrada. Na Seção 3.2, são apresentados gráficos que são obtidos a partir dos dados, com um exemplo, para um maior entendimento do modelo estudado. Na Seção 3.3, estarão os procedimentos necessários para a identificação da ordem de um modelo periódico autorregressivo. Na Seção 3.4, será realizada a aplicação dos procedimentos na identificação da ordem do modelo da série temporal do exemplo. Na Seção 3.5 será apresentado um teste para a verificação da periodicidade do modelo.

3.1 Definição do modelo

O modelo periódico autorregressivo - PAR é utilizado em séries temporais que apresentam um comportamento periódico. Ou seja, conforme mostrado em Hipel e McLeod (1994), esse tipo de modelo pode ser aplicado para obter análises de séries temporais que possuam uma estrutura de autocorrelação entre as observações, em que estas observações dependem do período de tempo observado. Sendo assim, tal modelo é utilizado em séries temporais não estacionárias.

O modelo $PAR(p_m)$, ou simplesmente $PAR(p)$, ajusta para cada período de tempo um modelo $AR(p)$, em que p representa a ordem do modelo, isto é, o número de termos autorregressivos do modelo em cada período m de tempo.

Matematicamente, tal modelo é dado por,

$$\left(\frac{X_t - \mu_m}{\sigma_m}\right) = \phi_1^m \left(\frac{X_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \left(\frac{X_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + \epsilon_t, \quad (3.1)$$

em que,

1. $X_t, X_t : t \in \mathbb{Z}$, é uma série sazonal de período m ;
2. t é o índice de tempo, $t = 1, 2, \dots, S$;
3. m corresponde a cada período, $m = 1, 2, \dots, s$ ($m = 12$ para séries mensais e $m = 4$ para séries trimestrais);
4. s é o tamanho do período;
5. μ_m é a média sazonal do período m , $m = 1, 2, \dots, s$;
6. σ_m é o desvio sazonal do período m , $m = 1, 2, \dots, s$;
7. ϕ_i^m é o i -ésimo coeficiente autorregressivo do período m ;
8. p_m é a ordem do período autorregressivo do período m ;
9. ϵ_t é a série de ruídos aleatórios independentes com média zero e variância σ_a^2 para cada período m .

De forma simplificada, o modelo PAR(p_m) calculado a partir de mínimos quadrados simples é dado por,

$$X_t = \alpha_{1m}X_{t-1} + \alpha_{2m}X_{t-2} + \dots + \alpha_{p_m}X_{t-p} + \epsilon_t, \quad (3.2)$$

em que, há a variável X_t em estudo, o ruído branco ϵ_t , e $m = 1, 2, \dots, s$ é a periodicidade da série temporal. Conseqüentemente, como há periodicidade na série, o modelo PAR tem como propriedade ajustar para cada um dos períodos da série temporal um modelo autorregressivo, utilizando para esta finalidade os parâmetros $\alpha_{1m}, \alpha_{2m}, \dots, \alpha_{p_m}$.

O modelo PAR(p_m), pode ser representado ainda como,

$$x_t - \alpha_m x_{t-1} = \beta_{1m}(x_{t-1} - \alpha_{m-1}x_{t-2}) + \dots + \beta_{(p-1)m}(x_{t-(p-1)} - \alpha_{m-(p-1)}x_{t-p}) + \epsilon_t, \quad (3.3)$$

que é chamada forma periodicamente diferenciada do modelo autorregressivo periódico, em que α_{pm} e β_{pm} são parâmetros.

3.1.1 Modelo autorregressivo periódico integrado - PIAR

De acordo com Franses *et al.* (1996), uma série temporal trimestral é tida como uma série periódica integrada de ordem 1 quando o filtro diferencial $(1 - \alpha_m B)$ é necessário para remover a tendência da série x_t , em que α_m são parâmetros sazonais, com a restrição de que $\alpha_1 \alpha_2 \alpha_3 \alpha_4 = 1$ e B é um operador passo atrás (Backward), que em sua forma mais simples é dado por $Bx_t = x_{t-1}$.

O modelo autorregressivo periódico pode ser obtido utilizando o método de mínimos quadrados não lineares, e é dado por,

$$X_t = \alpha_m X_{t-1} + \beta_m (X_{t-1} - \alpha_{m-1} X_{t-2}) + \epsilon_t, \quad (3.4)$$

em que ϵ_t é ruído branco, e sob a restrição de que $\prod_{i=1}^s \alpha_i = 1$ para $m = 1, \dots, s$ e $i = 1, 2, \dots, p$, onde m denota os períodos da série temporal. Interessante notar que para um modelo PIAR de ordem 1, pela Equação 3.4, o β é igual a zero.

É interessante observar que a restrição acima se estabelece em dois casos, quando $\alpha_m = 1$ ou $\alpha_m = -1$. Portanto a restrição, neste caso, pode envolver a raiz 1 ou a raiz sazonal -1.

Para verificar a hipótese de integração, é construído o seguinte teste de hipótese,

1. $H_0 : \alpha_m = 1$, para $m = 1, \dots, s - 1$;
2. $H_0 : \alpha_m = -1$, para $m = 1, \dots, s - 1$.

A estatística teste, segundo López-de Lacalle (2005), a ser utilizada segue uma distribuição $F(s - 1, n - k)$, em que n é o número de observações e k o número de regressores.

Assim, quando a primeira hipótese nula não é rejeitada, o modelo PAR contém raiz unitária 1, e o filtro diferencial periódico é somente $(1 - B)$, onde B é o operador passo atrás. Em contrapartida, quando a segunda hipótese nula não é rejeitada, o modelo PAR contém raiz unitária -1 e o filtro diferencial periódico é $(1 + B)$.

Nos dois casos, se aceita as hipóteses nulas, se diz que os dados se comportam como uma série sem integração, conhecida como PAR ou PARI, em que ambos filtros diferenciais

entram em colapso. Se ambas as hipóteses forem rejeitadas, então o modelo correspondente é chamado de modelo autorregressivo periódico integrado, PIAR e precisará de uma diferenciação.

3.2 Exemplo de aplicação do modelo PAR

Para demonstrar o uso do modelo autorregressivo periódico, foi realizada a escolha de utilizar o artigo López-de Lacalle (2005), pois nele há a aplicação do modelo em um conjunto de dados reais. Todos os processos citados nele foram refeitos no software estatístico R.

Os dados utilizados para aplicação são os de Produto Interno Bruto (PIB) trimestrais provenientes da Alemanha, entre os anos de 1960 a 1990, e é empregada a função logarítmica na série para melhor visualização e obtenção de estimativas, visto que a série apresenta grande variabilidade em seus valores.

Esta série temporal pode ser obtida através do pacote *partsm* do R e foi feita a utilização de tal pacote ao longo de todo o estudo. Alguns gráficos do pacote *ggplot2* do *software* R também serão utilizados para identificação e visualização de algumas características que serão discutidas previamente, e que posteriormente poderão refletir em alguns resultados obtidos na análise.

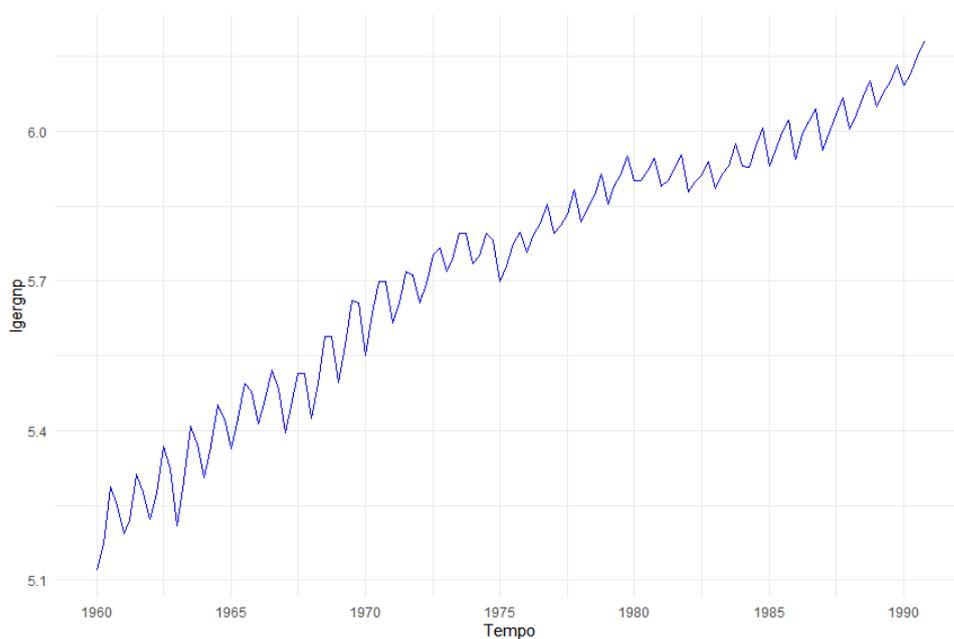


Figura 3.1: Gráfico do logaritmo do PIB real da Alemanha (1960-1990).

O gráfico da série temporal da Figura 3.1 apresenta tendência de crescimento do PIB da Alemanha conforme os anos passam, o que atende os requisitos do uso do modelo autorregressivo periódico, visto que a série não é estacionária.

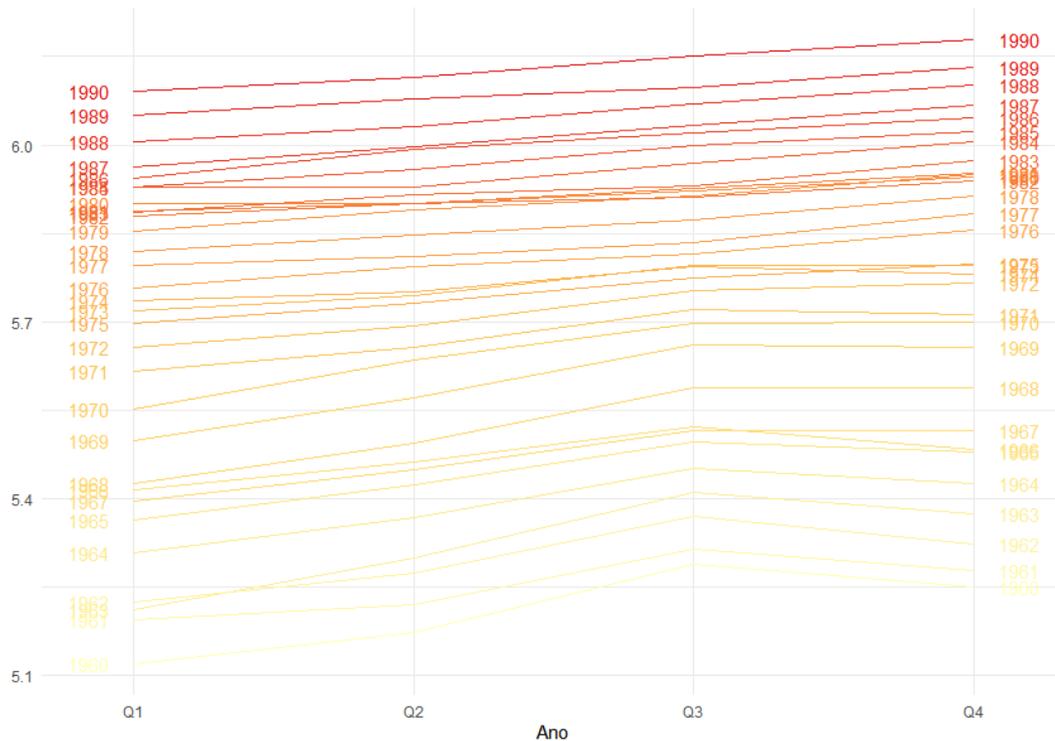


Figura 3.2: Gráfico sazonal do PIB da Alemanha (1960-1990).

Na Figura 3.2 está um gráfico sazonal de todos os anos separados por trimestre referentes ao PIB da Alemanha no período de tempo já mencionado. Com este gráfico é possível identificar se em determinado ano, o padrão encontrado nos dados passa por alguma mudança.

Neste caso em específico, apenas é possível concluir que há tendência na série, pois conforme visto na Figura 3.1, o PIB na Alemanha só aumenta com o decorrer dos anos. Não há, portanto, como diferenciar os trimestres entre si, em cada um dos anos distintos.

Outro gráfico interessante é uma variação do gráfico sazonal que utiliza coordenadas polares para a construção.

