



**Universidade de Brasília
Departamento de Estatística**

**Análise Preditiva da Receita Bruta da Contribuição sobre Bens e Serviços
com Modelo Autorregressivo de Ordem Longa**

Daniel Miranda Rocha

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Daniel Miranda Rocha

**Análise Preditiva da Receita Bruta da Contribuição sobre Bens e Serviços
com Modelo Autorregressivo de Ordem Longa**

Orientador(a): Prof(a). Raul Yukihiro Matsushita

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Agradecimentos

Gostaria de iniciar agradecendo profundamente aos meus pais, à minha irmã e ao meu sobrinho Noah. Vocês sempre foram meu maior alicerce e inspiração, me apoiando incondicionalmente em todos os momentos da minha vida. Cada conquista que alcanço é resultado do amor, dedicação e suporte que recebi de vocês. Se hoje sou quem sou, devo isso inteiramente aos valores e ao carinho que sempre me ofereceram. Além disso, Noah, com apenas 2 anos, me proporcionou momentos de alegria e leveza em meio à pressão do TCC, me ajudando a relaxar e a encarar os desafios de forma mais tranquila. Esses momentos de fuga foram essenciais para que eu seguisse adiante com mais felicidade e menos estresse.

Aos meus amigos do curso, Stephany, Marcelo, Rubens, Irenaldo, Sabrina e Gabriela, que nos momentos mais difíceis, encontrei em vocês o incentivo e a parceria que tornaram essa caminhada mais leve. Ter a oportunidade de dividir essa experiência acadêmica com amigos tão especiais fez toda a diferença.

Um agradecimento especial à Julia Birbeire, que esteve comigo durante toda a trajetória do curso. Que foi mais do que uma amiga, foi um verdadeiro pilar de suporte, acompanhando de perto meu crescimento e sempre me incentivando a alcançar o melhor de mim. Sua presença foi essencial para que eu pudesse concluir esta etapa.

Agradeço também à UnB, que me proporcionou as oportunidades e o ambiente necessários para me desenvolver tanto como profissional quanto como pessoa.

À minha namorada, Fabiana, meu agradecimento especial. Esteve ao meu lado com uma paciência infinita e um apoio incansável, especialmente neste momento tão importante do TCC. Sua presença constante, seu suporte emocional e seu incentivo foram fundamentais para que eu pudesse concluir esta etapa com tranquilidade. Ter você ao meu lado durante este processo fez toda a diferença, e sou imensamente grato por tudo o que você representa na minha vida.

Não poderia deixar de lembrar do meu primo Kaio, que infelizmente não está mais aqui. Durante muitas madrugadas no ensino médio, sonhamos juntos em estar na UnB, imaginando todas as experiências que viveríamos. Embora ele não tenha tido a chance de viver isso, sinto que, de certa forma, vivi esse sonho por nós dois. Cada conquista, cada momento especial que tive na universidade, eu dedico a ele. Se cheguei até aqui, foi porque mantive vivo o nosso sonho de aproveitar ao máximo o que a UnB tinha a oferecer.

Por fim, deixo um agradecimento especial ao meu orientador, Prof. Raul Yukihiro Matsushita, por todo o apoio, confiança e orientação durante a elaboração deste trabalho. Sua sabedoria e dedicação foram fundamentais para a concretização deste projeto.

Resumo

Este estudo propõe a aplicação do Modelo Autorregressivo de Ordem Longa (AROL) para a projeção da receita bruta da Contribuição sobre Bens e Serviços (CBS). Utilizando dados tributários do Relatório do Tesouro Nacional (RTN) e técnicas de análise no software *R*, os dados foram transformados em uma função logarítmica para estabilizar a variabilidade e reduzir assimetria nos dados. Os valores também foram deflacionados com base no Índice Nacional de Preços ao Consumidor Amplo (IPCA) para torná-los comparáveis no tempo. Em contraste com os modelos de Box&Jenkins, que pertencem a uma classe particular de filtros lineares propostos na década de 1970 sob as limitações computacionais, os princípios clássicos da parcimônia e da redução de dimensionalidade, o modelo (AROL) é um caso geral do filtro linear que se aproveita da capacidade computacional moderna para proporcionar alto desempenho preditivo de maneira relativamente simples e direta. Desse modo, assim como performam os complexos modelos computacionalmente intensivos (que não são abordados neste trabalho), o modelo AROL não leva em consideração o princípio da parcimônia em sua formulação. Os resultados sugerem que o modelo AROL é aplicável diretamente sobre uma série temporal não estacionária e com *outliers*, tendo proporcionado erro médio percentual absoluto de 1,9%, o que é considerado excelente. O método também permite a obtenção de intervalos de previsão empíricos que permitem acomodar ocorrência de *outliers*.

Palavras-chaves: Modelo Autorregressivo de Ordem Longa, Previsão, Receita Bruta, CBS, Tesouro Nacional, Séries Temporais, MAPE

Lista de Tabelas

1	Algoritmo do Modelo Autorregressivo de ordem longa	15
2	Medidas resumo da Receita Bruta da Possível CBS	18
3	Métrica de desempenho do modelo	21

Lista de Figuras

1	Evolução Mensal da Receita Bruta Possível CBS	17
2	Boxplot da Receita Bruta da Possível CBS	18
3	Decomposição MSTL da Receita Bruta da Possível CBS	19
4	Gráfico da previsão um passo à frente	21
5	Gráfico dos erros padrão da previsão um passo à frente	22
6	Gráfico do MAPE acumulado da previsão um passo à frente	23
7	Previsão do Modelo Autorregressivo de Ordem Longa	24

Sumário

1 Introdução	8
2 Referencial Teórico	10
2.1 Componente das Séries Temporais	10
2.2 Transformação Logarítmica em Séries Temporais	10
2.3 Modelos Autorregressivos: AR(p)	10
2.3.1 Preditor Linear Empírico	11
2.4 Função de Autocorrelação (FAC)	12
2.5 Erro Percentual Absoluto Médio (MAPE)	12
3 Metodologia	13
3.1 Conjunto de dados	13
3.2 Técnicas Utilizadas	13
3.2.1 Algoritmo	14
4 Resultados	17
4.1 Análise Descritiva	17
4.2 Análise da Série	19
4.3 Definição do Modelo	20
4.4 Estudo do Desempenho Preditivo	23
5 Conclusão	25
Referências	26

1 Introdução

Ao longo dos anos, é notável o aumento da carga tributária no Brasil, o que tem gerado questionamentos acerca da adequação dessa carga ao perfil socioeconômico do país, bem como em relação à complexidade envolvida nesse assunto (RABELLO; OLIVEIRA, 2015).

