



DOCHMO ANALYTICS

MODELAGEM QUANTITATIVA

RELATÓRIO TÉCNICO DE ESTATÍSTICA APLICADA

MODELAGEM E PREVISÃO DA PRODUÇÃO DE PETRÓLEO NO BRASIL

UMA ABORDAGEM MULTIMETODOLÓGICA

DOUGLAS CHAVES MOURA

FORTALEZA - CE

JULHO DE 2026

VERSÃO 2.0

DOCHMO REPORT SERIES

Nº 001/2026

Douglas Chaves Moura

MODELAGEM E PREVISÃO DA PRODUÇÃO DE PETRÓLEO NO BRASIL

UMA ABORDAGEM MULTIMETODOLÓGICA

Relatório técnico especializado em modelagem estatística e análise preditiva de séries temporais. Desenvolvido como estudo aplicado de previsão e inferência no setor petrolífero brasileiro.

Fortaleza - Ceará

2026

RESUMO

Este trabalho analisa a série temporal da produção mensal de petróleo no Brasil no período de janeiro de 1997 a dezembro de 2025, totalizando 348 observações. O objetivo é compreender o processo gerador dos dados por meio da caracterização de sua tendência de longo prazo, de seus componentes cíclicos e sazonais, além de avaliar a estabilidade estrutural da série sob a influência de choques geopolíticos, macroeconômicos e sanitários. A modelagem estatística seguiu uma abordagem multimetodológica abrangendo a metodologia de Box-Jenkins para modelos SARIMA, análise espectral no domínio da frequência, testes de quebras estruturais (Bai-Perron) e modelos de intervenção (Box-Tiao). Adicionalmente, comparou-se a acurácia de previsões fora da amostra (janeiro de 2023 a dezembro de 2025) sob desenhos de validação cruzada do tipo holdout e rolling-origin, incluindo modelos lineares (SARIMA, ARIMAX, Holt-Winters) e não lineares (Prophet, TBATS, Redes Neurais Autorregressivas - NNAR). A análise espectral confirmou a forte presença de um ciclo sazonal anual, cuja energia espectral se distribui de forma significativa por seus harmônicos subanuais, refletindo a assimetria e a não-senoidalidade da dinâmica sazonal observada. A modelagem de intervenção identificou quebras de nível permanentes e transitórias significativas, com destaque para a crise do apagão de 2001 e a pandemia de COVID-19. Nos testes preditivos, o modelo Holt-Winters (exponencial no espaço log) apresentou a melhor capacidade de acompanhar a aceleração recente da produção nacional, seguido de perto pelo modelo auto.arima com deriva determinística. O trabalho conclui que a dinâmica de produção petrolífera brasileira é governada por uma tendência determinística robusta impulsionada pela maturação das reservas do pré-sal, cuja previsão exige a incorporação de modelagens de intervenção para acomodar a volatilidade de choques estruturais.

Palavras-chave: Séries Temporais. SARIMA. Análise de Intervenção. Análise Espectral. Produção de Petróleo.

ABSTRACT

This study analyzes the monthly petroleum production time series in Brazil from January 1997 to December 2025, comprising 348 observations. The objective is to understand the underlying data-generating process by characterizing its long-term trend and cyclical seasonal components, as well as evaluating structural stability under the influence of geopolitical, macroeconomic, and sanitary shocks. The statistical modeling followed a multi-methodological approach encompassing the Box-Jenkins methodology for SARIMA models, spectral analysis in the frequency domain, structural break tests (Bai-Perron), and intervention analysis (Box-Tiao). Additionally, we compared the accuracy of out-of-sample forecasts (January 2023 to December 2025) using holdout and rolling-origin validation designs across linear (SARIMA, ARIMAX, Holt-Winters) and non-linear models (Prophet, TBATS, Neural Network Autoregression - NNAR). Spectral analysis confirmed the strong presence of an annual seasonal cycle, whose spectral energy is significantly distributed across its sub-annual harmonics, reflecting the asymmetry and non-sinusoidal nature of the observed seasonal dynamics. The intervention modeling identified significant permanent level shifts and transient changes, highlighting the 2001 energy crisis and the COVID-19 pandemic. In predictive testing, the Holt-Winters model (exponential in log space) showed the best capability to track the recent acceleration of national production, closely followed by the auto.arima model with deterministic drift. The study concludes that the Brazilian oil production dynamics are governed by a robust deterministic trend driven by the maturation of pre-salt reserves, whose forecasting requires the incorporation of intervention modeling to accommodate the volatility of structural shocks.

Keywords: Time Series. SARIMA. Intervention Analysis. Spectral Analysis. Petroleum Production.