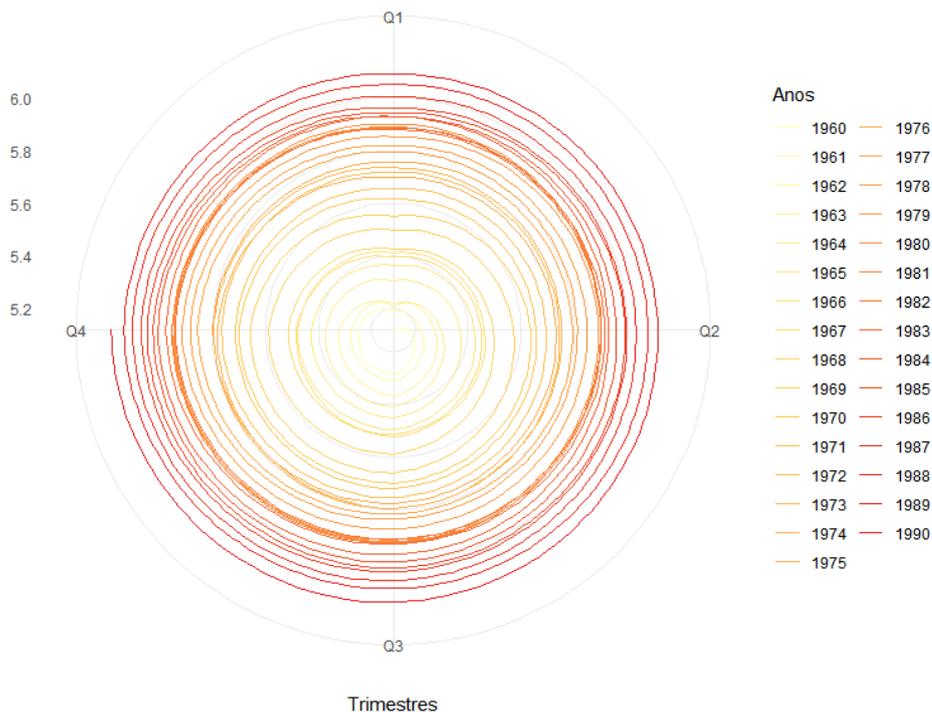


Figura 3.3: Gráfico sazonal polar PIB da Alemanha (1960-1990).

A Figura 3.3 também mostra que há um crescimento na série de PIB conforme o tempo passa e é possível perceber que não há como distinguir os trimestres entre si.

Na Figura 3.4, está representado o gráfico em que os dados são coletados juntos em gráficos menores separados por cada trimestre correspondentes. As linhas horizontais indicam as médias de cada trimestre. Esta forma de gráfico permite a visualização do padrão sazonal e também mostra as mudanças da sazonalidade ao longo do tempo. É especialmente utilizado para verificação de mudanças em determinados trimestres.

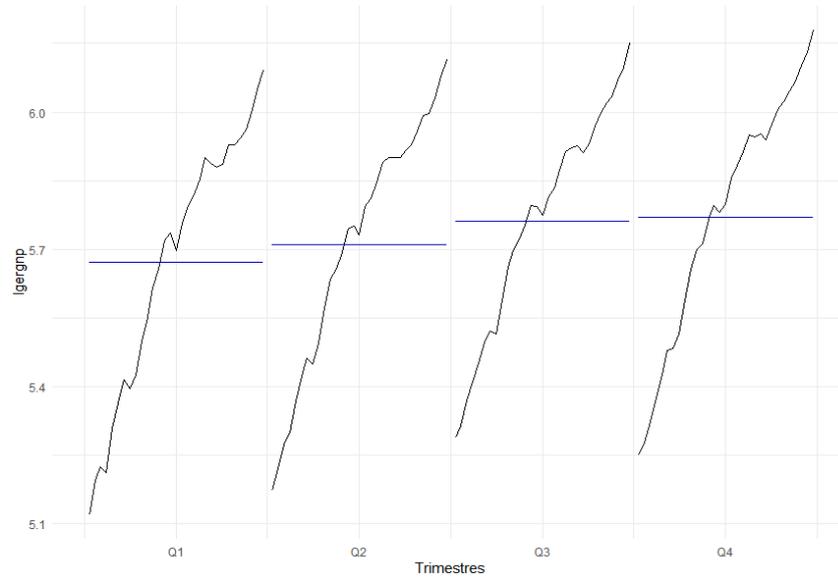


Figura 3.4: Gráfico da subsérie sazonal do PIB da Alemanha (1960-1990).

Devido à tendência, não há como afirmar que há diferença nos quatro trimestres utilizados da série temporal. Cada trimestre possui uma média de PIB semelhante aos demais, e os dados possuem aproximadamente o mesmo padrão.

Já a Figura 3.5 mede a relação direta entre os trimestres de cada ano no período acima citado, utilizando a dispersão dada pelos períodos de tempo. O eixo horizontal mostra os valores defasados da série temporal. Assim, cada gráfico dessa Figura mostra x_t contra x_{t-k} , para diferentes valores de lag k .

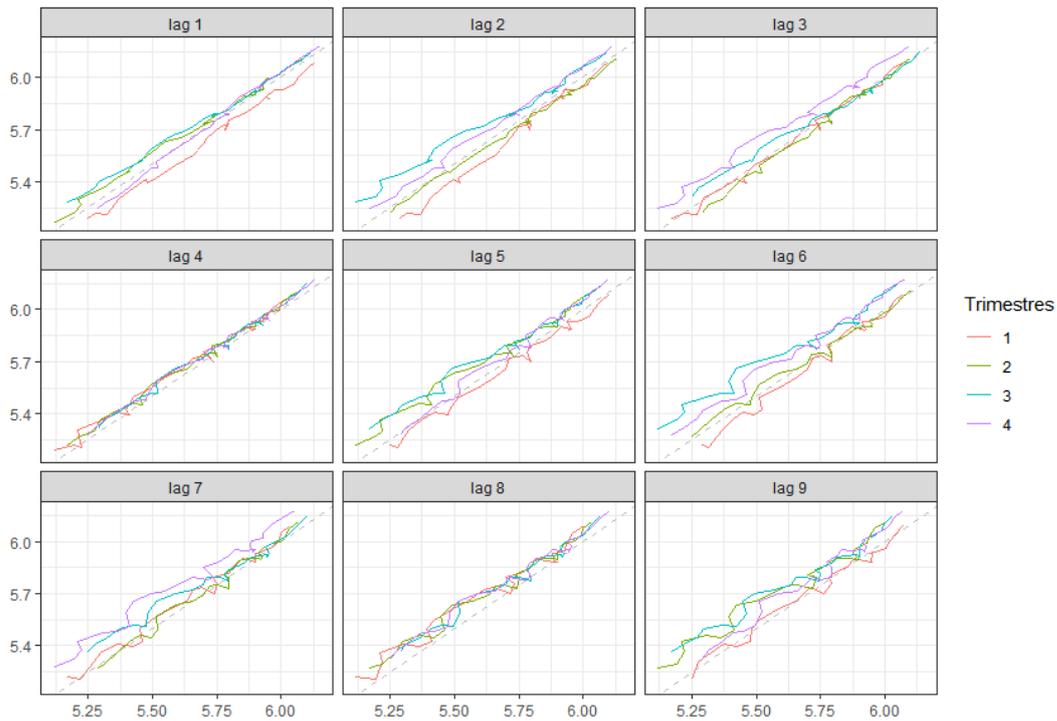


Figura 3.5: Gráfico de dispersão entre os trimestres do PIB da Alemanha (1960-1990), nos lags 1 até 9.

Cada cor de linha, presente nos gráficos, medem os valores de PIB dentro do trimestre e conectam pontos de dispersão em ordem cronológica. Entretanto, pode-se observar que não tem como diferenciar nenhum dos trimestres, pelo mesmo motivo já visto anteriormente. Há uma tendência forte na série, o que causa problemas, e trabalhar com a série diferenciada é uma solução.

Abaixo está a decomposição multiplicativa da série temporal, já que a variação temporal também cresce de forma sazonal.

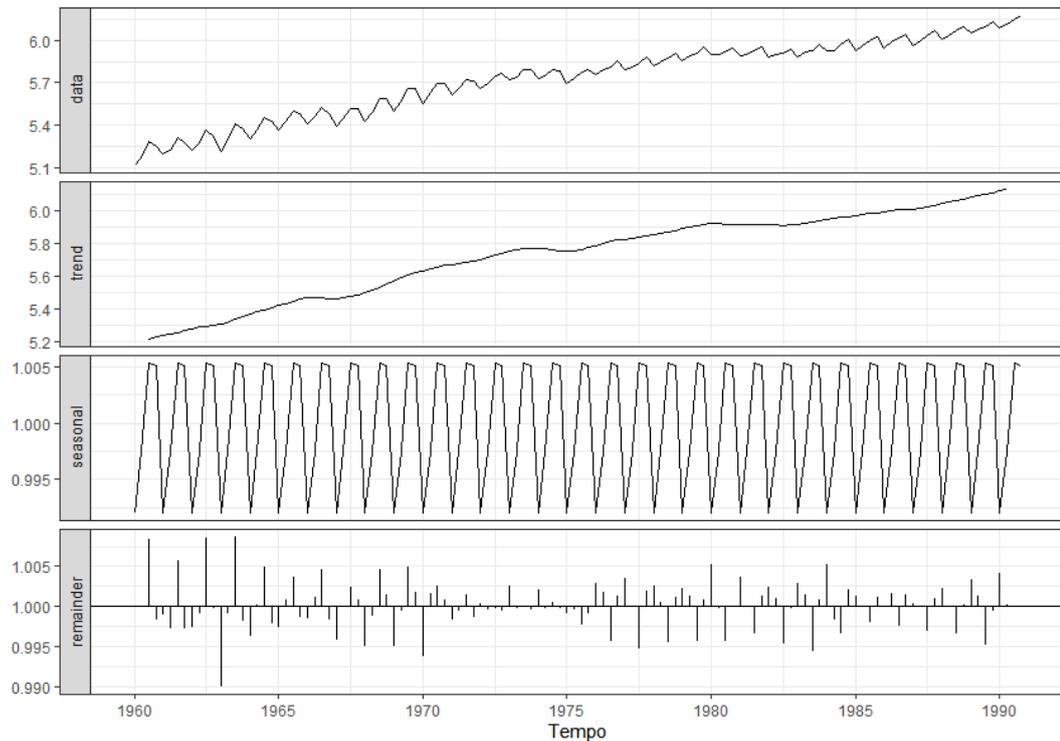


Figura 3.6: Decomposição multiplicativa da série PIB da Alemanha (1960-1990).

Na Figura 3.6, pode-se perceber que a componente tendência é bastante elevada e os resíduos assumem valores razoavelmente altos também. Podemos notar que há sazonalidade, apesar que ela apresenta uma magnitude menor que a tendência apresentada na série.

3.2.1 Série temporal diferenciada - Z_t

Aplicando uma diferenciação na série temporal, com a finalidade de excluir o efeito de tendência da série do PIB, obtém-se resultados diferentes dos anteriormente citados. Para facilitar o entendimento, chamaremos a série diferenciada de Z_t , ou seja, $Z_t = \text{diff}(\log(X_t))$.

Utilizando a mesma série de gráficos e o logaritmo da série temporal, realiza-se abaixo algumas conclusões acerca dos diferentes trimestres. No Capítulo 4, será apresentado como estimar os parâmetros desse modelo, e será apresentada conjuntamente tais estimativas.

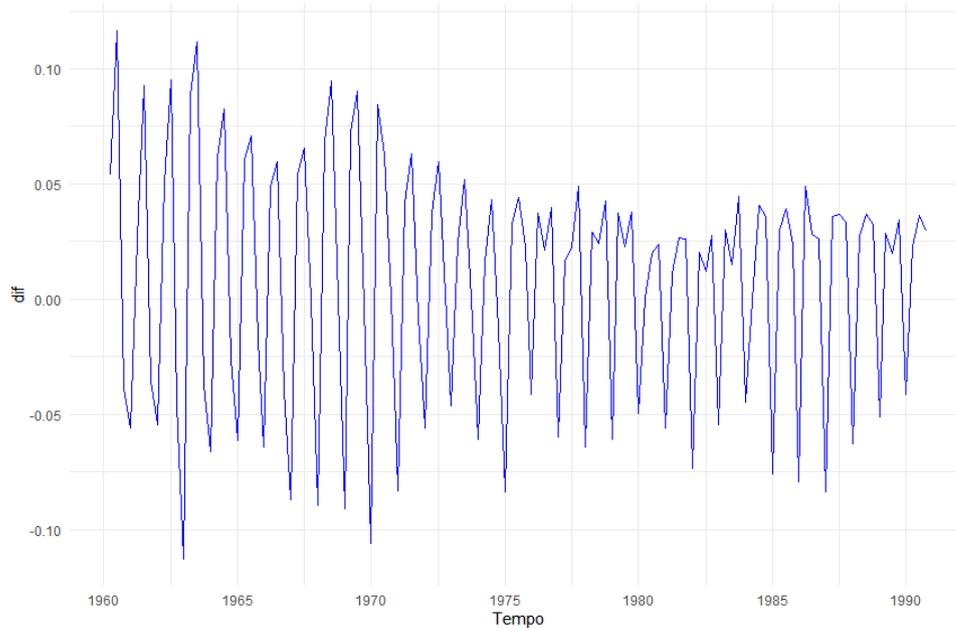


Figura 3.7: Gráfico do logaritmo da série Z_t do PIB real da Alemanha (1960-1990).

Pode-se perceber pela Figura 3.7 que a série Z_t não apresenta mais tendência e crescimento acentuado. De forma que ela permanece com valores em torno do valor zero, com o passar do tempo.