Define-se como carga tributária, a soma da arrecadação tributária federal, estadual e municipal dividida pelo Produto Interno Bruto (PIB) (Receita Federal do Brasil, 2024), no qual é composta por impostos sobre bens e serviços, como: Imposto sobre Produtos Industrializados (IPI), Contribuição para Financiamento da Seguridade Social (Cofins) e Imposto sobre Circulação de Mercadorias e Serviços (ICMS); impostos sobre renda, lucros e ganho de capital, por exemplo: Imposto de Renda (IR); impostos sobre a propriedade, como: Imposto sobre Propriedade Territorial Urbana (IPTU); impostos sobre a folha de pagamento e mão de obra, como exemplo: Previdência Social (INSS) e Fundo de Garantia por Tempo de Serviço (FGTS); e impostos sobre comércio e transações internacionais, por exemplo: Imposto sobre Produtos Industrializados (IPI) que também contribui diretamente para o Pis/PASEP e Cofins (Presidência da República - Casa Civil, 2004).

Em 2023, o Brasil apresentou uma carga tributária bruta de 32,44% do PIB. Dentre as maiores arrecadações, destacam-se os Impostos sobre bens e serviços. De acordo com os dados, os impostos PIS, Cofins e IPI fazem parte das maiores participações desses impostos (Tesouro Nacional, 2023).

Dada a grande complexidade do sistema tributário do país, a reforma tributária tem sido um tema de grande relevância nos últimos anos. Em Julho de 2023, foi aprovado na Câmara dos Deputados o texto-base da PEC 45/2019, que visa simplificar e modernizar o sistema tributário no Brasil. Uma das propostas mais significativas em discussão, é a criação da CBS (Contribuição sobre Bens e Serviços), no qual serão substituídos os impostos: PIS, Cofins e IPI.

Os tributos que irão compor a CBS, possuem um papel fundamental na implementação de políticas públicas e sociais, visto que os valores arrecadados tem como objetivo financiar áreas essenciais como saúde, seguridade social e desenvolvimento econômico. Sendo assim, é importante entender o comportamento futuro da arrecadação para um melhor planejamento econômico do país e auxiliar na formulação de políticas públicas mais eficazes (GADELHA; ANDRADE; FIGUEIREDO, 2023).

Neste contexto, surgem desafios na aplicação de técnicas convencionais de análise de séries temporais por conta da presença de *outliers*, mudanças de regime e a não-normalidade dos dados.

Com o intuito de superar essas limitações, o estudo a seguir utiliza o modelo autorregressivo de ordem longa (AROL), uma abordagem robusta para previsão, que se mostrou eficaz em situações de variabilidade estrutural dos dados, conforme discutido por Sklarz et al. Diferente dos modelos tradicionais de Box&Jenkins, que foram criados com o intuito de facilitar o processamento de máquinas e que tem como princípios a parcimônia e a redução de dimensionalidade, o modelo AROL trata-se de um caso geral que aproveita da capacidade computacional atual para proporcionar alto desempenho preditivo de uma maneira relativamente simples e direta.

Portanto, a questão que este estudo busca responder é: "Como se comportará a arrecadação tributária unificada pela CBS no futuro, considerando um horizonte de 12 meses?". O objetivo geral deste trabalho é utilizar o Modelo Autorregressivo de Ordem Longa para gerar previsões da Receita Bruta da CBS, com foco exclusivamente preditivo. Para isso, foram seguidos alguns objetivos específicos: (i) deflação e transformação logarítmica da série de dados, (ii) ajuste do modelo AROL aos dados históricos, (iii) avaliação das previsões geradas pelo modelo em termos de erro e acurácia, e (iv) análise dos resultados em termos de seu uso para prever a arrecadação futura.

2 Referencial Teórico

Nesta seção, serão abordados os principais conceitos e técnicas necessárias para uma melhor compreensão do estudo realizado.

2.1 Componente das Séries Temporais

Uma série temporal é formada por quatro elementos principais: tendência, variações sazonais, variações cíclicas e variações aleatórias também conhecidas como resíduos. A tendência é definida como um movimento contínuo regular de longo prazo, que pode apresentar um movimento crescente ou decrescente (FISCHER, 1982).

A sazonalidade é descrita como eventos que apresentam características distintas em diferentes momentos do tempo (FRANSES, 1998). As variações cíclicas correspondem a oscilações de longo prazo que geralmente caracterizam os ciclos econômicos, ou seja, representam as flutuações em torno da curva de tendência ao longo do tempo. E por fim, os resíduos que se referem a eventos aleatórios imprevisíveis, como: fenômenos da natureza, pandemias, etc (FISCHER, 1982).

2.2 Transformação Logarítmica em Séries Temporais

A transformação logarítmica em séries temporais tem como objetivo estabilizar a variância e adequar os dados para modelagem e previsão. Consiste em aplicar o logaritmo natural aos valores reais da série.

2.3 Modelos Autorregressivos: AR(p)

Um modelo autorregressivo (AR) é um modelo estatístico usado para prever valores futuros de uma série temporal com base em valores passados. Eles são amplamente utilizados na análise técnica para prever preços futuros de títulos (BOX et al., 2015).

A ideia por trás dos modelos autorregressivos é que o valor atual de uma série pode ser interpretado como uma combinação linear de seus valores passados. Logo, um modelo autorregressivo de primeira ordem (AR(1)) pode ser escrito como:

$$X_t = c + \phi X_{t-1} + \varepsilon_t,$$

no qual X_t é o valor atual da série, c é uma constante, ϕ é o coeficiente autorregressivo que mede a relação entre X_t e X_{t-1} , e ε_t é um termo de erro aleatório.

De forma geral, os Modelos Autorregressivos seguem a seguinte fórmula:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

Para determinar o modelo AR(p), é necessário determinar a ordem do modelo (p), ou seja, o número de valores passados que serão incluídos na previsão. Isso pode ser feito observando a função de autocorrelação parcial (PACF) da série temporal. O PACF geralmente exibe um limite acentuado após o p-ésimo lag, sugerindo que o modelo AR(p) pode ajustar os dados.

Os modelos autorregressivos de longa ordem são uma extensão dos modelos AR, onde se considera um número maior de valores passados para prever os valores futuros da série temporal (SKLARZ; MILLER; GERSCH, 1987).