LISTA DE FIGURAS

3.1	Série temporal da produção mensal de petróleo no Brasil (jan/1997 – dez/2025).	11
3.2	Gráfico de subséries sazonais da produção de petróleo (jan/1997–dez/2025).	12
3.3	Distribuição mensal da produção de petróleo por mês do ano (1997–2025).	13
3.4	Variância da produção mensal por mês do ano (1997–2025).	13
3.5	Variação (%) média da produção de petróleo em relação ao mês anterior (1997–2025).	14
3.6	Volume Total Anual de Produção de Petróleo no Brasil (1997–2025).	16
3.7	Volatilidade Intra-Anual da Produção de Petróleo (1997–2025).	17
3.8	Variação Percentual Anual da Produção de Petróleo (YoY).	17
3.9	Gráfico Sazonal Polar — Ciclicidade Anual da Produção.	18
3.10	Gráfico sazonal da produção de petróleo (1997–2025).	19
3.11	Decomposição STL da série de produção de petróleo (1997–2025).	20
3.12	Decomposição clássica multiplicativa da série de produção de petróleo (1997–2025).	20
4.1	Série temporal após transformação logarítmica, $\ln(Y_t)$.	22
4.2	Série após primeira diferenciação.	24
4.3	Comparação das transformações para estacionariedade.	25
4.4	Série log-diferenciada $\Delta \ln(Y_t)$.	27
4.5	Partição da Série em Conjuntos de Treino e Teste.	28
5.1	FAC da série $\Delta \ln(Y_t)$ (treino, 1997–2022).	30
5.2	FACP da série $\Delta \ln(Y_t)$ (treino, 1997–2022).	30
5.3	FAC e FACP dos resíduos do modelo SARIMA(1,1,1)(1,0,1) ₁₂ (campeão por BIC).	32
5.4	FAC e FACP dos resíduos do modelo SARIMA(8,1,1)(1,0,1) ₁₂ .	34
5.5	Valores- p do teste de Ljung-Box dos resíduos — modelo SARIMA(8,1,1)(1,0,1) ₁₂ .	34
5.6	Resíduos do modelo campeão ao longo do tempo.	35
5.7	Gráfico Q-Q dos resíduos do modelo campeão.	36
5.8	Densidade empírica dos resíduos do modelo campeão.	36
6.1	Outliers identificados automaticamente pelo procedimento $t_{so}()$.	39
6.2	Efeito da remoção dos outliers detectados.	40
6.3	Pontos de intervenção manual sobre a série observada de produção de petróleo.	41

Lista de Figuras

6.4	Série $\ln(Y_t)$ com três quebras estruturais estimadas pelo método de Bai–Perron (série completa).	42
7.1	Periodograma bruto da série $\Delta \ln(Y_t)$, em função do período (meses).	48
7.2	Espectro suavizado (kernel de Daniell) versus periodograma bruto.	50
7.3	Ajuste da regressão harmônica de Fourier ($K = 6$) sobre a série completa. . . .	52
8.1	Comparação dos principais modelos no conjunto de teste (jan/2023–dez/2025). . .	55
8.2	Previsão da produção de petróleo no Brasil para 2026 — Holt-Winters.	60
8.3	Previsão da produção de petróleo no Brasil para 2026 — ARIMA + Fourier (parc.).	62

LISTA DE TABELAS

2.1	Taxonomia dos dados atípicos em séries temporais e suas equações de efeito. . .	6
3.1	Estatísticas descritivas da série mensal de produção de petróleo no Brasil, jan/1997–dez/2025 (em m ³).	10
3.2	Estatísticas descritivas mensais da produção de petróleo no Brasil (m ³), jan/1997 a dez/2025.	11
3.3	Desvio percentual de cada mês frente à média geral, antes e depois do ajuste por dias-no-mês.	15
4.1	Testes formais de raiz unitária e estacionariedade.	23
4.2	Testes formais para a ordem de diferenciação sazonal D , e sua adesão pelos candidatos SARIMA.	26
5.1	Comparação dos modelos SARIMA candidatos por AICc e BIC (conjunto de treino, 1997–2022).	31
5.2	Coeficientes estimados do modelo SARIMA(8,1,1)(1,0,1) ₁₂	33
5.3	Testes de diagnóstico dos resíduos do modelo SARIMA(8,1,1)(1,0,1) ₁₂	35
5.4	Teste de heterocedasticidade condicional (ARCH) nos resíduos do modelo SARIMA(8,1,1)(1,0,1) ₁₂	37
6.1	<i>Outliers</i> detectados na série de produção de petróleo (escala logarítmica, treino).	39
6.2	Datas de quebra estrutural identificadas pelo procedimento de Bai-Perron restrito ao conjunto de treino (jan/1997–dez/2022).	43
6.3	Fator de Inflação de Variância (VIF) dos regressores de intervenção.	44
6.4	Comparação por AICc: modelo sem intervenção vs. os três candidatos ARIMAX (treino).	45
6.5	Efeitos estimados — ARIMAX (5 manuais).	46
6.6	Efeitos estimados — ARIMAX (5 manuais + 3 quebras estruturais sem vazamento).	46
6.7	Efeitos estimados — ARIMAX parcimonioso (3 regressores).	46
7.1	Cinco maiores picos de densidade espectral do periodograma bruto.	49
7.2	AICc por número de harmônicos K — seleção restrita ao treino, busca exaustiva.	51
8.1	Acurácia preditiva no conjunto de teste (jan/2023–dez/2025), origem única, horizonte de 36 meses.	54