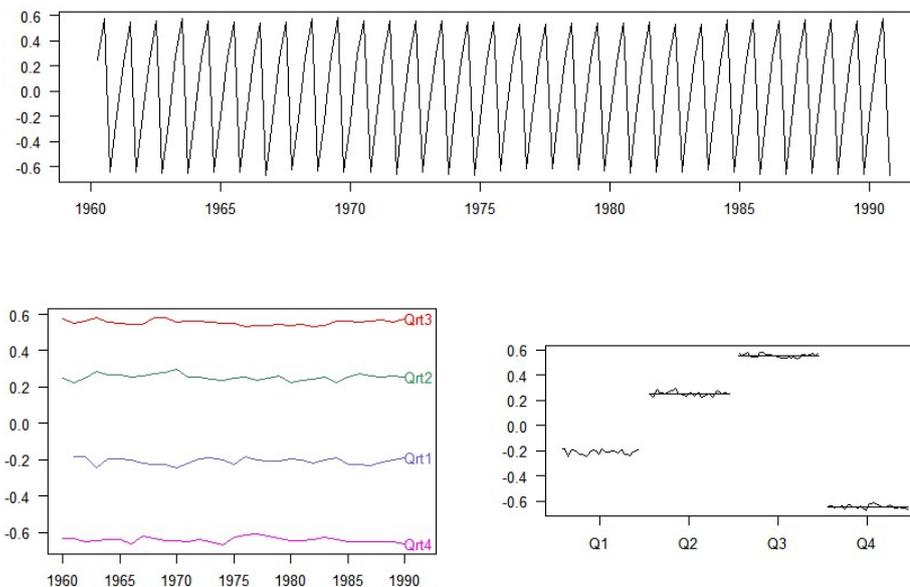


Figura 3.8: Representação periódica e sazonal da série Z_t .

Da representação da Figura 3.8, obtida a partir de Z_t , pode-se observar que o segundo e terceiro trimestre apresentam os maiores valores de PIB.

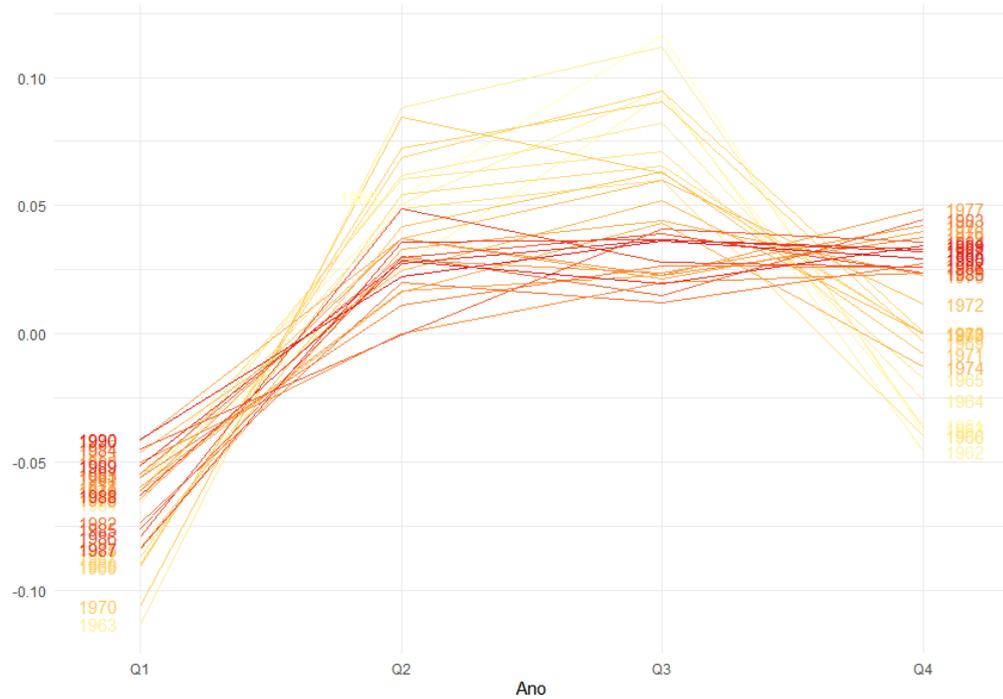


Figura 3.9: Gráfico sazonal da série temporal diferenciada do PIB da Alemanha (1960-1990).

Baseando na Figura 3.9, há evidência que os anos aparentam um padrão em que os segundos e terceiros trimestres apresentam valores maiores que os demais, e que o primeiro trimestre é o que apresenta menor valor de PIB em todos os anos.

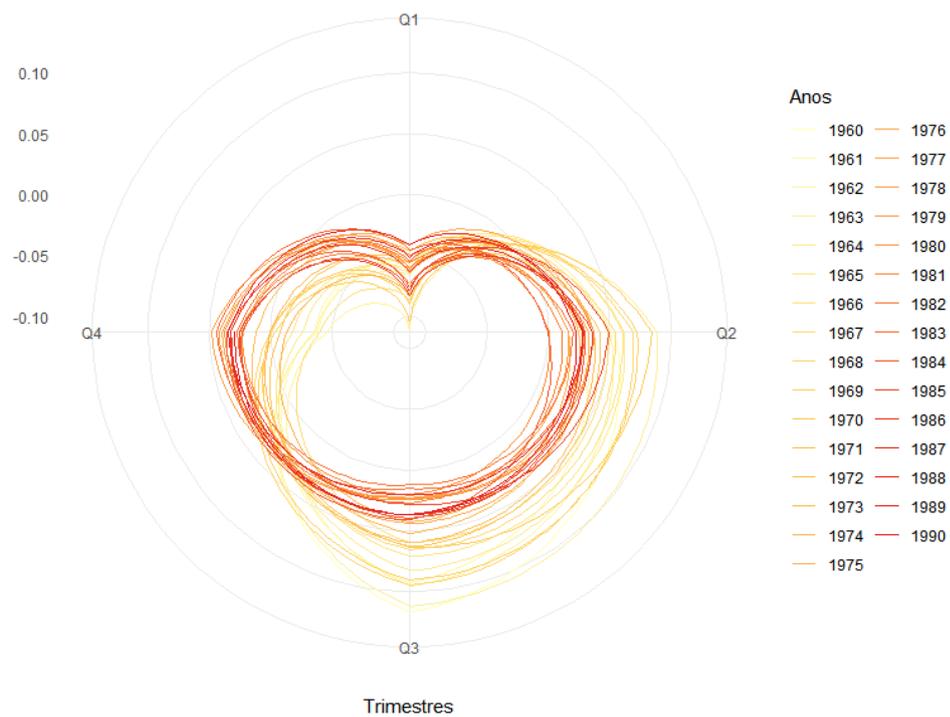


Figura 3.10: Gráfico sazonal polar do PIB da Alemanha (1960-1990).

Novamente, com a Figura 3.10, é possível perceber, pelas curvas do gráfico, que o segundo e o terceiro trimestre apresentam os maiores valores na grande maioria dos anos de PIB obtidos e o primeiro trimestre apresenta os menores valores de PIB. Os gráficos sazonais, Figura 3.9 e Figura 3.10, apresentam um padrão específico, em que, o PIB cresce do primeiro para o segundo trimestre e decai do terceiro trimestre até o último trimestre.

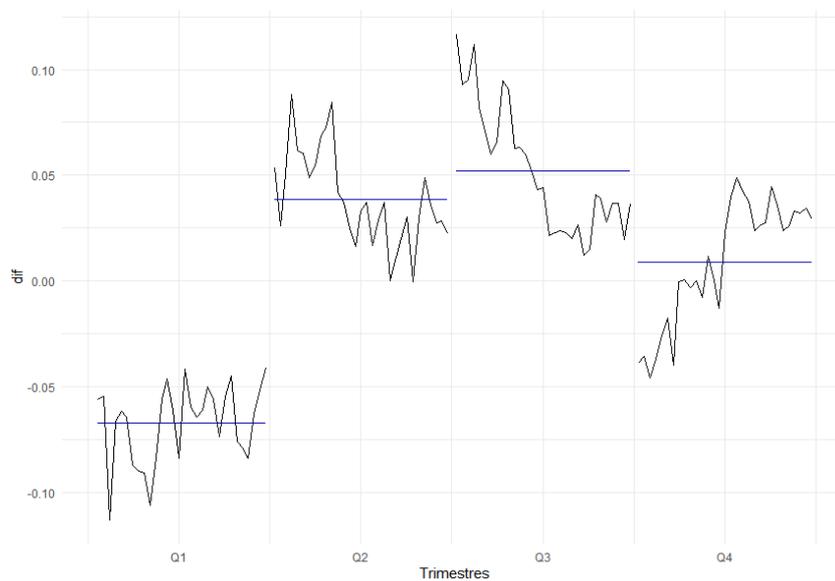


Figura 3.11: Gráfico da subsérie sazonal do PIB da Alemanha (1960-1990).

Por meio da Figura 3.11 observamos que, em média, o primeiro trimestre realmente apresenta menores valores que os demais trimestres, enquanto que os outros três trimestres apresentam em média valores semelhantes de PIB.

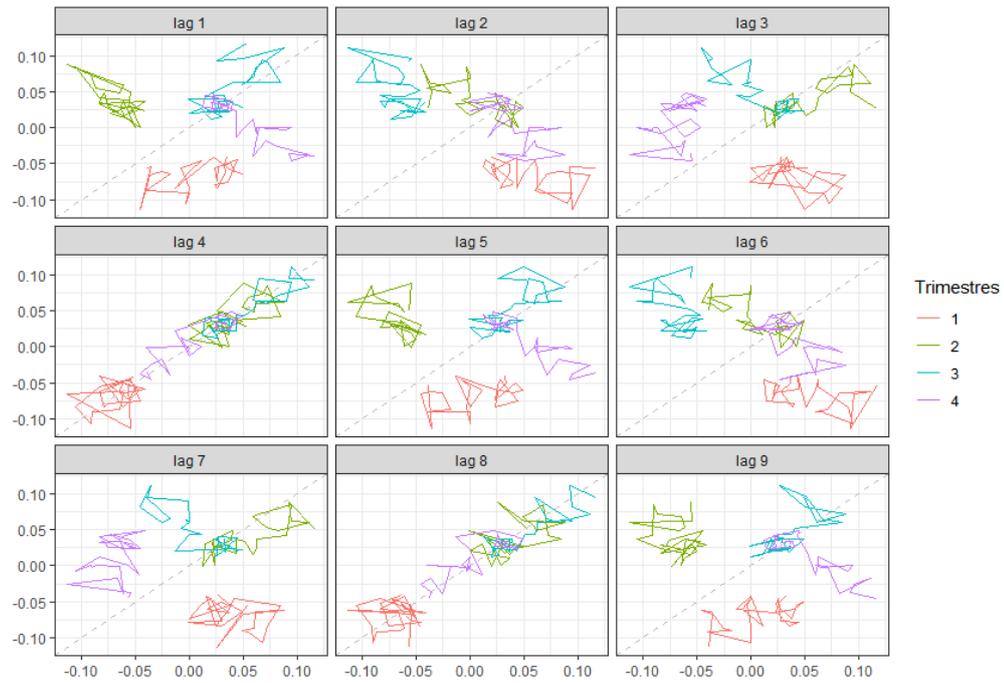


Figura 3.12: Gráfico de dispersão entre os trimestres do PIB da Alemanha (1960-1990), nos lags 1 até 9.

Na Figura 3.12, as linhas conectam os pontos de cada trimestre em ordem cronológica. É possível extrair algumas conclusões. A primeira delas é que o primeiro trimestre apresenta os menores valores em termo de dispersão nos períodos de tempo observados do PIB da Alemanha. Além disso, percebe-se um comportamento sazonal de ordem 4 na série, visto que depois do lag 4 há a repetição do padrão de dispersão entre os trimestres nos próximos lags.

Abaixo está a representação da decomposição multiplicativa da série Z_t .

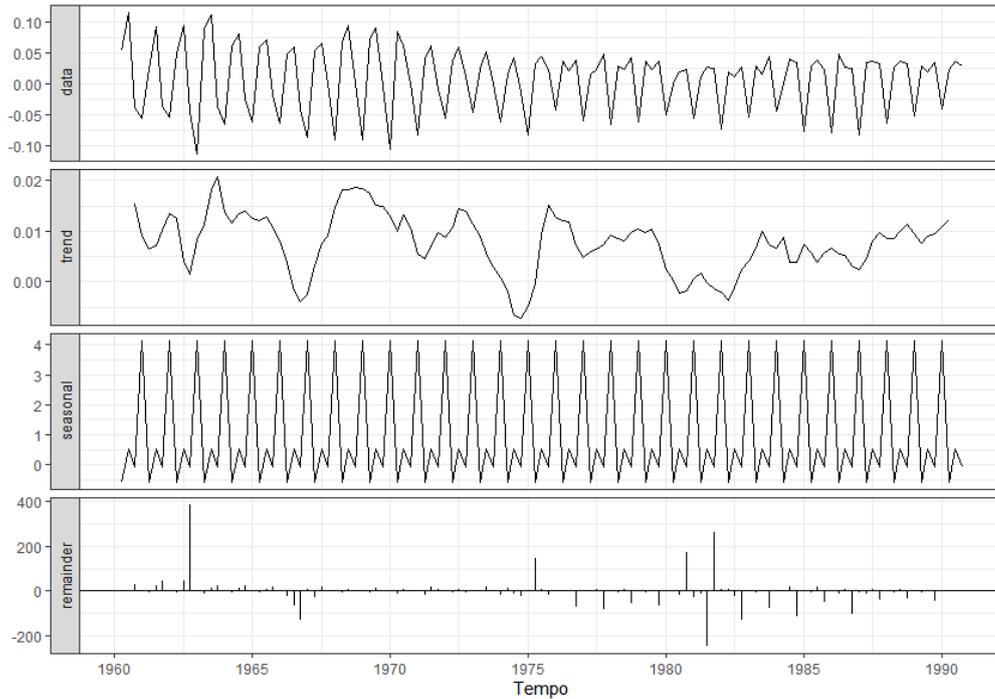


Figura 3.13: Decomposição multiplicativa da série Z_t do PIB da Alemanha (1960-1990).

Pela Figura 3.13, é possível perceber que não há mais tendência na série Z_t . A componente de sazonalidade recebeu uma maior dimensão, e os resíduos receberam valores menores em comparação com os obtidos anteriormente, quando utilizou-se a série temporal sem a diferenciação.