2.3.1 Preditor Linear Empírico

O preditor linear empírico é uma técnica utilizada para prever valores futuros de uma série temporal com base em suas autocorrelações. É calculado por meio das seguintes matrizes, que representam as relações entre as variáveis ao longo do tempo:

$$\begin{pmatrix} \rho(h) \\ \rho(h+1) \\ \vdots \\ \rho(h+n-1) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \cdots & \rho(n-1) \\ \rho(1) & 1 & \cdots & \rho(n-2) \\ \vdots & \rho(1) & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \cdots & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix},$$

na qual, a autocorrelação da série temporal no lag h , representada por $\rho(h)$, e os coeficientes β_n são utilizados para obter o preditor. Os valores matriciais observados da variável ρ resultam de uma combinação linear entre a matriz dos coeficientes derivados de ρ e a matriz de coeficientes β , estabelecendo uma relação direta entre a autocorrelação e os coeficientes preditivos.

A matriz dos coeficientes derivados de ρ é uma matriz de Toeplitz, pois a autocorrelação $\rho(h)$ possui uma estrutura onde os valores ao longo de cada diagonal dependem

apenas da diferença entre os índices, ou seja, $\rho(i, j) = \rho(|i - j|)$. Essa característica resulta em uma matriz de coeficientes de autocorrelação com elementos constantes em suas diagonais descendentes. A matriz de Toeplitz é obtida a partir da Função de Autocorrelação (FAC) e a partir da combinação linear entre essa matriz e a matriz de coeficientes β , é possível capturar de forma eficaz a dinâmica da série temporal e o comportamento preditivo da variável observada (BROCKWELL; DAVIS, 2002).

Conforme o lag h aumenta, pode ocorrer alguns problemas com o preditor empírico, como por exemplo: uma menor precisão nas estimativas de autocorrelação devido à escassez de dados; maior incerteza e erro nas previsões; amplificação do ruído e variações aleatórias; decaimento do poder preditivo das observações anteriores; resultando em previsões menos informativas (SURAKHI et al., 2021).

2.4 Função de Autocorrelação (FAC)

A Função de Autocorrelação (FAC) é expressa pela fórmula:

$$FAC = \frac{COV(X_t, X_{t-k})}{Var(X_t)}$$

Essa função quantifica a relação linear entre dois pontos em uma série temporal, com valores que variam entre -1 e +1. Caso o coeficiente de autocorrelação não diminua rapidamente à medida que a defasagem aumenta, isso sugere que a série não é estacionária (JOSE, 2022).

2.5 Erro Percentual Absoluto Médio (MAPE)

O MAPE é uma medida de precisão de previsão em um modelo. Ele calcula a média dos erros percentuais absolutos ao longo de todas as observações (SKLARZ; MILLER; GERSCH, 1987). Pode ser escrito como:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100$$

Por ser expresso em formato de percentual, a métrica facilita o entendimento e pode ser utilizada para comparar previsões em diferentes escalas e unidades.

3 Metodologia

Uma vez que esse trabalho foca na avaliação do Modelo Autorregressivo de Ordem Longa para prever a receita bruta da possível CBS, essa seção diz respeito ao banco e as técnicas utilizadas durante o estudo.

3.1 Conjunto de dados

Os dados utilizados neste projeto foram obtidos no Relatório do Tesouro Nacional (RTN), que disponibiliza informações sobre a tributação brasileira desde agosto de 1995. Para este estudo, foram selecionados os dados mensais de janeiro de 1997 até o presente, referentes à junção dos impostos PIS, Cofins e IPI. A consulta à base de dados foi realizada em 2 de setembro de 2024, disponível em: <https://www.gov.br/tesouronacional/pt-br/estatisticas-fiscais-e-planejamento/resultado-do-tesouro-nacional-rtn>.

3.2 Técnicas Utilizadas

Em relação as técnicas utilizadas, vale ressaltar que, assim como as manipulações realizadas no banco de dados, todas as análises foram realizadas utilizando o software *R*.

A partir do banco definido na Seção 3.1, após a análise exploratória, os dados foram deflacionados e transformados para função logarítmica com o objetivo de reduzir a variabilidade dos dados, linearizar as relações entre variáveis e tornar proporcional os dados que apresentam grandes taxas de variação.

A deflação ajusta os valores monetários de uma série temporal para um período de referência, removendo os efeitos da inflação. Isso permite comparar valores ao longo do tempo com um poder de compra constante. Para realizar a deflação, utiliza-se um índice de preços que reflete as variações no nível de preços de bens e serviços ao longo do tempo. Neste caso, foi utilizado o Índice de Preços ao Consumidor Amplo (IPCA) disponibilizados pelo pacote *sidrar*. A fórmula para deflacionar um valor monetário é dada por:

$$X_{t,b} = \frac{I_t}{I_b} \times Y_t,$$

na qual

- $X_{t,b}$ representa o valor deflacionado do mês t referente à data-base b ;
- Y_t denota o valor original da série no instante t ;
- I_t é o índice de preços do IPCA no instante t ;
- I_b é o índice de preços do IPCA na data-base b .

Com a série deflacionada, é obtido os valores ajustados de acordo com o preço constante, o que permite uma análise mais precisa ao longo do tempo, sem o efeito distorcido da inflação.

O Modelo Autorregressivo de Ordem Longa que foi aplicado aos dados, tem como objetivo considerar como secundário o conceito estatístico de redução de dimensionalidade e suficiência utilizados nos modelos conhecidos de Box&Jenkins, quando o objetivo da modelagem é a obtenção de previsões. Tendo em vista que com o avanço tecnológico e as ferramentas disponíveis nos dias atuais, é possível a execução de modelos mais robustos com uma maior flexibilidade (SKLARZ; MILLER; GERSCH, 1987).

3.2.1 Algoritmo

A tabela 1 apresenta o algoritmo do modelo em formato de código, explicando o processo de previsão com base em autocorrelações e erros históricos.