Lista de Tabelas

8.2	MAPE médio (%) obtido na validação rolling-origin, sete origens (dez/2022 a dez/2024, passo de 4 meses), horizonte de 12 meses.	56
8.3	Cobertura empírica dos intervalos de previsão nominais (holdout, 36 meses). . .	58
8.4	Cenários de previsão mensal para 2026 (m^3), por modelo.	60

SUMÁRIO

Resumo

Abstract

Lista de Figuras

Lista de Tabelas

1	Introdução	1
2	Metodologia	3
2.1	Pré-processamento e Engenharia de Dados	3
2.2	Metodologia de Box-Jenkins (SARIMA)	3
2.3	Análise de Dados Atípicos e Análise de Intervenção	6
2.4	Análise Espectral	7
2.5	Testes de Quebra Estrutural	8
2.6	Diagnóstico de Heterocedasticidade Condicional	8
2.7	Validação Preditiva e Métricas de Acurácia	9
3	Análise Exploratória	10
3.1	Descrição e Visualização dos Dados	10
3.2	Quantificação do Efeito de Calendário	15
3.3	Visão Anual	16
3.4	Decomposição da Série	19
4	Análise de Estacionariedade e Transformações	22
4.1	Transformação Logarítmica	22
4.2	Testes Formais de Raiz Unitária e de Tendência	23
4.3	Diferenciação Sazonal	25
4.4	Divisão Treino/Teste	27
5	Identificação e Estimação dos Modelos SARIMA	29
5.1	Identificação via FAC, FACP e EACF	29
5.2	Modelos Candidatos	31
5.3	Diagnóstico de Resíduos do Modelo Campeão	34

5.4	Heterocedasticidade Condicional dos Resíduos (ARCH)	37
6	Análise de Dados Atípicos, Intervenção e Quebra Estrutural	38
6.1	Detecção Automática de Outliers	38
6.2	Construção dos Regressores de Intervenção	40
6.3	Testes Formais de Quebra Estrutural	41
6.3.1	Caracterização Descritiva: Bai-Perron sobre o treino	42
6.4	Diagnóstico de Multicolinearidade	44
6.5	Os Três Modelos ARIMAX	45
7	Análise Espectral	48
7.1	Periodograma	48
7.2	Teste de Fisher para Significância Espectral	50
7.3	Regressão Harmônica de Fourier	50
8	Previsão e Avaliação de Modelos	53
8.1	Desenho de Avaliação	53
8.2	Resultado sob o Holdout Estático (36 meses)	54
8.3	Validação Rolling-Origin	56
8.4	Discussão: Por que os Desenhos Discordam, e o que Isso Significa	57
8.5	Cobertura Empírica dos Intervalos de Previsão	58
8.6	Diagnóstico de Sobreajuste do Modelo Campeão SARIMA	59
8.7	Previsão para 2026	60
9	Conclusão	63
9.1	Síntese dos Achados	63
9.2	Limitações do Estudo	64
9.3	Distinção Frente às Projeções Estruturais Oficiais	65
9.4	Recomendações e Trabalhos Futuros	65
	Referências Bibliográficas	65
A	Apêndice de Reprodutibilidade	68
A.1	Ambiente de Execução	68
A.2	Controle de Aleatoriedade	68
A.3	Sensibilidade Numérica do Procedimento $t_{SO}()$	68
A.4	Recomendação para Reexecução Futura	69

CAPÍTULO 1

INTRODUÇÃO

A análise de séries temporais constitui uma das fronteiras mais produtivas da estatística aplicada, situando-se na intersecção entre a teoria estocástica, a inferência estatística e a modelagem preditiva. Ao contrário de dados transversais, em que as observações são assumidas independentes entre si, as séries temporais caracterizam-se pela presença de dependência temporal: a covariância entre Y_t e Y_{t-k} é, em geral, não nula e decai de forma estruturada com a defasagem k . É essa estrutura de dependência que confere às séries temporais seu potencial analítico: em vez de obstruir a inferência, ela carrega informação sobre o processo gerador dos dados e, portanto, sobre o comportamento futuro da variável (BOX; JENKINS, 1970).

Do ponto de vista epistemológico, a análise de séries temporais resolve dois tipos de problemas fundamentais. O primeiro é descritivo: identificar, decompor e interpretar os componentes latentes da série — tendência de longo prazo, ciclos sazonais, flutuações irregulares e, quando presentes, descontinuidades estruturais provocadas por choques exógenos. O segundo é preditivo: construir modelos probabilísticos capazes de produzir previsões quantificadas, acompanhadas de intervalos de incerteza, sobre valores futuros ainda não observados (SHUMWAY; STOFFER, 2017).