3.3 Identificação da ordem do modelo

De acordo com Maceira (1989), a identificação da ordem p , em cada um dos períodos m , dos modelos $PAR(p_m)$ também será feita por meio da função de autocorrelação e principalmente por meio da função de autocorrelação parcial, com o único diferencial que será calculado a ordem para cada período m de tempo.

Será utilizado um teste de hipótese por meio da função de autocorrelação parcial e serão considerados importantes o uso do critério de informação de Akaike (AIC) e o critério de informação bayesiano (BIC) para definição da ordem do modelo adequado.

3.3.1 Função de autocorrelação (FAC)

Para o cálculo da função de autocorrelação - FAC de um modelo PAR, primeiramente tem-se que ρ_k^m a correlação entre x_t e x_{t-k} , sendo que t corresponde especificamente ao

período m e σ_m o desvio padrão sazonal do período m ,

$$\rho_k^m = E\left[\left(\frac{X_t - \mu_m}{\sigma_m}\right) \cdot \left(\frac{X_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] = E[x_t x_{t-k}]. \quad (3.5)$$

As funções de autocorrelação da serie temporal, ρ_k^m , para cada $k > 1$,

$$\rho_k^m = \phi_1 E[x_{t-1} x_{t-k}] + \dots + \phi_{p_m} E[x_{t-p_m} x_{t-k}], \quad (3.6)$$

em que,

1. $\rho_0^m = 0$.

Conhecidos os parâmetros do modelo $\text{PAR}(p_m)$, as funções ρ_k^m podem ser calculadas pela Equação 3.6. Essas funções têm formas de decaimentos exponenciais ou ondas senoidais, o que faz com que em cada período m , ρ_k^m , tenda a zero a medida em que k aumenta.

3.3.2 Função de autocorrelação parcial (FACP)

Da mesma forma que no modelo autorregressivo, para determinar as funções de autocorrelação parcial do modelo $\text{PAR}(p_m)$ é necessário obter as equações de Yule-Walker. Para um período m qualquer fixado, tem-se que,

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \dots & \rho_{p_m-1}^{m-1} \\ \rho_1^{m-1} & 1 & \dots & \rho_{p_m-2}^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p_m-1}^{m-1} & \rho_{p_m-2}^{m-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k}^m \\ \phi_{2k}^m \\ \vdots \\ \phi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \vdots \\ \rho_{p_m}^m \end{bmatrix}. \quad (3.7)$$

O conjunto de valores ϕ_{kk} são as funções de correlação parciais do período m do modelo autorregressivo periódico. Portanto, a identificação do modelo consiste em identificar as ordens mais apropriadas dos operadores autorregressivos de cada período. Em um modelo $\text{PAR}(p_m)$ a função de correlação parcial, assim como no modelo $\text{AR}(p)$, será igual a zero para valores $k > p_m$ e diferentes de zero para valores de k menores que p_m .

3.3.3 Teste de hipótese

Utilizando as funções FAC e FACP é possível identificar a ordem do modelo $\text{PAR}(p_m)$ que deverá ser utilizada na série temporal estudada. Um dos meios de verificação é fazer uso de um teste de hipótese. Estabelece-se como hipótese nula, ou seja, a hipótese que irá ser testada, $H_0 : \phi_{kk} = 0$ para todos os valores de $k = 1, 2, \dots$, para assim testar a significância das estimativas no modelo.

Para realização do teste utiliza-se o resultado assintótico, que afirma que, se a ordem da estimativa autorregressiva de um período m é p_m , então o estimador $\hat{\phi}_{kk}$, obtido através das equações de Yule-Walker, de ϕ_{kk} , para $k > p_m$, segue uma distribuição aproximadamente normal com média 0 e variância $1/n$, sendo n o tamanho da amostra da série. Por consequência, não se rejeita H_0 quando zero estiver contido no intervalo $\hat{\phi}_{kk} \pm \frac{2}{\sqrt{n}}$. Neste caso, a estimativa $\hat{\phi}_{kk}$ não é significativa. O procedimento consiste em verificar para cada período m a maior ordem de p_m tal que todas as estimativas não sejam mais significativas.

Um dos riscos de usar tal método é que ele pode permitir a aceitação de processos como sendo ruído branco quando o processo pode não ser considerado assim. Por este motivo, Stedinger (2001) sugeriu o uso de um critério de defasagens j significantes, para encontrar a maior ordem j de cada período tal que todas as estimativas anteriores a j sejam significativas.

Para a construção da ordem do modelo que será utilizado no exemplo, será utilizada a hipótese nula como $\phi_{p+1} = 0$, que define quais parâmetros $p + 1$ serão iguais a zero. Serão considerados valores significativos das estimativas $\hat{\phi}_{p+1}$ além de serem considerados conjuntamente o Critério de informação Akaike e o Critério de informação Bayesiano para a escolha da ordem do modelo.

3.3.4 Critério de informação Akaike (AIC)

O critério de informação Akaike foi proposto por Akaike (1974) e é um dos procedimentos utilizados para a seleção da ordem dos modelos em geral. Tratando-se de modelos autorregressivos periódicos, o AIC é calculado pela Equação 3.8, para cada período de tempo m ,

$$AIC_m = n \log(\hat{\tau}_m^2) + 2p_m, \quad (3.8)$$

em que $\hat{\tau}_m^2$ são as estimativas dos resíduos para cada período m e p_m é a ordem do modelo. É selecionada as ordens que fazem com que seja mínimo o valor de $AIC = \sum_{m=1}^s AIC_m$.

3.3.5 Critério de informação Bayesiano (BIC)

Proposto por Akaike *et al.* (1977), Rissanen (1978) e Schwarz (1978), o critério de informação Bayesiano, BIC_m , para seleção de modelos foi adaptado para realização de seleção das ordens p_m dos modelos autorregressivos de cada período m ,

$$BIC_m = n \log(\hat{\tau}_m^2) + p_m \log(n), \quad (3.9)$$

o qual representa a ponderação entre as estimativas dos resíduos $\hat{\tau}_m^2$ e a ordem p_m do modelo, ou seja, é selecionado as ordens, tal que $BIC = \sum_{m=1}^s BIC_m$ seja mínimo.

3.4 Ordem do modelo do exemplo

Na Figura 3.14, estão os gráficos da função de autocorrelação - FAC, e da função de autocorrelação parcial - FACP, obtidos a partir da série temporal X_t , a série original, do exemplo.

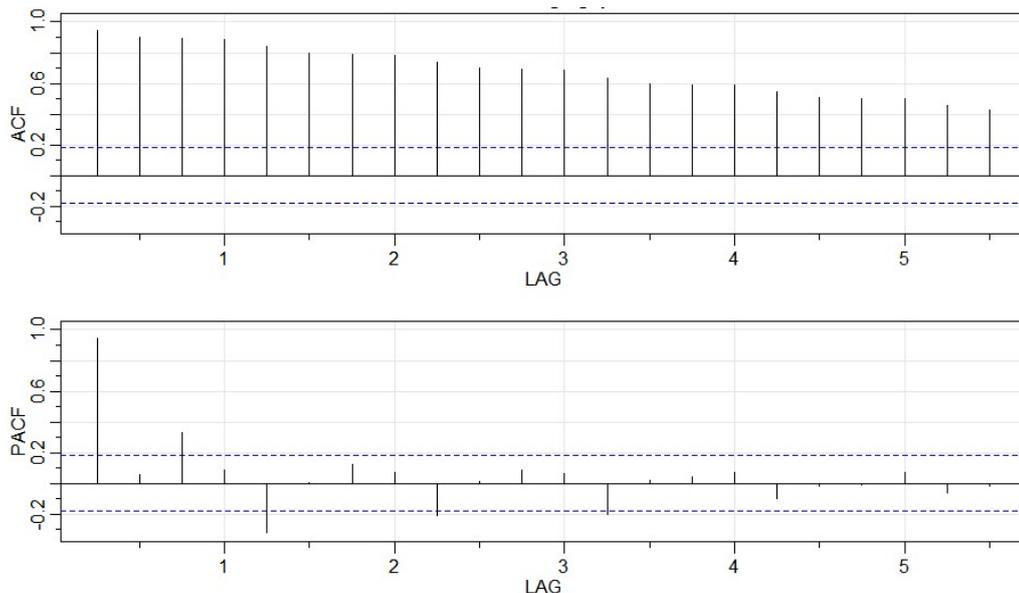


Figura 3.14: Gráficos da função de autocorrelação e da função de autocorrelação parcial, respectivamente.

Pode-se observar que a FAC decresce lentamente depois de alguns lags, isto é, períodos de tempo, e pelo gráfico da FACP pode-se notar que há correlações significativas nos primeiros lags, seguidas por correlações não significativas nos demais lags. As duas funções indicam que há termos autorregressivos a serem utilizados no modelo da série.

Na Tabela 3.1, estão os valores dos critérios correspondentes aos testes de hipótese, AIC e BIC, para até a quarta ordem do modelo periódico autorregressivo. Além disso, inicialmente foi já ajustado um modelo que inclui interceptos sazonais, ou seja, o modelo que será ajustado será trimestral, e é utilizado o logaritmo da série.

Tabela 3.1: Seleção da ordem do modelo periódico autorregressivo.

Critério	Ordem do modelo			
	1	2	3	4
AIC	-661,60	-680,89	-669,84	-661,54
BIC	-636,29	-644,44	-622,31	-603,00
F ($\phi_{p+1} = 0$)	8,54	0,80	1,35	2,91
Valor-p	0,00	0,53	0,26	0,03

De acordo com a Tabela 3.1, os parâmetros PAR de ordem 2 são significativos, além de que os critérios AIC e BIC correspondentes a esta ordem apresentam os valores mais baixos, então o modelo selecionado para as análises é o modelo autorregressivo de segunda ordem com interceptações sazonais.

3.5 Periodicidade no modelo

Como o modelo escolhido é o periódico, é interessante verificar, por meio de um teste de hipótese, se a série é realmente periódica autorregressiva, ou se somente um modelo autorregressivo seria o mais adequado a ser utilizado neste caso. Foi o procedimento que o autor López-de Lacalle (2005) realizou em seu artigo sobre a biblioteca do R chamada *partsm*.

Com base no modelo a seguir, já apresentado anteriormente, será construído o teste de hipótese, onde $m = 1, 2, \dots, s$ e m é a periodicidade da série temporal,

$$x_t = \alpha_{1m}x_{t-1} + \alpha_{2m}x_{t-2} + \dots + \alpha_{pm}x_{t-p} + \epsilon_t. \quad (3.10)$$

A hipótese nula do teste é $H_0 : \alpha_{im} = \alpha_i$, para $m = 1, 2, \dots, s$ e $i = 1, 2, \dots, p$. Ou seja, se a hipótese nula for aceita, o uso de um modelo autorregressivo será mais adequado. Se a hipótese for rejeitada, indica que há periodicidade na série, e um modelo periódico será mais adequado.

Para testar a periodicidade do modelo foi realizado a estatística teste, que é calculada, segundo López-de Lacalle (2005), com base na soma dos resíduos quadrados do modelo PAR(p) estimado. A estatística de teste que será utilizada, para esta finalidade, segue a

distribuição F com $((m-1)p, n - (m + mp))$ graus de liberdade, onde m é a periodicidade da série e n o número de observações.

No exemplo, tem-se como hipótese nula a não periodicidade do modelo, ou seja, ele se trata de um $AR(2)$, e como hipótese alternativa o modelo a ser utilizado é um $PAR(2)$. A estatística F é calculada com base na soma residual dos quadrados de cada modelo. Quando quatro interceptações sazonais são incluídas, a estatística segue um Distribuição F com $(3p, n - (4 + 4p))$ graus de liberdade, onde n é o número de observações.

O resultado do teste informa que a periodicidade do modelo não poderá ser rejeitada, pois o p-valor é bem abaixo do nível de significância de 5%, o que confirma que o modelo adequado é um autorregressivo periódico de ordem 2 com interceptos sazonais, como já foi indicado anteriormente.

Capítulo 4

Estimação dos parâmetros do modelo

Neste capítulo, na Seção 4.1, será abordado o procedimento de estimação dos parâmetros do modelo autorregressivo periódico utilizando as equações de Yule-Walker. Na Seção 4.2, será abordado um tipo diferente de estimação dos parâmetros, com representação matricial e a aplicação do procedimento matricial na série temporal do exemplo. Na Seção 4.3, será mostrada uma das formas de verificação da integração do modelo do exemplo, e as estimativas dos parâmetros do modelo integrado.

4.1 Método de Yule-Walker

Como já foi dito anteriormente, o modelo $PAR(p_m)$ é dado por,

$$\left(\frac{X_t - \mu_m}{\sigma_m}\right) = \phi_1^m \left(\frac{X_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \left(\frac{X_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + \epsilon_t. \quad (4.1)$$

Os parâmetros $\phi_i^m, i = 1, \dots, p^m$ do modelo são calculados a partir das equações de Yule Walker (Equação 3.7). Eles podem ser obtidos de maneira independente dos parâmetros de qualquer outro período. Cada uma das equações podem ser resolvidas a partir do método de Decomposição de Cholesky.