Tabela 1: Algoritmo do Modelo Autorregressivo de ordem longa

Código do Modelo AR de ordem longa

1. Inicialize variáveis:
 $x_hat = NULL$, $x_obs = NULL$, $error_h = NULL$, $n_size =$ comprimento de $x.t$
 $delta_h = 1$

2. Para t de $n_training$ até n_size faça:
 - 2.1. Defina x_fit como os valores de $x.t$ até t
 - 2.2. Calcule rho (autocorrelação) de x_fit com lag $(t - 1)$
 - 2.3. Construa a matriz $Omega$ como uma matriz Toeplitz de rho
 - 2.4. Calcule o vetor β como $\beta = inv(Omega) \times rho[(1 + f.h) : (t - delta.h)]$
 - 2.5. Calcule $\beta.0$ como a média de x_fit multiplicado por $(1 - sum(\beta))$
 - 2.6. Estime $x_hat[t + f.h] = \beta.0 + sum(rev(\beta) \times x_fit[(f.h + delta.h) : (t - delta.h)])$
 - 2.7. Armazene o valor observado $x_obs[t + f.h] = x.t[t + f.h]$
 - 2.8. Calcule o erro $error_h[t + f.h] = x_obs[t + f.h] - x_hat[t + f.h]$

3. Calcule os erros de predição:
 $MAPE = 100 \times$ média do valor absoluto dos erros divididos pelos valores observados

4. Calcule os intervalos de previsão usando os quantis do erro

5. Retorne os valores:
 Horizonte de previsão ($f.h$), MAPE, dataset, previsão (x_hat), intervalos

O algoritmo pode ser descrito da seguinte forma:

1. **Inicialização das Variáveis:** Inicialmente, são definidas as variáveis-chave do modelo: `x_hat` para armazenar as previsões, `x_obs` para registrar os valores observados, e `error_h` para armazenar os erros de previsão. O comprimento da série temporal é armazenado em `n_size`, e a variável `delta_h` é definida como 1, que indica o intervalo de atraso (lag).
2. **Loop de Previsão:** O algoritmo realiza iterações começando no ponto `n_training` até o fim da série (`n_size`). Para cada ponto no tempo, as seguintes operações são realizadas:
 - (a) A série temporal até o ponto `t` é armazenada em `x_fit`.
 - (b) A função de autocorrelação (`acf`) é calculada para a série `x_fit`, gerando o vetor `rho`, que contém as autocorrelações para cada lag.
 - (c) A matriz de covariância (`Omega`) é construída usando o vetor `rho` em uma forma de matriz de Toeplitz.

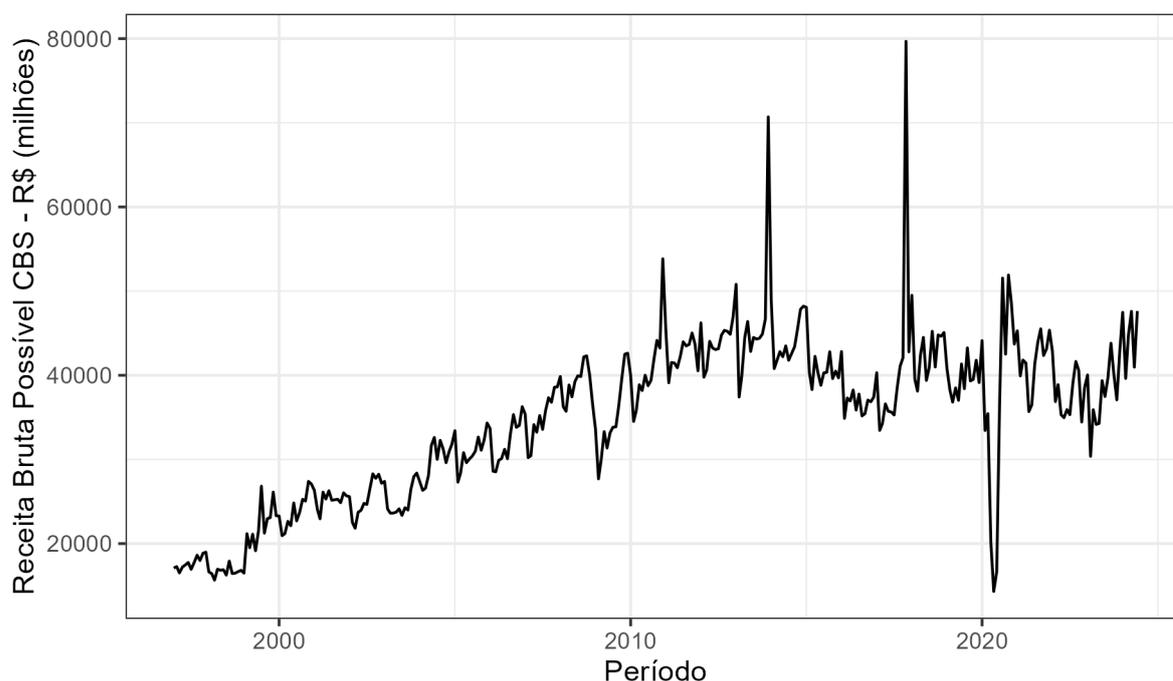
- (d) O vetor de coeficientes de predição **beta** é obtido resolvendo a equação linear, onde se utiliza a matriz inversa de **Omega** multiplicada por uma parte dos valores de autocorrelação. Esse subconjunto corresponde aos valores de ρ para os lags entre $1 + f.h$ e $t - \delta.h$, garantindo que somente as autocorrelações dentro desse intervalo temporal sejam utilizadas no cálculo dos coeficientes.
 - (e) O intercepto (**beta.0**) é calculado como a média da série multiplicada pela soma ponderada dos coeficientes de predição.
 - (f) A previsão para o próximo ponto no tempo, **x_hat[t+f.h]**, é então calculada utilizando os coeficientes **beta** e os valores passados da série temporal.
 - (g) O valor real observado **x_obs[t+f.h]** é registrado e o erro de predição **error_h[t+f.h]** é computado como a diferença entre o valor observado e o valor previsto.
3. **Cálculo dos Erros de Predição:** Após as iterações, o modelo calcula o Erro Percentual Absoluto Médio, que expressa o erro médio em termos percentuais em relação aos valores observados.
 4. **Intervalo de Previsão:** Utilizando os quantis do vetor de erros, o modelo gera intervalos de confiança para as previsões futuras, oferecendo uma margem de incerteza.
 5. **Saída:** O modelo retorna o horizonte de previsão (**f.h**), o valor MAPE, o conjunto de dados contendo os valores observados, as previsões e os erros, a previsão final, e os intervalos de previsão calculados.

4 Resultados

4.1 Análise Descritiva

Na Figura 1 é apresentada a evolução da arrecadação da receita dos impostos que compõem a Possível CBS de Janeiro de 1997 a Junho de 2024.