O setor petrolífero brasileiro oferece um caso de estudo particularmente rico para a análise de séries temporais. O petróleo ocupa posição estratégica singular na economia nacional: responde por parcela significativa da balança comercial, impulsiona receitas fiscais via *royalties* e participações especiais, e constitui a principal fonte de energia primária do parque industrial brasileiro. A descoberta das reservas do pré-sal, anunciada em 2006 e progressivamente comprovada entre 2008 e 2010, representou um ponto de inflexão histórico: projeções iniciais da Agência Nacional do Petróleo (ANP) situavam o potencial acumulado em mais de 50 bilhões de barris, impulsionando a trajetória que posicionaria o Brasil, anos mais tarde, entre os dez maiores produtores mundiais de petróleo. A trajetória de produção nacional desde então passou a ser moldada, simultaneamente, pela curva de maturação tecnológica do pré-sal, pelas flutuações do preço internacional do Brent, pelas decisões estratégicas da Petrobras e pelas intervenções regulatórias do Estado.

A versão 2.0 deste trabalho — que representa uma atualização substancial em relação à edição anterior — nasce de três motivações complementares. Em primeiro lugar, a extensão do período de análise: a série agora cobre o intervalo de janeiro de 1997 a dezembro de 2025, incorporando 348 observações mensais e, de forma inédita, o ano de 2025 completo — o que permite avaliar se o novo patamar de produção atingido em 2023–2024 consolidou-se como

tendência ou representa um platô temporário. Em segundo lugar, a inclusão de duas metodologias inéditas em relação à versão anterior: a **análise de dados atípicos** e **análise de intervenção** (seguindo o arcabouço de Box e Tiao (1975)), que permite modelar formalmente os efeitos de eventos históricos sobre a série, e a **análise espectral**, que investiga a série no domínio da frequência e quantifica a contribuição de cada ciclo periódico para a variabilidade total da produção. Em terceiro lugar, refinamentos metodológicos na identificação dos modelos SARIMA, com a incorporação da Função de Autocorrelação Estendida (EACF) como ferramenta adicional de diagnóstico.

Cabe notar, desde já, que a previsão oficial da produção de petróleo brasileira — publicada anualmente pela Empresa de Pesquisa Energética no âmbito do Plano Decenal de Expansão de Energia — segue um paradigma estrutural e ascendente (*bottom-up*), fundamentado em cronogramas de projetos e engenharia de reservatório, não em modelagem estatística da série histórica agregada; é precisamente essa lacuna metodológica que a abordagem estatística aqui conduzida busca preencher.

A questão central que orienta este estudo pode ser enunciada da seguinte forma: *existe um ciclo anual — e possivelmente sub-anual — na produção mensal de petróleo no Brasil, e em que medida esse ciclo é modulado por choques estruturais de natureza geopolítica, regulatória e sanitária?* A resposta não é trivial: a presença de sazonalidade determinística coexiste com descontinuidades estruturais que alteram o nível da série de forma permanente (*level shifts*) ou transitória (*transient changes*), e a evidência reunida neste relatório indica que essa coexistência é mais complexa — e a instabilidade temporal mais pronunciada — do que uma leitura superficial da série sugeriria.

O relatório está organizado da seguinte forma. O Capítulo 2 apresenta o arcabouço teórico: pré-processamento de dados, metodologia de Box-Jenkins, análise de intervenção, análise espectral e testes de quebra estrutural. O Capítulo 3 realiza a caracterização estatística da série. O Capítulo 4 aplica testes formais de raiz unitária e transformações para estacionariedade. O Capítulo 5 conduz o ciclo de Box-Jenkins. O Capítulo 6 modela os choques históricos e apresenta a evidência formal de instabilidade estrutural. O Capítulo 7 investiga a série no domínio da frequência. O Capítulo 8 compara os modelos preditivos sob dois desenhos de validação distintos e projeta a produção para 2026. O Capítulo 9 sintetiza os achados, discute limitações e propõe trabalhos futuros.

CAPÍTULO 2

METODOLOGIA

2.1 Pré-processamento e Engenharia de Dados

Os dados brutos utilizados neste estudo provêm da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), disponibilizados através do Portal de Dados Abertos do Governo Federal. A base original apresenta granularidade mensal e estratificação por Unidade Federativa (UF) e por localização (terrestre ou marítima). O pré-processamento, implementado em R (versão 4.4.2, (R Core Team, 2024)), envolveu:

- **Agregação nacional:** para cada competência mensal, os volumes de produção foram somados ao longo de todas as UFs e de ambas as localizações (terra + mar), obtendo-se a série do volume total de petróleo produzido no Brasil em metros cúbicos (m³).
- **Recorte temporal:** a série foi filtrada para o intervalo de janeiro de 1997 a dezembro de 2025, resultando em $n = 348$ observações mensais contínuas.
- **Exportação em formato Parquet:** o dataset processado foi serializado em formato Apache Parquet com o pacote `nanoparquet`, que implementa a leitura e a escrita do formato de maneira autocontida, sem dependência da biblioteca nativa `libarrow`, favorecendo a portabilidade e a reprodutibilidade do fluxo de dados entre diferentes ambientes de execução.