As estimativas de $\sigma_a^{(m)}$ podem ser obtidas da seguinte maneira: fazendo a substituição de ϕ_i^m por seus correspondentes estimadores,

$$\sigma_a^{(m)} = 1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m. \quad (4.2)$$

4.2 Método de representação matricial

Segundo López-de Lacalle (2005), em modelos periódicos autorregressivos de séries temporais trimestrais de ordem p menor ou igual a 4, o modelo com $m = 1, 2, 3, 4$, e ϵ_t é o ruído branco com variância constante e igual a σ_m^2 , para cada trimestre, pode ser reescrito como um modelo autorregressivo de ordem P , em que $P = 1 + [(p - 1)/4]$.

Para este objetivo, considera-se um modelo com vetor autorregressivo de 4 dimensões $X_T = (X_{1T}, X_{2T}, X_{3T}, X_{4T})$ e X_{sT} como a observação da variável no período s do ano T , com $T = 1, 2, \dots, N$. Tal equação é dada por,

$$\phi_0 X_T = \phi_1 X_{T-1} + \phi_2 X_{T-2} + \dots + \phi_P X_{T-P} + \epsilon_T, \quad (4.3)$$

em que $\epsilon_T = (\epsilon_{1T}, \epsilon_{2T}, \epsilon_{3T}, \epsilon_{4T})$, é o valor do ruído branco ϵ_{sT} no período s do ano T .

A Equação 4.3 pode ser reescrita na forma matricial utilizando a seguinte equação,

$$\Phi_0 X_T = \Phi_1 \mathbf{X}_{T-1} + \epsilon_T. \quad (4.4)$$

Em que as matrizes Φ_0 e Φ_1 são matrizes que contém os parâmetros estimados e são dadas por,

$$1. \phi_0(i, j) = \begin{cases} 1, & \text{se } i = j, \\ 0, & \text{se } j > i, \\ -\phi_{i-j, i}, & \text{se } j < i. \end{cases}$$

$$2. \phi_k(i, j) = \begin{cases} \phi_{i+4k-j, i}. \end{cases}$$

Com $i = 1, 2, 3, 4$; $j = 1, 2, 3, 4$; $k = 1, 2, \dots, P$ e $P = 1 + [(p - 1)/4]$.

Note que Φ_0 e Φ_1 são matrizes $S \times S$, ou seja, neste caso, 4×4 , já que $m = 1, 2, 3, 4$.

Com os parâmetros α_{is} estimados, elas são da maneira apresentada a seguir,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\alpha_{12} & 1 & 0 & 0 \\ -\alpha_{23} & -\alpha_{13} & 1 & 0 \\ -\alpha_{34} & -\alpha_{24} & -\alpha_{14} & 1 \end{bmatrix}, \quad (4.5)$$

e

$$\Phi_1 = \begin{bmatrix} \alpha_{41} & \alpha_{31} & \alpha_{21} & \alpha_{11} \\ 0 & \alpha_{42} & \alpha_{32} & \alpha_{22} \\ 0 & 0 & \alpha_{43} & \alpha_{33} \\ 0 & 0 & 0 & \alpha_{44} \end{bmatrix}. \quad (4.6)$$

Tomando como exemplo, as estimativas do modelo autorregressivo periódico de ordem 2 tem-se que as matrizes de estimativas serão dadas por,

$$x_t = \alpha_{1m}x_{t-1} + \alpha_{2m}x_{t-2} + \epsilon_t, \quad (4.7)$$

que pode ser reescrito como,

$$\Phi_0 X_T = \Phi_1 \mathbf{X}_{T-1} + \epsilon_T, \quad (4.8)$$

em que as matrizes estimadas de coeficientes serão iguais a,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\alpha_{12} & 1 & 0 & 0 \\ -\alpha_{23} & -\alpha_{13} & 1 & 0 \\ 0 & -\alpha_{24} & -\alpha_{14} & 1 \end{bmatrix}, \quad (4.9)$$

e

$$\Phi_1 = \begin{bmatrix} 0 & 0 & \alpha_{21} & \alpha_{11} \\ 0 & 0 & 0 & \alpha_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4.10)$$

4.2.1 Estimativas do modelo PAR do exemplo

Utilizando o pacote *partsm* do R e tendo como base um modelo autorregressivo periódico de ordem 2, temos que as matrizes das estimativas do exemplo são,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0,279 & 1 & 0 & 0 \\ 0,320 & -1,237 & 1 & 0 \\ 0 & -0,422 & -0,637 & 1 \end{bmatrix}, \quad (4.11)$$

e

$$\Phi_1 = \begin{bmatrix} 0 & 0 & -0,375 & 1,360 \\ 0 & 0 & 0 & 0,684 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4.12)$$

Portanto, os coeficientes do modelo autorregressivo periódicos, neste caso, são dados na Tabela 4.1,

Tabela 4.1: Coeficientes do modelo PAR

	m=1	m=2	m=3	m=4
α_{1m}	1,36	0,28	1,24	0,64
α_{2m}	-0,38	0,68	-0,32	0,42

4.3 Verificação de diferenciação no modelo do exemplo

A verificação da aplicação de diferenciação do modelo pode se dar previamente com base nos gráficos obtidos na Seção 2.4. Neles há evidência de que a série possui uma tendência e que para um modelo ser ajustado corretamente, é necessário lidar com ela antes.

Outra possível verificação se dá a partir dos autovalores gerados a partir da multiplicação das matrizes obtidas no modelo PAR ajustado. A confirmação de tal utilização pode ser feita a partir do teste de hipótese mencionado na Seção 3.1.1.

Em modelos periódicos derivados de ordem 2, com $m = 1, 2, 3, 4$, a equação do modelo se resume em,

$$x_t - \alpha_m x_{t-1} = \mu_s + \beta_m (x_{t-1} - \alpha_{m-1} x_{t-2}) + \epsilon_t. \quad (4.13)$$

As matrizes Φ_0 e Φ_1 de raízes do modelo para este caso são definidas abaixo como,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\alpha_2 & 1 & 0 & 0 \\ 0 & -\alpha_3 & 1 & 0 \\ 0 & 0 & -\alpha_4 & 1 \end{bmatrix}, \quad (4.14)$$

e

$$\Phi_1 = \begin{bmatrix} 0 & 0 & 0 & \alpha_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4.15)$$

4.3.1 Autovalores da matriz estimada

Os autovalores da matriz estimada são dados por Γ , que é a matriz construída a partir da multiplicação de $\phi_0^{-1} \phi_1$, dadas nas Equações 4.11 e 4.12.

Os valores dos autovalores da matriz estimada são $\lambda_1 = 0,965$, $\lambda_2 = 0,036$, $\lambda_3 = 0$ e $\lambda_4 = 0$. De acordo com este resultado há apenas um autovalor próximo de 1, portanto há evidências de que há uma raiz igual a 1 no modelo ajustado. Além disso, não há autovalores complexos, o que mostra evidências para esta mesma conclusão. É importante, portanto, fazer o teste de hipótese para verificar tal evidência de integração.

4.3.2 Teste de hipótese para o modelo PIAR

A verificação da evidência se dá por meio do teste de hipótese mencionado na Seção 2.1.1. Como visto anteriormente, na Seção 3.1.1, um modelo PIAR se dá quando há a restrição $\prod_{i=1}^s \alpha_i = 1$ para $m = 1, \dots, s$ e $i = 1, 2, \dots, p$, em que m denota os períodos da série temporal é satisfeita. Isso se dá quando $\alpha_1, \alpha_2, \alpha_3$ ou α_4 são iguais a 1 ou -1 .

Com base no modelo autorregressivo periódico, de ordem 2 e com interceptos sazonais, para $m = 1, \dots, s$,

$$x_t = \mu + \alpha_m x_{t-1} + \beta_m (x_{t-1} - \alpha_{m-1} x_{t-2}). \quad (4.16)$$

É construído o seguinte teste de hipótese,

1. $H_0 : \alpha_m = 1$, para $m = 1, \dots, s - 1$,
2. $H_0 : \alpha_m = -1$, para $m = 1, \dots, s - 1$.

Em resumo, quando ambas as hipóteses nulas são aceitas, se diz que a série trata-se de um modelo PAR, ou conhecido como PARI. E quando ocorre a rejeição delas, trata-se de um modelo autorregressivo periodicamente integrado - PIAR, e que precisará de uma diferenciação na série.

No caso do exemplo do PIB, tem-se as hipóteses a seguir, já que $m = 4$,

1. $H_0 : \alpha_m = 1$, para $m = 1, \dots, 3$,
2. $H_0 : \alpha_m = -1$, para $m = 1, \dots, 3$.

Utilizando o pacote *partsm* da biblioteca do R, obtém-se que ambas as hipóteses são rejeitadas por p-valores menores que 5%, portanto, trata-se de um processo autorregressivo periodicamente integrado - PIAR. É importante fazer tal verificação conjuntamente com os

gráficos, pois o modelo adequado neste caso, é o PIAR, ou seja é necessário a diferenciação na série, com a utilização de $Z_t = \text{diff}(\log(X_t))$, e novas estimativas dos parâmetros terão que ser calculadas a seguir.

4.3.3 Estimativas do modelo PIAR

Utilizando novamente o pacote *partsm* do R e tendo como base um modelo autorregressivo periódico integrado de ordem 2, temos que as matrizes das estimativas do exemplo são do mesmo formato que as matrizes das Equações 4.14 e 4.15, que neste caso são dadas por,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0,962 & 1 & 0 & 0 \\ 0 & -0,912 & 1 & 0 \\ 0 & 0 & -1,113 & 1 \end{bmatrix}, \quad (4.17)$$

e

$$\Phi_1 = \begin{bmatrix} 0 & 0 & 0 & 1,025 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4.18)$$

Com base no modelo autorregressivo periódico, de ordem 2 e com intercepções sazonais, para $m = 1, 2, 3, 4$ a seguir,

$$x_t = \mu + \alpha_m x_{t-1} + \beta_m (x_{t-1} - \alpha_{m-1} x_{t-2}). \quad (4.19)$$

Os coeficientes do modelo PIAR, neste caso, são, aproximadamente, dados pela Tabela 4.2.

Tabela 4.2: Coeficientes do modelo PIAR.

	m=1	m=2	m=3	m=4
α_m	1,025	0,962	0,912	1,113
β_m	0,338	-0,676	0,351	-0,429

4.3.4 Matriz acumulada de choques aleatórios

A matriz acumulada de choques aleatórios é construída com base na multiplicação das matrizes $\phi_0^{-1}\phi_1\phi_0^{-1}$. A linha que possui valores mais altos, indica o valor correspondente ao período s que sofre impactos mais severos no acúmulo de choques aleatórios. Consequentemente, é mais provável que essa linha exiba tendência estocástica na série temporal, e que flutuações sazonais também possam ser observadas a partir do uso de tal matriz.

Da mesma forma, a coluna com os valores mais altos está relacionada ao período que possui o maior impacto de choques a longo prazo.

Na Tabela 4.20 aparece a matriz acumulada de choques aleatórios dado pela multiplicação das matrizes ϕ_0 e ϕ_1 do modelo do exemplo, a partir do uso do pacote *partsm* do R. Por padrão, os valores iniciais são calculados para o modelo não linear. No entanto, nesta versão, pode haver casos em que as estimativas não convergem, dando uma mensagem de erro.

$$\begin{bmatrix} 1,000 & 1,040 & 1,140 & 1,025 \\ 0,962 & 1,000 & 1,097 & 0,985 \\ 0,877 & 0,912 & 1,000 & 0,898 \\ 0,976 & 1,015 & 1,113 & 1,000 \end{bmatrix}. \quad (4.20)$$

Considerando as colunas da matriz presentes na Equação 4.20, pode-se observar que a coluna que contém os valores mais altos é a terceira coluna, o que pode indicar que o terceiro trimestre de PIB é mais suscetível a mudanças na tendência estocástica na série temporal.

Capítulo 5

Diagnóstico do modelo autorregressivo periódico ajustado

Na Seção 5.1, será abordada a forma que será realizada a verificação de adequação do modelo autorregressivo periódico, baseado no diagnóstico deste modelo e acerca de alguns requisitos básicos que o modelo deverá ter. Na Seção 5.2, será apresentado um teste de hipótese para testar a heterocedasticidade sazonal dos resíduos do modelo ajustado. Na Seção 5.3, estará um teste para a hipótese de independência dos resíduos. Por fim, na Seção 5.4, estará presente um teste cuja hipótese nula é se os resíduos seguem uma distribuição Normal.

5.1 Requisitos para a adequabilidade do modelo ajustado

Para realizar a verificação de que o modelo autorregressivo periódico ajustado é adequado, uma série de testes de hipóteses são sugeridos. A maioria deles baseados na verificação da hipótese de que os resíduos ajustados são ruídos brancos, independentes e identicamente distribuídos, e que por vezes seguem uma distribuição Normal com média zero e variância constante. Além disso, eles não podem possuir nenhuma estrutura de dependência temporal.

5.2 Teste de hipótese para heterocedasticidade sazonal

Com o objetivo de verificar se os resíduos apresentam comportamento de heterocedasticidade sazonal no modelo ajustado, pode ser realizado um teste de hipótese. Apresentado por López-de Lacalle (2005), o teste é obtido do modelo pela Equação 5.1,

$$x_t^2 = \omega_0 + \omega_1 D_{1,t} + \dots + \omega_{S-1} D_{S-1,t} + \epsilon_t, \quad (5.1)$$

em que, m é a periodicidade dos dados e a hipótese nula testada é dada por $\omega_i = 0$ para $i = 1, \dots, m - 1$. A estatística teste, segundo López-de Lacalle (2005), segue uma distribuição F com $((m-1), n-k)$ graus de liberdade, em que n é o número de observações e k o número de parâmetros.