Figura 1: Evolução Mensal da Receita Bruta Possível CBS



Fonte: Relatório do Tesouro Nacional, 2024

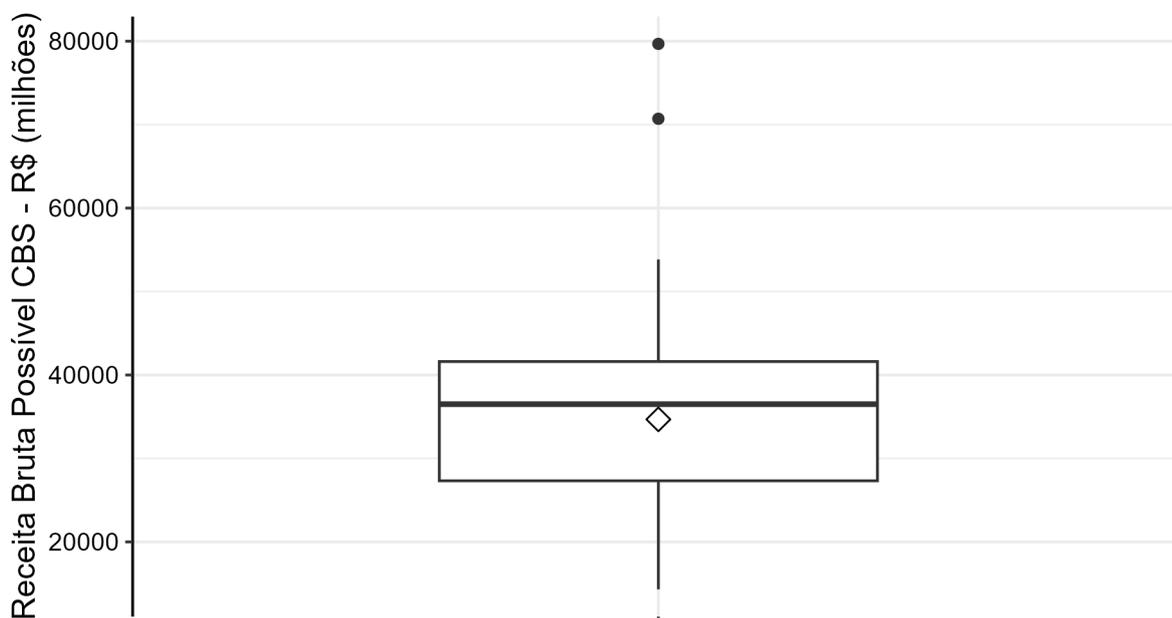
Observa-se que a série mostra o crescimento da arrecadação dos tributos com o passar dos meses até o final de 2013. A partir de 2013, é perceptível a estabilização da arrecadação que pode ser explicada por uma combinação de fatores, como a desaceleração econômica, que reduziu a base de arrecadação devido ao menor faturamento das empresas; a diminuição do consumo das famílias causada pelo aumento do desemprego e perda de renda; políticas fiscais de desoneração que reduziram alíquotas e impactaram diretamente a arrecadação; além da inflação elevada e aumento das taxas de juros, que limitaram o crédito e o consumo. Mesmo com uma recuperação lenta da economia após 2016, a base de arrecadação não retomou o ritmo de crescimento anterior, resultando em uma estabilização, apesar dos picos observados em Dez/13, Nov/2017 e Mai-Jun/20.

O aumento que ocorreu em Dezembro de 2013, está relacionado ao lançamento

de uma Multa de Ofício por Falta de Pagamento dos Tributos. O pico que ocorreu em Novembro de 2017, deve-se ao fato da elevação das alíquotas do PIS e Cofins sobre os combustíveis. Já a queda na arrecadação em Maio e Junho de 2020 ocorreu por conta da prorrogação do prazo para o recolhimento dessas contribuições que venceram em maio e junho de 2020, para outubro e novembro do mesmo ano (Portaria nº 139, de 3 de abril de 2020), em razão da pandemia relacionada ao coronavírus.

A Figura 2 e a Tabela 2, apresenta a distribuição dos dados da série por meio do boxplot e suas respectivas medidas resumo.

Figura 2: Boxplot da Receita Bruta da Possível CBS



Fonte: Relatório do Tesouro Nacional, 2024

Tabela 2: Medidas resumo da Receita Bruta da Possível CBS

Estatística	Valor
Média	34.681,66
Desvio Padrão	9.493,54
Mínimo	14.314,59
1º Quartil	27.323,71
Mediana	36.505,80
3º Quartil	41.612,68
Máximo	79.680,89

Por meio das medidas resumo (Tabela 2), observa-se que a média das arrecadações da Possível CBS é de aproximadamente 34,6 bilhões e possui um desvio padrão de aproximadamente 9,4 bilhões. Logo, pode-se considerar que ocorre uma maior dispersão dos valores em relação à média. Ou seja, há uma grande amplitude nos valores arrecadados.

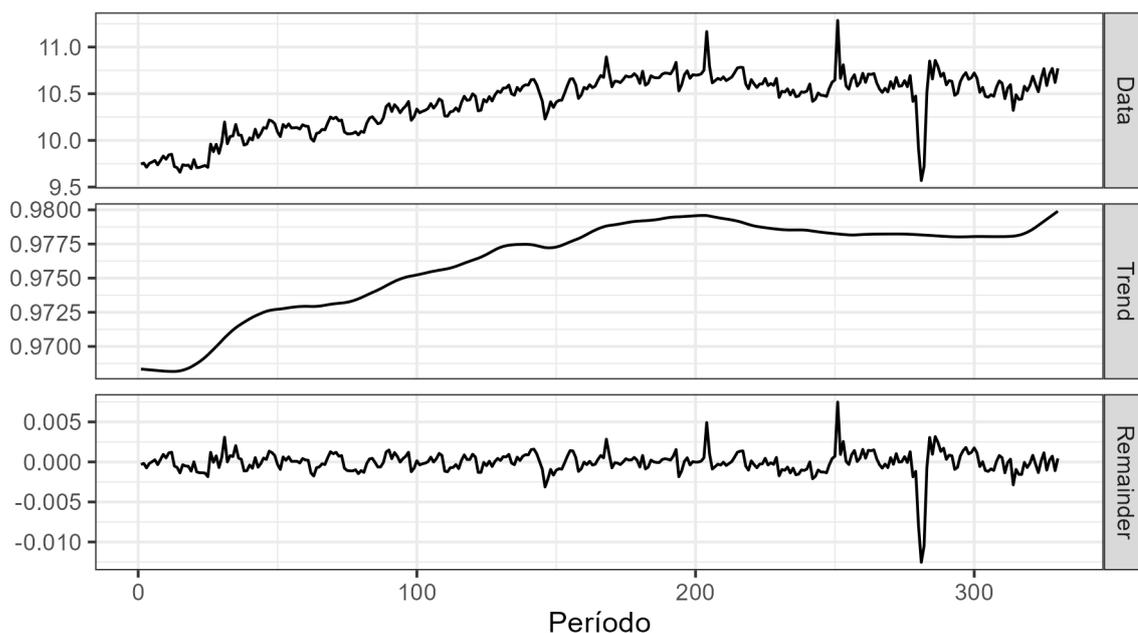
Além disso, é notável que os valores observados indicam que 25% das receitas brutas estão abaixo de aproximadamente 27,3 bilhões, conforme indicado pelo 1º quartil. Já o valor de 36,5 bilhões corresponde ao 2º quartil (ou mediana), o que significa que metade dos valores da receita bruta está abaixo desse valor, enquanto a outra metade está acima. Por fim, o 3º quartil revela que 75% dos valores estão abaixo de aproximadamente 41,6 bilhões, com os 25% restantes situando-se acima desse valor.