2.2 Metodologia de Box-Jenkins (SARIMA)

A modelagem preditiva é fundamentada na metodologia de Box-Jenkins (BOX; JENKINS, 1970), consolidada como referência canônica na literatura de séries temporais econômicas (HAMILTON, 1994; MORETTIN; TOLOI, 2006; WEI, 2006). O método consiste em um procedimento iterativo de quatro estágios: identificação, estimação, verificação diagnóstica e previsão.

Um modelo SARIMA(p, d, q)(P, D, Q)_s para uma série $Z_t = \log(Y_t)$ é descrito pelo operador:

$$\phi_p(B) \Phi_P(B^s) (1 - B)^d (1 - B^s)^D Z_t = \theta_q(B) \Theta_Q(B^s) a_t, \quad a_t \sim \text{RB}(0, \sigma_a^2), \quad (2.1)$$

em que B é o operador de retardo (*backshift*), (p, d, q) e (P, D, Q) são as ordens não sazonais e sazonais, $s = 12$, e a_t é ruído branco.

A seleção entre modelos candidatos é guiada por critérios de informação:

$$\text{AIC} = -2\ell(\hat{\psi}) + 2k, \quad (2.2)$$

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}, \quad (2.3)$$

$$\text{BIC} = -2\ell(\hat{\psi}) + k\ln(n), \quad (2.4)$$

em que $\ell(\hat{\psi})$ é a log-verossimilhança maximizada, k o número de parâmetros e n o tamanho da amostra (HYNDMAN; ATHANASOPOULOS, 2018). O AICc é preferível ao AIC para n/k moderado.

Importante — escopo do que o AICc mede: os critérios de informação (2.2) – (2.4) avaliam exclusivamente o ajuste *dentro da amostra de treino*, penalizado pela complexidade. Eles não medem, e não devem ser interpretados como medindo, a capacidade de generalização do modelo para períodos fora da amostra. Como ficará evidente no Capítulo 8, o modelo SARIMA com menor AICc neste estudo não é o modelo com menor erro no conjunto de teste — um resultado perfeitamente possível e consistente com a teoria, mas que exige que conclusões sobre “o melhor modelo” sejam sempre qualificadas pelo critério (*in-sample* ou *out-of-sample*) utilizado.

Robustez da seleção ao critério de informação utilizado. O AIC e o AICc penalizam a complexidade do modelo de forma constante ($2k$), independentemente do tamanho da amostra; o BIC penaliza com peso $k\ln(n)$, crescente em n — para $n = 312$ (tamanho do treino utilizado neste estudo), $\ln(312) \approx 5,74$, mais que o dobro do peso 2 usado pelo AIC/AICc. Quando a margem de vitória do melhor modelo por AICc sobre o segundo colocado é pequena (inferior a 2 pontos, limiar frequentemente citado como “diferença substancial” em Burnham e Anderson (2002)), é metodologicamente relevante verificar se o BIC — que penaliza mais fortemente os parâmetros adicionais — aponta para o mesmo modelo. Esta verificação é incorporada na Seção 5.2 e revela, neste estudo, uma reversão: o BIC favorece um modelo mais parcimonioso do que o AICc.

Observação sobre a deriva em modelos com dupla diferenciação: Quando $d = 1$ e $D = 1$ são usados simultaneamente, um termo de deriva determinística (`include.drift = TRUE` em `forecast::Arima()`) torna-se, na prática, não identificável: uma tendência linear no tempo, diferenciada uma vez de forma regular e uma vez de forma sazonal, é levada a zero antes de entrar na equação do modelo. Por essa razão, os modelos SARIMA com diferenciação sazonal completa ($D = 1$) estimados neste trabalho foram ajustados com `include.drift = FALSE`. Já os modelos com componente autorregressiva sazonal e sem diferenciação sazonal ($D = 0$) permitem a inclusão identificável de deriva e foram, portanto, estimados com `include.drift = TRUE`. Consequentemente, apenas os modelos com $D = 1$ têm função

de previsão de longo prazo isenta de componente determinística de crescimento, dependendo inteiramente da dinâmica transitória dos termos AR/MA; os modelos com $D = 0$ e deriva incluída — entre eles o modelo com melhor AICc (Capítulo 5) — carregam uma tendência linear determinística em sua extrapolação. Essa distinção é importante para interpretar corretamente os resultados do Capítulo 8 e é resgatada explicitamente naquele capítulo.