A heterocedasticidade sazonal do modelo ajustado é calculada no R e tem-se como resultado que ela é rejeitada com um p-valor aproximado de 0,044, o que nos mostra que não há uma heterocedasticidade sazonal entre os dados da série temporal tida no modelo do exemplo.

5.3 Teste de Ljung-Box

O teste de hipótese proposto por Ljung-Box é realizado com a finalidade de testar se os resíduos são independentes. As hipóteses são,

1. H_0 : os resíduos são independentes,
2. H_1 : os resíduos não são independentes e por isso os dados presentes na série temporal apresentam correlação entre si.

A estatística de teste é dada por,

$$Q(k) = n(n+2) \sum_{j=1}^K \frac{\hat{\rho}_j^2}{n-j}, \quad (5.2)$$

em que $\hat{\rho}_j^2$ são as estimativas de autocorrelação na defasagem j , n o tamanho da amostra e K é o número total de defasagens a serem testadas. Tal estatística possui

distribuição χ^2 com $K - p - q$ graus de liberdade, em que p e q são as ordens do modelo ajustado.

Realizando o teste de hipótese por meio do Software R, obtém-se que a hipótese não é rejeitada, com um p-valor aproximado de 0,3224 e, por isso, pode-se afirmar que os resíduos são independentes, não possuindo, portanto, uma estrutura de correlação entre os resíduos do modelo ajustado.

5.4 Teste de Jarque-Bera

O teste de hipótese de Jarque e Bera (1980), tem como hipótese nula a normalidade dos dados. Ele se baseia na diferença entre os coeficientes de assimetria e curtose dos dados. A estatística teste é dada por,

$$JB = n \left(\frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24} \right), \quad (5.3)$$

em que,

$$\alpha_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{ns^3}, \quad (5.4)$$

$$\alpha_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns^4}, \quad (5.5)$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (5.6)$$

Nos quais, \bar{x} é a média amostral e s^2 , α_3 e α_4 o segundo, terceiro e quarto momentos centrais respectivamente. A estatística teste, segundo López-de Lacalle (2005), tem distribuição assintótica χ^2 sob a hipótese nula.

Para o modelo dado no exemplo, apresentado na Seção 3.2, o resultado do teste tem como resposta a não rejeição da hipótese nula, com um p-valor próximo de 0,8291. Assim,

há evidências de que os resíduos seguem uma distribuição Normal.

Capítulo 6

Previsão pelo modelo PAR

Inicialmente, é comentado, na Seção 6.1, o método de previsão para modelos ARIMA. Na Seção 6.2, é apresentado como realizar o método de previsão do modelo autorregressivo periódico de forma geral e como encontrar o erro de previsão de tais estimativas. Já na Seção 6.3, é representado para o exemplo do PIB da Alemanha (1960-1990), as estimativas futuras com base em tal método de previsão.

6.1 Método de previsão de modelos ARIMA

Um dos maiores objetivos da análise de séries temporais, muitas das vezes, é a previsão para observações futuras de tal série. Resumidamente, segundo Morettin e Tolo (2004), na classe de modelos ARIMA, o método de previsão se dá tendo como objetivo a minimização do erro quadrático médio (EQM) de previsão. Ou seja, queremos que, $E \left\{ (X_{t+h} - \hat{X}_t(h))^2 \right\}$ seja o valor mínimo, sendo $\hat{X}_t(h)$ uma função linear das observações. Ver Morettin e Tolo (2004), que apresentam como são construídas as estimativas de EQM através da decomposição de Wold.

6.2 Método de previsão do modelo PAR

No modelo autorregressivo periódico a metodologia de cálculo para previsão se difere. Segundo López-de Lacalle (2005), o método de previsão do modelo autorregressivo periódico, deve ser realizado seguindo os seguintes passos. De acordo com a representação multivariada do modelo PAR, tem-se que,

$$\Phi_0 X_t = \Psi + \Phi_1 X_{T-1} + \dots + \Phi_P X_{T-P} + \epsilon_T. \quad (6.1)$$

Assim, como visto na Seção 5.2, $T = 1, 2, \dots, N$ é considerado um período maior de tempo e P é tal que $P = 1 + [(p - 1)/4]$, onde p é a ordem do modelo. Ψ se trata da matriz estimada de interceptos do modelo e Φ_i , com $i = 1, 2, \dots, P$, são as matrizes de dimensão $s \times s$ que contém os parâmetros do modelo estimados.

As estimativas diretas de um passo à frente, para o período $T + 1$, são obtidas pela Equação 6.2,

$$X_t = \Phi_0^{-1} \Psi + \Phi_0^{-1} \Phi_1 X_{T-1} + \dots + \Phi_0^{-1} \Phi_P X_{T-P} + \Phi_0^{-1} \epsilon_T. \quad (6.2)$$

As demais estimativas são obtidas de forma semelhante a esta, recursivamente. Os erros preditos para tal previsão serão aqueles que estarão na diagonal da matriz, a qual é dada por, $\sigma^2 \Phi_0^{-1} (\Phi_0^{-1})'$.

Utilizando $h = 2, 3, \dots$ como a representação da previsão de anos à frente, a equação de erros previstos fica representada como, $\sigma^2 \Phi_0^{-1} (\Phi_0^{-1})' + (h - 1)(\gamma \Phi_0^{-1})(\Gamma \Phi_0^{-1})'$, em que $\Gamma = \Phi_0^{-1} \Phi_1$.

6.3 Previsão do modelo do exemplo

A previsão utilizada no exemplo considera, como foi estudado e verificado, o modelo PIAR com ordem 2 para dados trimestrais. Por padrão, as intercepções sazonais são incluídas no modelo como componentes determinísticos. O número de observações que poderão ser previstas deverá ser múltiplo de 4 para satisfazer o período s observado.

Utilizando a função *predictpiar* do software R, presente no pacote *partsm*, é possível obter os coeficientes estimados futuros, os valores dos limites superiores e inferiores dos intervalos de 95% de confiança para tais estimativas previstas, que são obtidos da derivação multivariada da representação de média das matrizes para o modelo PIAR multivariado e os erros padrão encontrados em tal previsão.

Abaixo encontra-se o gráfico da previsão dos coeficientes da série temporal do PIB da Alemanha para os 24 períodos futuros,

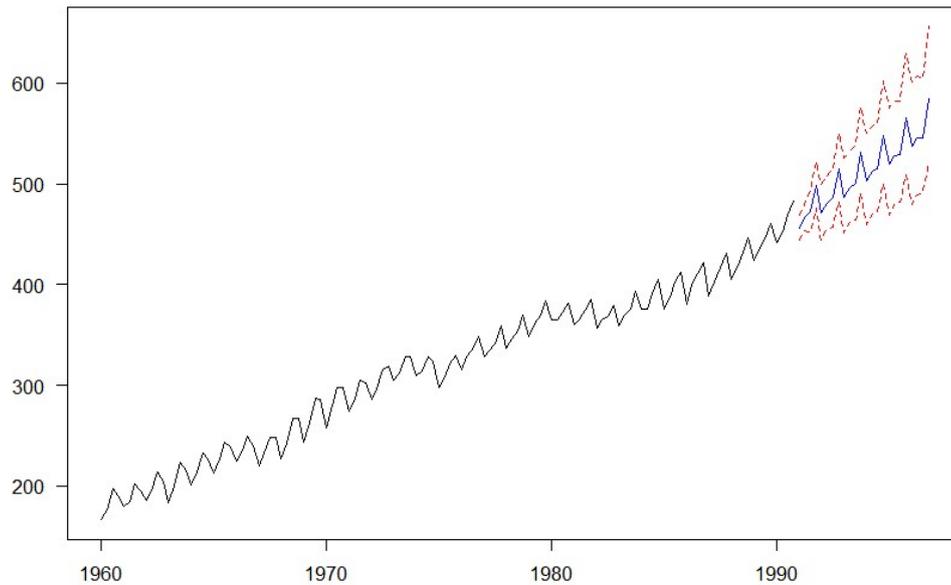


Figura 6.1: Previsão e intervalos de confiança do PIB da Alemanha (1991-1996)

Através da Figura 6.1 é possível observar, para o exemplo em questão, os valores previstos para os períodos seguintes, isto é, do período que vai do primeiro trimestre de 1991 até o quarto trimestre de 1996, conjuntamente dos intervalos de confiança, do PIB da Alemanha, baseado no método de previsão destacado anteriormente e nos períodos anteriores desta série de valores.

Nota-se que o PIB, nesta previsão, continuará subindo com o decorrer do tempo, até o quarto trimestre de 1996, mantendo certa periodicidade e tendência, apresentada anteriormente na série temporal dada em tal estudo.

Capítulo 7

Modelo de partículas inaláveis

Na Seção 7.1, foi ajustado um modelo autorregressivo periódico para uma série temporal de poluição. Na seção 7.2, está construído o modelo sazonal para esta mesma série e uma comparação entre os dois modelos.

7.1 Partículas inaláveis

Segundo CETESB (2021), qualquer substância presente no ar, e que possa oferecer risco à saúde, causando dano material, à fauna e a flora, são considerados poluentes. O nível de poluição é medido pela quantidade de poluentes presentes no ar, a variedade deles é tão grande que torna difícil a classificação. Existem poluentes classificados como primários, ou seja, aqueles que são emitidos diretamente pelas fontes de emissão e poluentes secundários, isto é, aqueles formados na atmosfera através de uma reação química entre poluentes primários e componentes naturais.

A concentração de poluentes está associada as condições meteorológicas. Um dos fatores que influenciam altos níveis de poluição, por exemplo, são ventos fracos, calma e inversão térmica.

Um dos elementos que constitui a poluição atmosférica são as partículas inaláveis. O diâmetro delas é menor do que 10 micrometros (μm) e elas causam efeitos negativos na saúde humana, penetrando no aparelho respiratório, provocando inúmeras doenças, como asma, bronquite e enfisema pulmonar.

Esta análise foi feita a partir dos dados de partículas inaláveis - MP_{10} medidos na estação RAMQAr de Enseada de Suá, Vitória. O período considerado de estudo foi de Janeiro de 2014 a Dezembro de 2015, retirando-se dados ausentes e utilizando as médias

das medidas horárias de concentrações em $\mu g/m^3$. Foi considerado o logaritmo da série temporal, para amenizar a variância presente nela.

O fator motivante deste estudo deve-se ao artigo escrito por Solci e Reisen (2017). Nele, os autores evidenciam a aplicação do modelo periódico autorregressivo nestes mesmo conjunto de dados de poluição. Entretanto algumas diferenças são percebidas na aplicação do modelo, o que pode acarretar em análises distintas. O objetivo do estudo desta série, deve-se ao fato de que é importante analisar a aplicação do modelo PAR em uma série temporal de valores reais e é interessante a realização de uma comparação com um tipo diferente de modelo existente.

A Figura 7.1 apresenta o comportamento da série temporal de partículas inaláveis - MP_{10} , com uma transformação logarítmica.

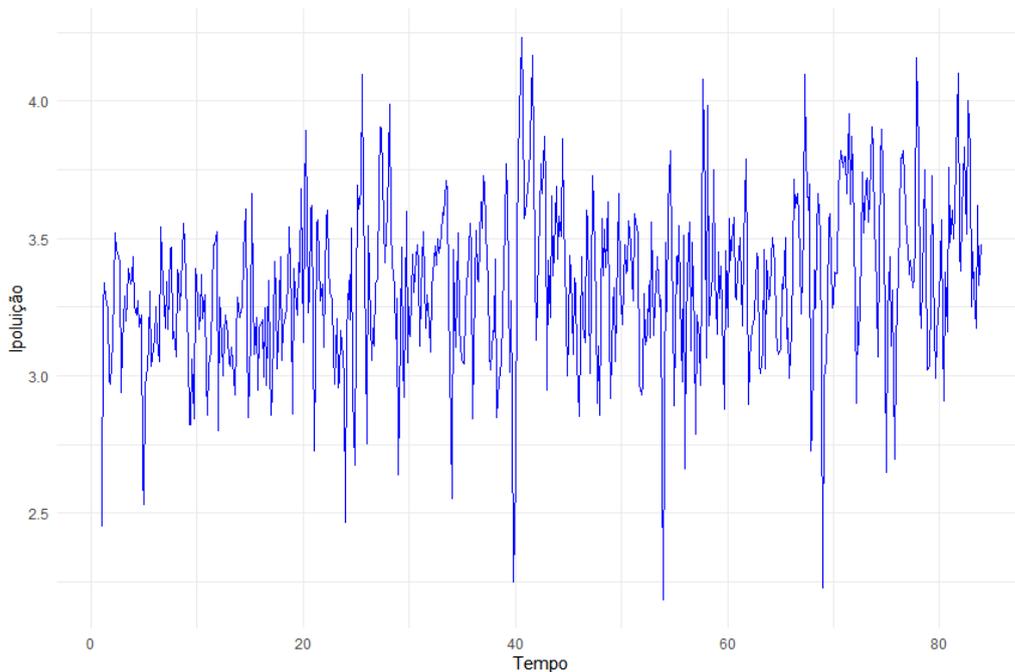


Figura 7.1: Gráfico das partículas inaláveis da Enseada de Suá (2014-2015).

Por meio dela é possível perceber que não há evidências de tendência crescente ou decrescente presente na série temporal de partículas inaláveis.