O valor mínimo da série, de aproximadamente 14,3 bilhões, está associado ao período da pandemia de COVID-19. Já o valor máximo, em torno de 79,6 bilhões, foi registrado devido às mudanças nas alíquotas do PIS e Cofins.

4.2 Análise da Série

Na Figura 3, é apresentada a decomposição MSTL da série da receita bruta da Possível CBS.

Figura 3: Decomposição MSTL da Receita Bruta da Possível CBS



Nota-se que, a análise da decomposição revela que não foi identificado um padrão

sazonal significativo na série temporal. Percebe-se que a série apresenta um padrão de crescimento moderado no início com pequenas flutuações no início seguido por períodos de estabilidade com algumas oscilações e quebras bruscas, as quais foram identificadas anteriormente na Seção 4.1. Essas quebras podem ser atribuídas a mudanças legislativas, alterações nas alíquotas e ao impacto da Pandemia do Covid-19.

Os resíduos da série variam em torno de uma média de zero, porém apresenta uma variância constante ao longo do tempo, o que indicava que os resíduos da série não exibem o comportamento esperado de um ruído branco. Observa-se também que tanto a série sem tendência quanto os resíduos apresentam uma quebra notável, no qual ocorreu por conta da Pandemia do Covid-19.

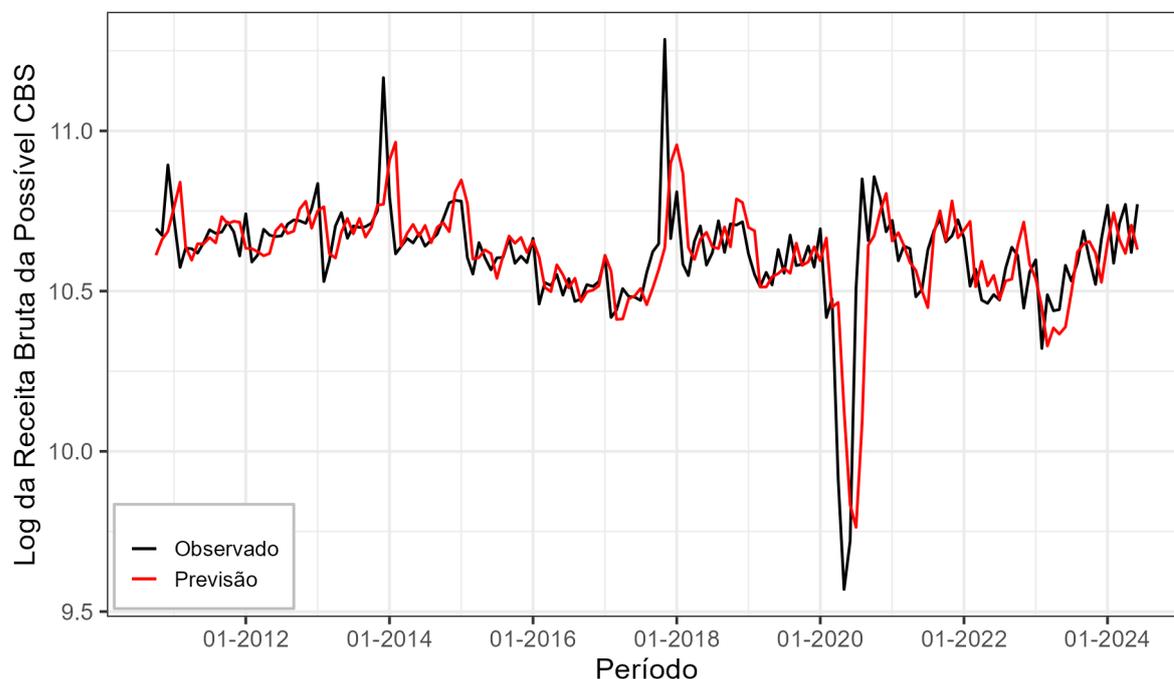
4.3 Definição do Modelo

Nesta seção, será abordado os resultados obtidos a partir da aplicação do modelo autorregressivo de ordem longa à receita bruta da possível CBS.

É importante destacar que os outliers presentes na série foram mantidos na análise, a fim de verificar a robustez do modelo. Além disso, foi aplicada a transformação logarítmica na receita bruta da possível CBS para melhorar o ajuste do modelo e lidar com a dispersão dos dados.

O gráfico abaixo, tem como objetivo apresentar a sua precisão preditiva e a capacidade de acompanhar a evolução da série original.

Figura 4: Gráfico da previsão um passo à frente



Observa-se na Figura 4 que o modelo conseguiu capturar bem as tendências e padrões da série. Em grande parte da série, a linha de previsão acompanha de perto os valores observados, indicando uma boa aderência do modelo aos dados.