A ordem de diferenciação regular d : é determinada por meio de três testes complementares de raiz unitária/estacionariedade, aplicados a cada representação candidata da série: o teste *Augmented Dickey-Fuller* (ADF) (DICKEY; FULLER, 1979), sob H_0 : a série possui raiz unitária; o teste *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) (KWIATKOWSKI et al., 1992), sob H_0 : a série é estacionária em torno de um nível determinístico — hipótese nula oposta à do ADF; e o teste *Phillips-Perron* (PP) (PHILLIPS; PERRON, 1988), sob a mesma hipótese nula do ADF, mas com correção não paramétrica para autocorrelação e heterocedasticidade nos resíduos, dispensando a especificação explícita do número de defasagens aumentadas exigida pelo ADF. A estacionariedade de uma representação candidata é declarada apenas quando os três testes concordam simultaneamente: ADF e PP rejeitam a hipótese de raiz unitária, e o KPSS não rejeita a hipótese de estacionariedade. Esse critério de concordância unânime — mais conservador do que uma regra de maioria simples — é adotado precisamente porque os três testes diferem em premissas (o KPSS, em particular, testa uma hipótese nula oposta à dos outros dois, e é conhecidamente mais sensível a quebras estruturais na série, Capítulo 6); a aplicação formal desse critério às quatro representações candidatas da série (nível, log, primeira diferença e log com primeira diferença) é conduzida na Seção 4.2.

A ordem D de diferenciação sazonal deve, ela própria, ser testada formalmente — não apenas inferida da presença de sazonalidade. A literatura disponibiliza testes específicos para essa decisão, distintos dos testes de presença de sazonalidade (Seção 4.3): o teste OCSB (OSBORN et al., 1988) e o teste de Canova-Hansen (CANOVA; HANSEN, 1995), ambos implementados em `forecast::nsdiffs()`. Os testes são aplicados formalmente (Capítulo 4) e discute-se as implicações de seu resultado para a escolha de especificação.

O parâmetro de transformação de Box-Cox λ também deve ser verificado quantitativamente, e não apenas assumido a partir de evidência visual de heterocedasticidade. Guerrero (1993) propõe um método que minimiza o coeficiente de variação das subséries sazonais — particularmente adequado para séries com sazonalidade, como a deste estudo — implementado em `forecast::BoxCox.lambda(method = "guerrero")`. Um segundo método, baseado na maximização da verossimilhança de uma regressão linear simples no tempo (`method = "loglik"`), é também reportado por completude, com a ressalva de que ele desconsidera a estrutura de autocorrelação da série e é, por isso, menos confiável para séries com dependência temporal forte como a presente.

2.3 Análise de Dados Atípicos e Análise de Intervenção

A taxonomia de Chen e Liu (1993) classifica os dados atípicos em quatro tipos: *Additive Outlier* (AO, pulso isolado), *Innovational Outlier* (IO, inovação propagada via estrutura ARMA), *Level Shift* (LS, mudança permanente de nível) e *Transient Change* (TC, mudança temporária com decaimento exponencial). A Tabela 2.1 resume as equações de efeito, em que ω denota a magnitude do efeito e, no caso específico do TC, δ (com $0 < \delta < 1$) denota a taxa de decaimento exponencial do choque — dois parâmetros distintos, estimados separadamente.

Tabela 2.1: Taxonomia dos dados atípicos em séries temporais e suas equações de efeito.

Tipo	Sigla	Efeito sobre a série
Additive Outlier	AO	$\omega \cdot I_t(T)$
Innovational Outlier	IO	$\omega \cdot \psi(B) \cdot I_t(T)$
Level Shift	LS	$\omega \cdot (1 - B)^{-1} \cdot I_t(T)$
Transient Change	TC	$\omega \cdot (1 - \delta B)^{-1} \cdot I_t(T)$

Fonte: Elaborado pelo Autor (2026).

Na Tabela 2.1, $I_t(T)$ denota a função indicadora do ponto T (igual a 1 se $t = T$ e 0 caso contrário), ω é a magnitude do efeito e δ é a taxa de decaimento exponencial específica do TC, e $\psi(B) = \theta_q(B) [\phi_p(B) (1 - B)^d]^{-1}$ é o polinômio de pesos de resposta ao impulso do modelo ARMA.

O modelo geral de intervenção de Box e Tiao (1975) representa a série observada como a soma de uma componente de sinal determinístico — função das variáveis de intervenção x_t — e uma componente estocástica modelada por um processo ARMA:

$$Y_t = \frac{\eta(B)}{\delta_r(B)} x_t + \frac{\theta_q(B)}{\phi_p(B) (1 - B)^d} a_t, \quad (2.5)$$

onde $\eta(B)$ e $\delta_r(B)$ são polinômios no operador de retardo (o subscrito r evita colisão notacional com o δ escalar do TC definido), e x_t pode ser um impulso unitário (modelando AO ou IO), uma função degrau $\xi_t^{(T)} = (1 - B)^{-1} I_t(T)$ (modelando LS) ou uma combinação das duas formas.