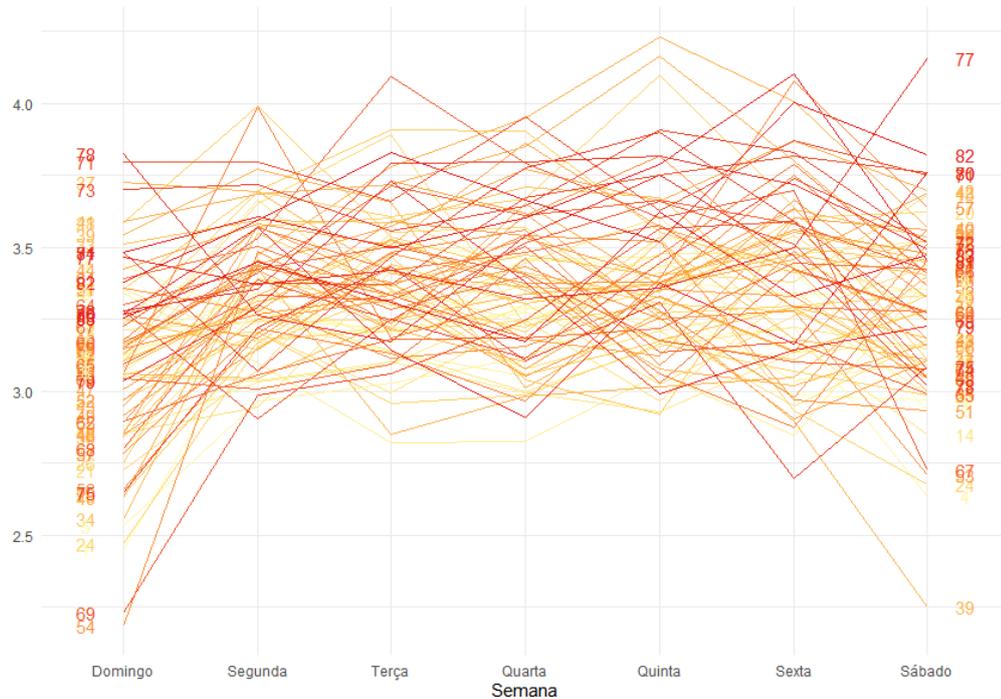


Figura 7.2: Gráfico sazonal das partículas inaláveis da Enseada de Suá (2014-2015).

Pela Figura 7.2, não há como distinguir qual dos dias há os menores valores, apenas que há uma variação de valores maior no sábado, e possível evidência de valores de partículas inaláveis maiores de quinta-feira. Para verificar se há indícios de diferença entre os dias, foi realizada a construção de boxplots com os dados semanais.

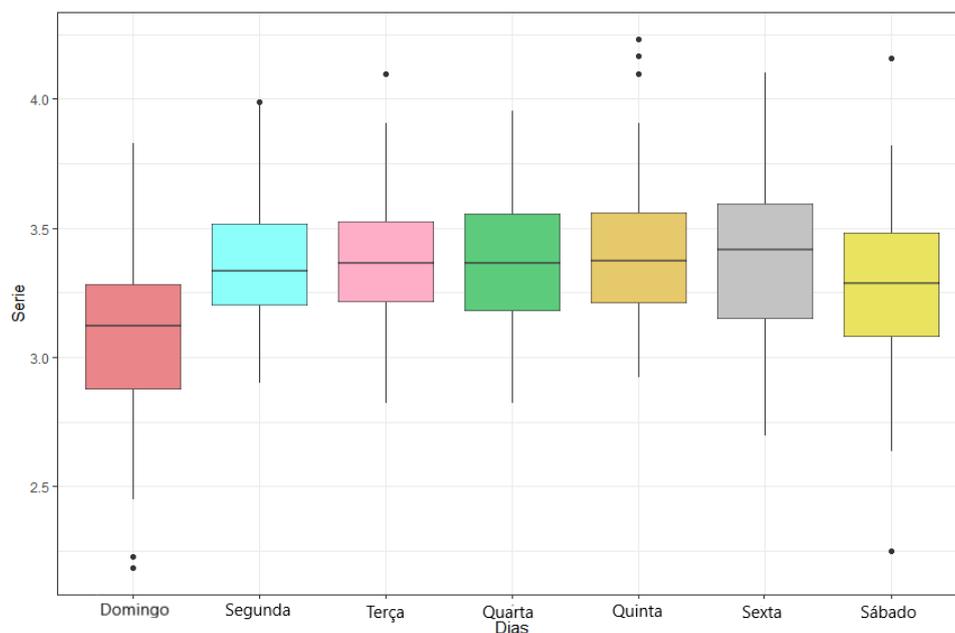


Figura 7.3: Boxplot das partículas inaláveis da Enseada de Suá (2014-2015).

Segundo a Figura 7.3, domingo aparenta ter os menores valores e mediana inferior aos

demais dias da semana.

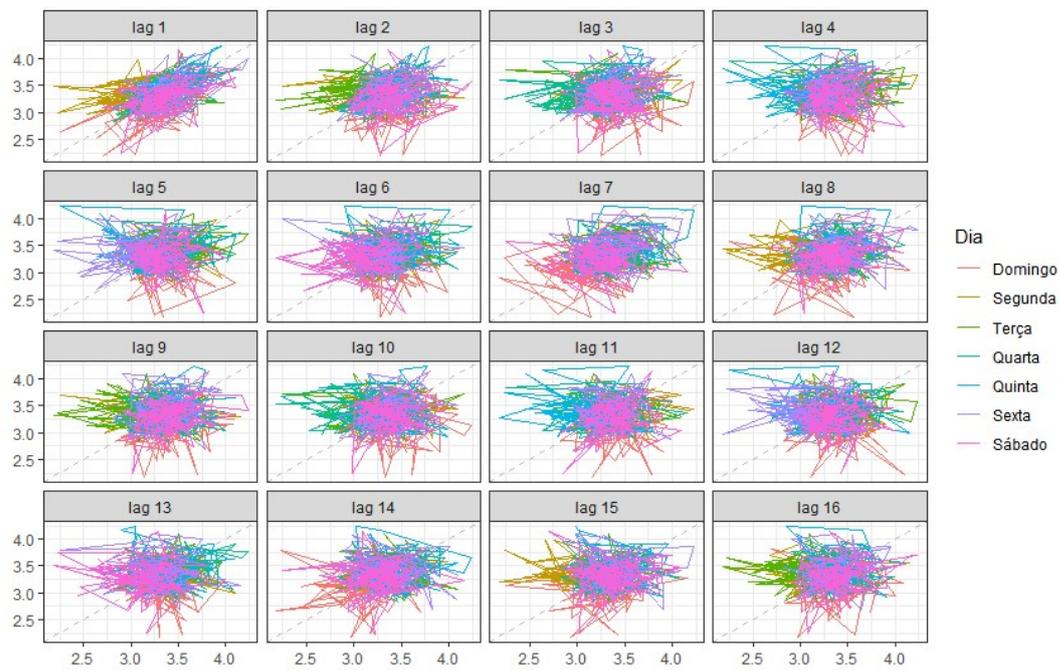


Figura 7.4: Gráfico de dispersão entre os dias das partículas inaláveis da Enseada de Suá (2014-2015).

Pela Figura 7.4, percebe-se um comportamento sazonal de ordem 7 na série, visto que depois do lag 7 há a repetição do padrão de dispersão entre os trimestres nos próximos lags.

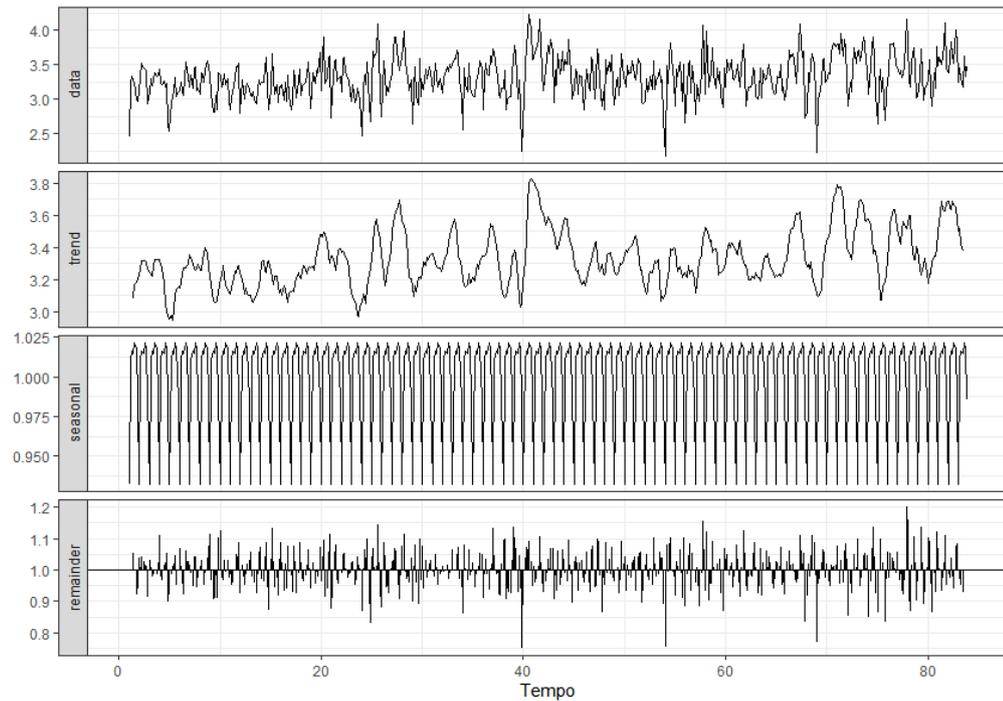


Figura 7.5: Decomposição multiplicativa da série temporal das partículas inaláveis da Enseada de Suá (2014-2015).

Pela Figura 7.5, é possível perceber que não há tendência na série do modelo. A componente de sazonalidade detém um padrão e os resíduos são pequenos.

7.1.1 Estimativas do modelo PAR do exemplo

Utilizando o pacote *partsm* do R e tendo como base um modelo autorregressivo periódico de ordem 1, temos que as matrizes das estimativas do exemplo são,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0,296 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0,466 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0,694 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0,63 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0,602 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0,46 & 1 \end{bmatrix}, \quad (7.1)$$

$$\Phi_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0,594 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (7.2)$$

Portanto, os coeficientes do modelo autorregressivo periódico, neste caso, aproximados pelas matrizes das Equações 7.1 e 7.2, são dados por,

Tabela 7.1: Coeficientes do modelo autorregressivo periódico - PAR.

	m=1	m=2	m=3	m=4	m=5	m=6	m=7
α_{1m}	0,59	0,3	0,47	0,69	0,63	0,6	0,46

Matriz acumulada de choques aleatórios

$$\begin{bmatrix} 0,010 & 0,034 & 0,072 & 0,104 & 0,165 & 0,273 & 0,594 \\ 0,003 & 0,010 & 0,021 & 0,031 & 0,049 & 0,049 & 0,176 \\ 0,001 & 0,005 & 0,010 & 0,014 & 0,023 & 0,038 & 0,082 \\ 0,001 & 0,003 & 0,007 & 0,010 & 0,016 & 0,026 & 0,057 \\ 0,001 & 0,002 & 0,004 & 0,006 & 0,010 & 0,016 & 0,036 \\ 0,000 & 0,001 & 0,003 & 0,004 & 0,006 & 0,010 & 0,022 \\ 0,000 & 0,001 & 0,001 & 0,002 & 0,003 & 0,005 & 0,010 \end{bmatrix} \quad (7.3)$$

Considerando as colunas da matriz presentes na Equação 7.3, pode-se observar que a coluna que contém os valores mais altos é a sétima coluna, o que pode indicar que sábado é mais suscetível a mudanças na tendência estocástica na série temporal.

7.1.2 Diagnóstico do modelo

Teste de hipótese para heterocedasticidade sazonal do modelo

A heterocedasticidade sazonal do modelo ajustado, sendo calculada pelo R, tem a hipótese nula do teste rejeitada, com p-valor aproximado de 0,03560248. Isto é, não há evidência de que há uma heterocedasticidade sazonal entre os dados da série temporal.

Teste de Ljung-Box

Realizando o teste de hipótese, obtém-se que a hipótese nula não é rejeitada com um p-valor aproximado de 0,7815. Ou seja, há evidências de que os resíduos são independentes, não possuindo uma estrutura de correlação entre eles.

Teste de Jarque-Bera

O resultado do teste de hipótese é dado pela não rejeição da hipótese nula, com um p-valor de 0,2089 há evidências de que os resíduos seguem uma distribuição normal.

7.2 Modelo sazonal

Segundo, Morettin e Tolo (2004), o modelo com um comportamento sazonal determinístico, por exemplo, de período 12 e com uma variável aleatória, X , pode ser escrito como

$$X_t = \mu_t + N_t, \quad (7.4)$$

em que, μ_t é uma função determinística periódica, que satisfaz $\mu_t - \mu_{t-12} = 0$ e N_t é um processo estacionário que pode ser modelado utilizando um $\text{ARMA}(p, q)$. Mais informações sobre o modelo sazonal poderão ser encontradas no livro do Morettin e Tolo (2004).

A principal diferença entre o modelo sazonal e o modelo PAR é que, no modelo periódico os parâmetros variam de forma periódica em vez de serem constantes para todos os instantes. Com o objetivo de comparar o modelo autorregressivo periódico construído e selecionar o melhor para a situação, alguns modelos sazonais foram ajustados. Dentre eles, o que obteve melhores resultados foi o modelo SARIMA $(1, 0, 0) \times (1, 0, 0)_{12}$.