Para confirmar a boa aderência do modelo realizado acima, foi calculado o MAPE (Erro médio Percentual Absoluto) com o objetivo de quantificar o desempenho do modelo:

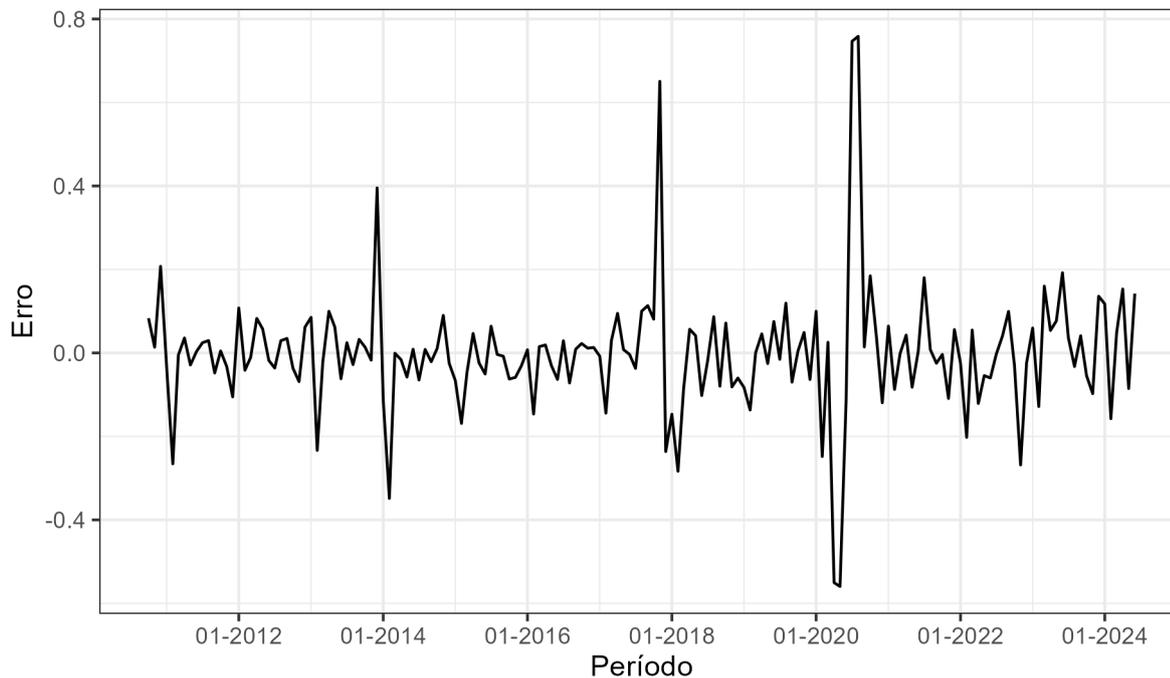
Tabela 3: Métrica de desempenho do modelo

Métrica	Resultado
MAPE	1,90%

O resultado apresentado na Tabela 3 indica que, em média, a previsão do modelo difere dos valores observados em apenas 1,90%. Isso sugere que o modelo possui uma alta acurácia, fornecendo previsões próximas dos valores reais.

A Figura 5 apresenta os erros de previsão, que é a diferença entre os valores observados e os valores previstos ao longo do tempo.

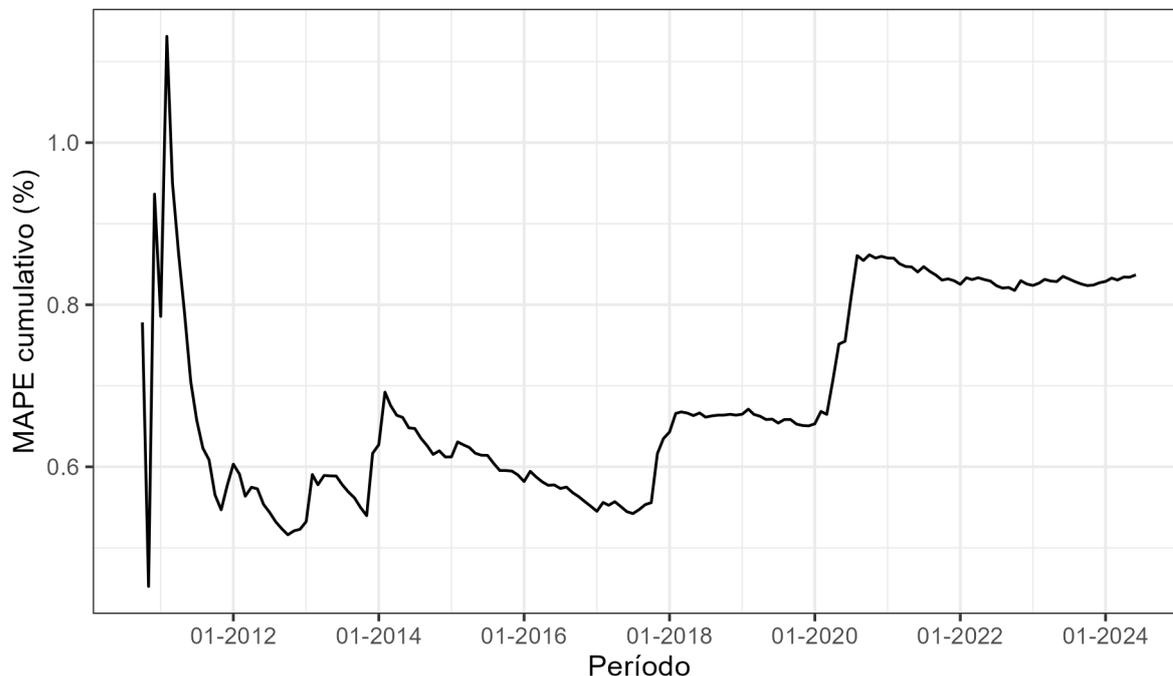
Figura 5: Gráfico dos erros padrão da previsão um passo à frente



Nota-se que, de maneira geral, os erros oscilam em torno de 0, tendo alguns picos em torno dos índices 200, 250 e 280. Isso mostra que o modelo teve dificuldades de prever as grandes oscilações existentes na série real observada pois apresentaram um maior erro, tanto positivos quanto negativos, em relação aos valores observados.

Com o intuito de verificar a evolução do erro percentual absoluto (MAPE) e a existência de padrões ou tendências no desempenho do modelo, foi elaborado o gráfico de MAPE Cumulativo.