Diagnóstico de multicolinearidade entre regressores de intervenção. Quando múltiplas variáveis de intervenção do tipo degrau (LS) são incluídas simultaneamente em um mesmo modelo — como ocorre neste estudo, combinando eventos históricos conhecidos com datas de quebra estimadas estatisticamente (Capítulo 6) —, a colinearidade entre regressores pode inflacionar artificialmente os erros-padrão estimados e comprometer a interpretabilidade individual dos coeficientes. A matriz de correlação par a par é insuficiente para detectar colinearidade decorrente de combinações lineares de três ou mais variáveis; por isso, o fator de inflação de variância (*Variance Inflation Factor*, VIF) é utilizado nesse trabalho como diagnóstico complementar, calculado

para cada regressor j como $VIF_j = 1/(1 - R_j^2)$, em que R_j^2 é o coeficiente de determinação de uma regressão de x_j contra todos os demais regressores do modelo. Convencionalmente, $VIF > 5$ indica colinearidade moderada e $VIF > 10$, colinearidade severa.

2.4 Análise Espectral

A análise espectral oferece uma perspectiva complementar à abordagem no domínio do tempo: em vez de descrever a série em termos de suas autocovariâncias, ela decompõe a variância total da série em contribuições de diferentes frequências (ou equivalentemente, de diferentes períodos). A densidade espectral de potência de uma série estacionária de segunda ordem é a transformada de Fourier da função de autocovariância:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega}, \quad \omega \in [-\pi, \pi], \quad (2.6)$$

estimada pelo periodograma $I(\omega_j) = \frac{1}{2\pi n} |\sum_t Y_t e^{-it\omega_j}|^2$ nas frequências de Fourier $\omega_j = 2\pi j/n$ (BROCKWELL; DAVIS, 1991; SHUMWAY; STOFFER, 2017).

O periodograma é assintoticamente não viesado ($E[I(\omega_j)] \rightarrow f(\omega_j)$), porém inconsistente: sua variância não decresce com n (SHUMWAY; STOFFER, 2017). Para obter estimativas espectrais consistentes, aplica-se suavização espectral via kernels de Daniell ou de Parzen, que calculam médias ponderadas dos ordinais do periodograma em vizinhanças de frequência, com o compromisso clássico entre viés e variância.

Ressalva sobre o periodograma de séries diferenciadas. O filtro de primeira diferença $(1 - B)$ possui função de transferência de ganho $|1 - e^{-i\omega}|^2 = 2(1 - \cos \omega)$, estritamente crescente em $\omega \in [0, \pi]$: diferenciar uma série amplifica mecanicamente a contribuição aparente de frequências altas (períodos curtos) em relação a frequências baixas. Como o periodograma deste estudo (Capítulo 7) é calculado sobre $\Delta \ln(Y_t)$, parte da concentração de energia espectral em harmônicos de período curto pode refletir esse ganho do filtro, e não exclusivamente a assimetria genuína do ciclo sazonal.

A Regressão Harmônica de Fourier constitui uma abordagem alternativa ao SARIMA puro para modelagem de sazonalidade: em vez de diferenciar sazonalmente a série, incorporam-se como regressores exógenos os pares de harmônicos $\cos(2\pi kt/s)$ e $\sin(2\pi kt/s)$ para $k = 1, \dots, K$ no modelo ARIMA:

$$Z_t = \mu + \sum_{k=1}^K \left[\alpha_k \cos\left(\frac{2\pi kt}{s}\right) + \beta_k \sin\left(\frac{2\pi kt}{s}\right) \right] + \varepsilon_t, \quad \varepsilon_t \sim \text{ARIMA}(p, d, q), \quad (2.7)$$

em que $s = 12$ e o número ótimo de harmônicos K^* é selecionado por minimização do AICc (SHUMWAY; STOFFER, 2017; HYNDMAN; ATHANASOPOULOS, 2018). Esse modelo é particularmente atraente por sua parcimônia: para muitos padrões sazonais, $K^* \ll s/2$, produzindo

modelos com muito menos parâmetros do que um SARIMA com diferenciação sazonal completa. Note-se que $K = s/2$ (aqui, $K = 6$) é o número máximo de pares harmônicos admissível para uma frequência $s = 12$ — a frequência de Nyquist — correspondendo a uma representação saturada, com tantos parâmetros sazonais quanto indicadores mensais.

2.5 Testes de Quebra Estrutural

A metodologia de Box-Jenkins pressupõe implicitamente que a relação dinâmica capturada pelo modelo é estável ao longo de todo o período de estimação. Essa suposição pode ser testada formalmente. O teste CUSUM de Brown, Durbin e Evans (1975), na variante OLS-CUSUM, acumula os resíduos recursivos (ou os resíduos de mínimos quadrados ordinários, no caso de uma regressão auxiliar contra uma tendência determinística) e testa H_0 : parâmetros constantes ao longo do tempo, contra H_1 : existe pelo menos uma quebra. O procedimento de Bai e Perron (2003) generaliza essa ideia para a detecção de *múltiplas* quebras em datas desconhecidas, estimando-as por minimização da soma de quadrados dos resíduos particionada.