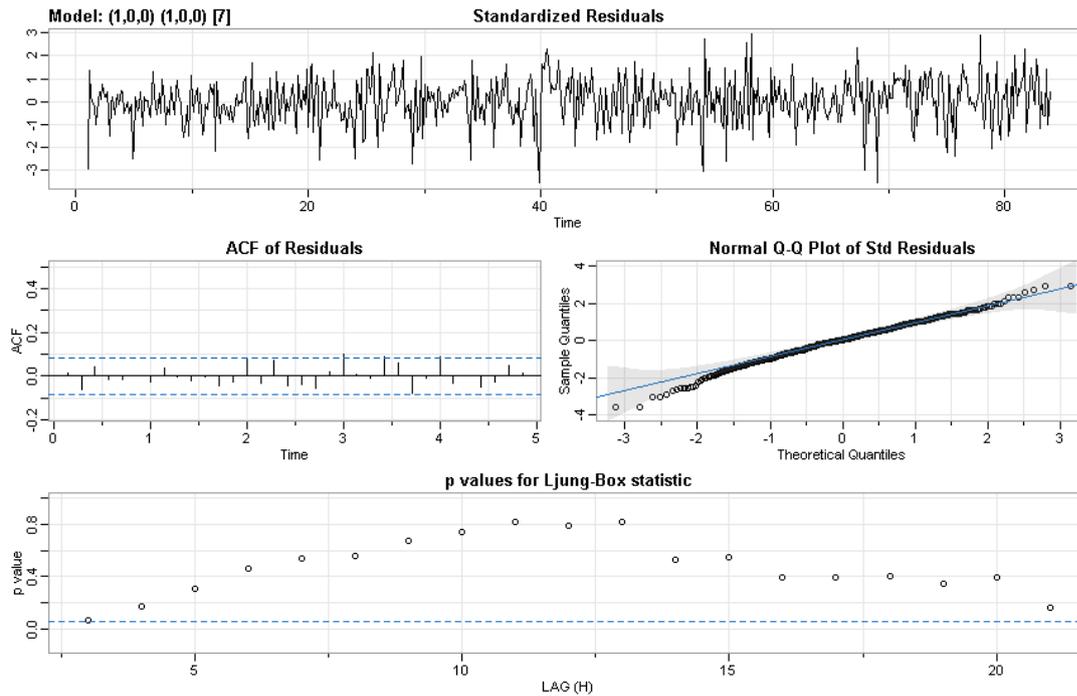


Figura 7.6: Gráfico do modelo sazonal das partículas inaláveis da Enseada de Suá (2014-2015)

Na Figura 7.6, encontra-se a análise de resíduos para este modelo. Nela, pode-se perceber, que os resíduos estão com um comportamento aceitável, os resíduos seguem uma distribuição Normal e os p-valores do teste de Lyung-Box indicam que os resíduos são independentes.

A estimação dos coeficientes é feita de forma similar aos modelos ARIMA, já descritos anteriormente. Os coeficientes do modelo sazonal obtido são,

Tabela 7.2: Coeficientes do modelo autorregressivo periódico diferenciado - PIAR

	s=1
ar_s	0,4809
sar_s	0,2058

É interessante comparar os modelos ajustados observando as medidas de AIC e BIC dos dois modelos obtidos.

Tabela 7.3: Tabela de comparação de AIC e BIC dos modelos.

Idade	Ano do Censo	
	AIC	BIC
Modelo PAR	7,369	103,31794
Modelo Sazonal	0,123	0,153

Por meio da Tabela 7.3, é possível perceber que o modelo sazonal recebeu menores AIC e BIC do que o modelo autorregressivo ajustado anteriormente. Isto indica que, para este caso, segundo estes critérios, um modelo sazonal seria melhor do que um modelo autorregressivo periódico. Cabe testar se para a maioria dos casos de séries sazonais e periódicas, tal correspondência continua sendo verdadeira.

Capítulo 8

Conclusões

Este trabalho apresentou o modelo autorregressivo periódico, que tem como característica principal a utilização em séries temporais que possuem uma estrutura de correlação entre observações vizinhas, que dependem apenas do tempo. Discutimos a identificação de series que se encaixam neste tipo de modelo, bem como as estimativas e fazer previsões. Além disso, um exemplo, utilizando a série do PIB da Alemanha entre os anos de 1960 a 1990, foi considerado, ilustrando o desenvolvimento do trabalho.

Por fim, uma aplicação com dados reais foi desenvolvida. Esta foi realizada utilizando dados de uma série temporal de poluição, também utilizados por Solci e Reisen (2017). Entretanto, a modelagem PAR não apresentou resultados melhores quando comparada ao modelo sazonal. Assim, como estudo futuro propomos que sejam analisadas quais outras características de séries temporais podem ser observadas para que se ajustem a este tipo de modelo.

Considero que para realização deste Trabalho de Graduação reestudei grande parte do conteúdo sobre série temporais, revisei conceitos de estimação de parâmetros, seleção de ordem e diagnóstico, e me deparei com alguns conceitos que não havia visto na graduação. Os objetivos deste trabalho foram atingidos, visto que o estudo do modelo autorregressivo periódico foi realizado.

Apêndice A

Códigos do R

```
#Carregando as bibliotecas necessárias
library(partsm)
library(ggplot2)
library(fpp)
library(forecast)
library(astsa)
library(tseries)

#Carregando os dados e aplicando o logaritmo
data("gergnp")
lgergnp = log(gergnp, base=exp(1))

#Gráficos utilizados
#Modelo sem diferenciação (com diferenciação é semelhante)

#Gráfico do logaritmo da série temporal
autoplot(lgergnp)+ theme_minimal()+
geom_line(colour = "blue")+ xlab("Tempo")

# Load RColorBrewer
library(RColorBrewer)

# Classic palette BuPu, with 4 colors
```

```

coul <- brewer.pal(4, "YlOrRd")

# Add more colors to this palette :
coul <- colorRampPalette(coul)(31)

# Plot it
pie(rep(1, length(coul)), col = coul , main="")

#Gráfico sazonal da série temporal
ggseasonplot(lgergnp, year.labels=TRUE, year.labels.left=TRUE, col = coul,
main = NULL, season.labels = c("Domingo","Segunda","Terça","Quarta","Quinta","Sexta",
"Sábado"))+theme_minimal()+ xlab("Ano")

#Gráfico sazonal polar
ggseasonplot(lgergnp, polar=TRUE, col = coul, main = NULL) + theme_minimal()+
xlab("Trimestres") +
labs(color = "Anos")

#Gráfico da subsérie sazonal
ggsubseriesplot(lgergnp) + theme_minimal()+
xlab("Trimestres")

#Gráfico de dispersão entre os trimestres
a = window(lgergnp, start = 1960)
gglagplot(a, seasonal = TRUE) + theme_bw() +
guides(color = guide_legend("Trimestres"))

#Decomposição multiplicativa da série temporal
lgergnp %>%
  decompose(type='multiplicative') -> pib_decomp

autoplot(pib_decomp, main = NULL) + theme_bw()+
xlab("Tempo")

```

```

#Função de autocorrelação e de autocorrelação parcial
acf2(lgergnp)

#Seleção da ordem do modelo autorregressivo periódico
detcomp = list(regular=c(0,0,0), seasonal=c(1,0), regvar=0)
aic = bic = Fnextp = Fpval = rep(NA, 4)
for(p in 1:4){
  lmpar = fit.ar.par(wts=lgergnp, detcomp=detcomp, type="PAR", p=p)
  aic[p] = AIC(lmpar@lm.par, k=2)
  bic[p] = AIC(lmpar@lm.par, k=log(length(residuals(lmpar@lm.par))))
  Fout = Fnextp.test(wts=lgergnp, detcomp=detcomp, p=p, type="PAR")
  Fnextp[p] = Fout@Fstat
  Fpval[p] = Fout@pval}

#Teste de periodicidade do modelo
dcsi = list(regular=c(0,0,0), seasonal=c(1,0), regvar=0)
out.Fparsi = Fpar.test(wts=lgergnp, detcomp=dcsi, p=2)
show(out.Fparsi)

dcsit = list(regular=c(0,0,0), seasonal=c(1,1), regvar=0)
out.Fparsit = Fpar.test(wts=lgergnp, detcomp=dcsit, p=2)
show(out.Fparsit)

#Representação matricial modelo PAR
out.par = fit.ar.par(wts=lgergnp, type="PAR", detcomp=detcomp, p=2)
out.MV = PAR.MVrepr(out.par)
out.MV

#Estimativas dos parâmetros modelo PAR
show(out.par)

#Teste para verificação de integração

```

```
out.LR = LRurpar.test(wts=lgergnp, detcomp=detcomp, p=2)
show(out.LR)

Fpari1.out = Fpari.piar.test(wts=lgergnp, detcomp=detcomp, p=2, type="PARI1")
show(Fpari1.out)

#Série diferenciada
dif =diff(lgergnp)

#Representação matricial modelo PIAR
out.piar = fit.piar(wts=lgergnp, detcomp=detcomp, p=2)
PAR.MVrepr(out.piar)

#Estimativas dos parâmetros modelo PIAR
show(out.piar)

#Teste para heterocedasticidade sazonal
par2 = fit.ar.par(wts=dif, type="PAR", p=2, detcomp=detcomp)
Fsh.out = Fsh.test(res=residuals(par2@lm.par), s=frequency(lgergnp))
show(Fsh.out)

#Teste de Ljung-Box
Box.test (residuals(par2@lm.par), lag = 1, type = "Ljung")

#Teste de Jarque-Bera
jarque.bera.test(residuals(par2@lm.par))

#Previsão do modelo
out.pred = predictpiar(wts=lgergnp, p=2, hpred=24)
show(out.pred)

out.pred@wts <- exp(1)^out.pred@wts
out.pred@fcst <- exp(1)^out.pred@fcst
```

```
out.pred@ucb <- exp(1)^out.pred@ucb
out.pred@lcb <- exp(1)^out.pred@lcb
plotpredpiar(out.pred)

#Boxplot
ts_plot_season <- function(x = x) {
  season <- cycle(x)
  season.factor <- factor(season)
  ggplot() +
    geom_boxplot(mapping = aes(x = season.factor,
                              y = x) ) +
    labs(x = "Periodo", y = "Serie")
}

ts_plot_season(lpoluição) + theme_bw()

#Modelo SARIMA
poluicao1 <- sarima(lpoluição,p = 1, d = 0, q = 0,
                  P = 1, D = 0, Q = 0,
                  S = 7)
```


Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.
- Akaike, H. (1979). A bayesian extension of the minimum aic procedure of autoregressive model fitting. *Biometrika*, **66**(2), 237–242.
- Akaike, H. *et al.* (1977). On entropy maximization principle. *Krishnaiah PR*.
- Cepel (2018). *GEVAZP – Modelo de Geração de Cenários de Energias e Vazões*. http://www.cepel.br/pt_br/produtos/gevazp-geracao-de-series-sinteticas-de-energias-e-vazoes-periodicas.htm
Acesso em 22 de maio de 2021.
- CETESB (2021). *Poluentes — Qualidade do Ar*. <https://cetesb.sp.gov.br/ar/poluentes/>. Acesso em 22 de maio de 2021.
- Franses, P. H. *et al.* (1996). Periodicity and stochastic trends in economic time series. *OUP Catalogue*.
- Gladyshev, E. (1963). Periodically and almost-periodically correlated random processes with a continuous time parameter. *Theory of Probability & Its Applications*, **8**(2), 173–177.
- Gujarati, D. N. e Porter, D. C. (2011). *Econometria Básica-5*. Amgh Editora.
- Hannan, E. J. (1980). The estimation of the order of an arma process. *The Annals of Statistics*, pages 1071–1081.
- Hipel, K. W. e McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier.

- Jarque, C. M. e Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, **6**(3), 255–259.
- Jones, R. H. e Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika*, **54**(3-4), 403–408.
- López-de Lacalle, J. (2005). Periodic autoregressive time series models in r: The partsm package. Technical report, BILCODEC 2005 working paper, Universidad del Pais Vasco UPV/EHU-Departamento.
- Maceira, M. (1989). Operação ótima de reservatórios com previsão de afluências. *Master's degree thesis, Federal University of Rio de Janeiro, Brazil*.
- Mendenhall, W. e Sincich, T. (2012). *A second course in statistics: regression analysis*, volume 7. Prentice Hall Upper Saddle River, NJ.
- Morettin, P. A. e Toloi, C. M. d. C. (2004). *Análise de séries temporais*. Edgard Blucher.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, **14**(5), 465–471.
- Sarnaglia, A. J. Q., Reisen, V. A. e Bondon, P. (2015). Periodic arma models: Application to particulate matter concentrations. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 2181–2185. IEEE.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.
- Solci, C. C. e Reisen, V. A. (2017). Estudo empírico de métodos de estimacao robusta de modelos par e aplicacao a dados de poluicao do ar.
- Souza, C. e Camargo, M. (2004). Análise e previsão de séries temporais: os modelos arima, s/e, 2ª edição. *Rio de Janeiro*.
- Stedinger, J. (2001). Report on the evaluation of cepel's par models. *Cornell University. School of Civil and Environmental Engineering. Ithaca, New York*.
- Tao, P.-C. e Delleur, J. W. (1976). Seasonal and nonseasonal arma models in hydrology. *Journal of the Hydraulics Division*, **102**(10), 1541–1559.
- Thomas, H. e Fiering, M. (1962). Design of water resources systems.

Tong, H. (1977). More on autoregressive model fitting with noisy data by akaike's information criterion (corresp.). *IEEE Transactions on Information Theory*, **23**(3), 409–410.