Figura 6: Gráfico do MAPE acumulado da previsão um passo à frente



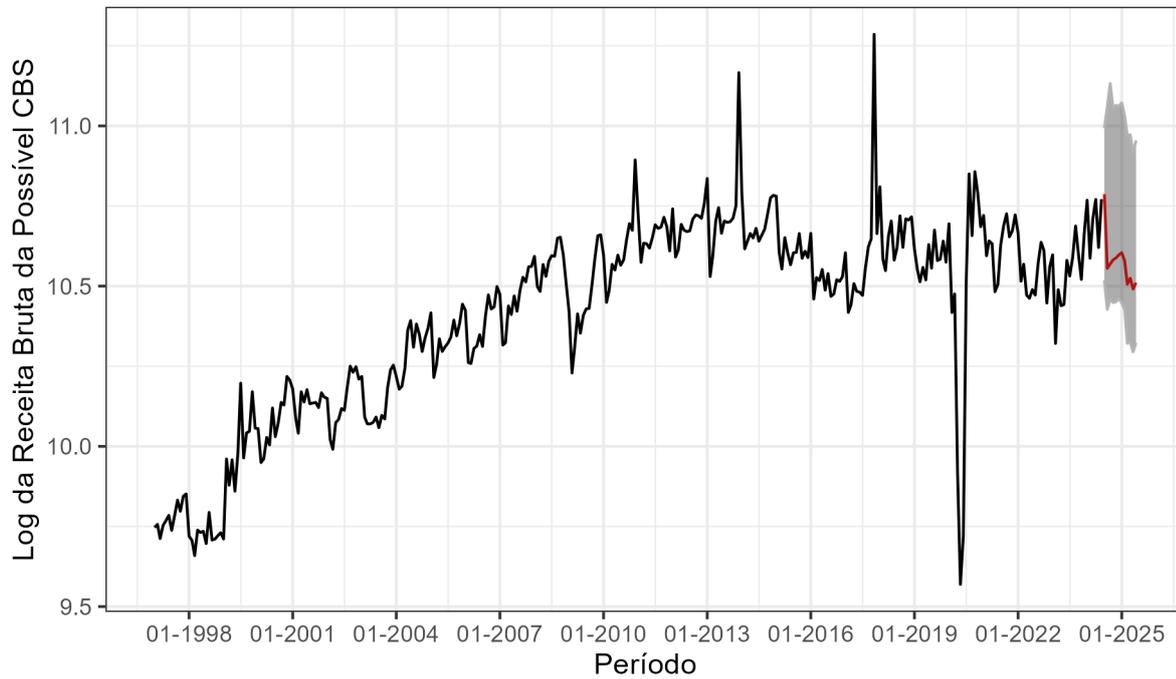
Ao analisar a Figura 6, é notável a presença de picos acentuados no início da série, indicando menor precisão nas primeiras previsões. Após o pico inicial, o gráfico apresenta uma queda, indicando uma melhora nas previsões e se mantém estável até por volta do índice 275, com pequenas flutuações que sugerem variações no desempenho da previsão. A partir do índice 280, há um aumento significativo para cerca de 0,8%, seguido por uma estabilização do valor permanecendo em torno desse percentual.

4.4 Estudo do Desempenho Preditivo

Levando em consideração a análise realizada na Seção acima (Seção 4.3), o modelo autorregressivo de ordem longa apresenta um excelente ajuste aos dados, portanto, o modelo é adequado para prever observações futuras da série.

A Figura 7 apresenta a série observada juntamente com a previsão para os próximos 12 meses, incluindo o intervalo de previsão, abrangendo até junho de 2025.

Figura 7: Previsão do Modelo Autorregressivo de Ordem Longa



Pode-se perceber que mesmo com a presença de outliers na série observada, a previsão segue a tendência, o que demonstra uma excelente adequação do modelo. Analisando o intervalo de previsão, nota-se uma faixa menor nos primeiros pontos. Conforme aumenta o número pontos previstos, o intervalo também aumenta, o que é esperado em previsões de longo prazo.

5 Conclusão

A aplicação do modelo autorregressivo de longa ordem à receita bruta da possível CBS demonstrou um ótimo resultado. Por meio da análise gráfica e da métrica de desempenho utilizada, pode-se concluir que o modelo possui uma alta acurácia, conseguindo capturar as tendências e padrões presentes na série temporal, mesmo com a presença de outliers.

Os resultados mostram que o modelo conseguiu acompanhar de perto os valores reais na maior parte da série, tendo um erro médio percentual absoluto de apenas 1,90%. Assim sugere que o modelo é robusto e adequado para previsões futuras, fornecendo estimativas confiáveis para a receita bruta da CBS.

Além disso, ao analisar os erros percebe-se que, embora o modelo tenha tido dificuldades de prever grandes oscilações, os erros se mantiveram oscilando em torno de zero grande parte do tempo. A evolução do MAPE cumulativo também indicou uma melhora nas previsões ao longo do tempo, com estabilização após os picos iniciais.

Por fim, a previsão para os próximos 12 meses, incluindo os intervalos de previsão, mostrou que o modelo continua a seguir a tendência da série observada, reforçando a adequação do modelo para previsões a longo prazo, embora seja esperado um aumento no intervalo de previsão conforme aumenta o horizonte temporal.

Em resumo, o modelo autorregressivo de longa ordem se mostrou eficiente e uma boa opção para a previsão da receita bruta da possível CBS, oferecendo uma ferramenta valiosa para análises e planejamentos futuros.

Referências

BOX, G. E. et al. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.

BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. 2nd. ed. [S.l.]: Springer, 2002.

FISCHER, S. *Séries univariantes de tempo-metodologia de Box & Jenkins*. [S.l.]: Fundacao de Economia e Estatistica, 1982.

FRANSES, P. H. *Time series models for business and economic forecasting*. [S.l.]: Cambridge university press, 1998.

GADELHA, S. R. d. B.; ANDRADE, B. B. de; FIGUEIREDO, W. W. L. *A Importância dos Marcos de Médio Prazo para o Planejamento Fiscal no Brasil: Instrumentos Orçamentários para um Novo Regime Fiscal*. [S.l.]: ENAP, 2023.

JOSE, J. Introduction to time series analysis and its applications. *Christ University, Bangalore*, 2022.

Presidência da República - Casa Civil. *Lei nº 10.865, de 30 de abril de 2004*. 2004. (https://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.865.htm).

RABELLO, G. G.; OLIVEIRA, J. M. d. Tributação sobre empresas no brasil: comparação internacional. Instituto de Pesquisa Econômica Aplicada (Ipea), 2015.

Receita Federal do Brasil. *Carga Tributária*. 2024. (https://www.gov.br/receitafederal/pt-br/assuntos/educacao-fiscal/educacao_fiscal/folhetos-orientativos/carga-tributaria-dig.pdf).

SKLARZ, M. A.; MILLER, N. G.; GERSCH, W. Forecasting using long-order autoregressive processes: An example using housing starts. *Real Estate Economics*, American Real Estate and Urban Economics Association, v. 15, December 1987.

SURAKHI, O. et al. Time-lag selection for time-series forecasting using neural network and heuristic algorithm. *Electronics*, MDPI, v. 10, 2021.

Tesouro Nacional. *Relatório Resumido da Execução Orçamentária - 6º Bimestre de 2023*. 2023. (https://sisweb.tesouro.gov.br/apex/f?p=2501:9:::9:P9_ID_PUBLICACAO:46589).