A datação de quebras estruturais usadas para construir regressores de modelos avaliados *out-of-sample* deve ser restrita ao conjunto de treino. Esse ponto, estruturalmente análogo ao que já se aplicava à seleção de K^* na regressão de Fourier, é particularmente relevante para quebras situadas próximas do fim da amostra: o próprio procedimento de Bai-Perron precisa de observações *posteriores* a uma quebra hipotética para estimar com confiança a inclinação do novo regime, de modo que datar uma quebra a poucos meses do fim do treino utilizando dados de teste tende a produzir uma data que parece “confirmada” precisamente porque foi estimada com informação que, na prática preditiva, não estaria disponível.

2.6 Diagnóstico de Heterocedasticidade Condicional

O diagnóstico de resíduos da metodologia de Box-Jenkins concentra-se, em sua forma tradicional, na estrutura de autocorrelação *em nível* (Ljung-Box) e na normalidade (Shapiro-Wilk, Jarque-Bera). Esses testes não detectam heterocedasticidade condicional: a possibilidade de que a variância dos resíduos, embora constante em média, varie sistematicamente ao longo do tempo, concentrando-se em *clusters* de alta e de baixa volatilidade. O teste ARCH-LM de Engle (1982) avalia formalmente essa hipótese; uma forma operacionalmente equivalente consiste em aplicar o teste de Ljung-Box à série de resíduos *ao quadrado* $\hat{a}_t^2$ — sob H_0 de ausência de heterocedasticidade condicional, $\hat{a}_t^2$ não deveria exibir autocorrelação significativa. Este diagnóstico é incorporado na Seção 5.3 e tem implicação direta sobre a confiabilidade dos intervalos de previsão (Seção 2.7 e Capítulo 8).

2.7 Validação Preditiva e Métricas de Acurácia

Além do *holdout* estático (treino e teste fixo de 36 meses), este estudo também avalia os modelos por meio de *validação rolling-origin* (HYNDMAN; ATHANASOPOULOS, 2018): a origem de previsão é deslocada ao longo do tempo, o modelo é reestimado em cada origem com os dados disponíveis até aquele ponto, e o erro de previsão para um horizonte fixo é calculado e depois agregado entre origens. Esse procedimento produz uma estimativa do erro de previsão com variância mais baixa e menos sensível à coincidência entre a fronteira do holdout e algum evento atípico do período de teste — risco discutido detalhadamente no Capítulo 8.

As métricas de acurácia utilizadas incluem o erro percentual absoluto médio (MAPE), a raiz do erro quadrático médio (RMSE) e o erro absoluto médio (MAE), calculados das formas usuais sobre o conjunto de teste. O erro absoluto escalonado médio (MASE) e a estatística U de Theil exigem maior cuidado em sua definição e merecem um destaque específico.

Definição canônica do MASE (HYNDMAN; KOEHLER, 2006):

$$\text{MASE} = \frac{\text{MAE}_{\text{modelo, teste}}}{\text{MAE}_{\text{naive sazonal, treino}}}, \quad \text{MAE}_{\text{naive sazonal, treino}} = \frac{1}{n_{\text{treino}} - s} \sum_{t=s+1}^{n_{\text{treino}}} |Y_t - Y_{t-s}|, \quad (2.8)$$

e, analogamente, a estatística U de Theil compara o RMSE do modelo no teste ao RMSE do mesmo predictor *naive* sazonal de um passo, igualmente calculado dentro do treino. O ponto essencial da definição (2.8) é que o termo de escala no denominador representa a dificuldade “natural” de prever a série, estimada com informação disponível no momento da previsão — e não a partir do próprio período sendo avaliado.

ANÁLISE EXPLORATÓRIA

3.1 Descrição e Visualização dos Dados

A série temporal em estudo consiste no volume mensal de petróleo produzido no Brasil, em metros cúbicos (m^3), entre janeiro de 1997 e dezembro de 2025: $n = 348$ observações mensais, frequência $s = 12$.

Tabela 3.1: Estatísticas descritivas da série mensal de produção de petróleo no Brasil, jan/1997–dez/2025 (em m^3).

Variável	Média	DP	EP	Mediana	Assim.	Curtose	Mín.	Máx.	n
Produção (m^3)	10.273.467	3.711.214	198.942	9.892.366	0,34	−0,63	3.629.331	19.861.174	348

Nota: DP = desvio padrão; EP = erro padrão; Assim. = coeficiente de assimetria; Mín. = valor mínimo; Máx. = valor máximo.

Fonte: Elaborado pelo Autor (2026).

A série apresenta um nível médio de produção da ordem de $10^7 \text{ m}^3/\text{mês}$, com coeficiente de variação de 36,1% — elevado, refletindo a forte tendência de longo prazo que atravessa toda a série e que comprime a utilidade da média como medida de posição central. A assimetria positiva moderada (0,34) é consistente com uma cauda direita mais alongada, formada pelos valores mais altos e mais recentes da série; a curtose negativa (−0,63, distribuição platicúrtica) indica menor concentração ao redor do centro do que a normal sugeriria — configuração típica de séries com tendência de crescimento de longo prazo.

A Tabela 3.2 desagrega as estatísticas por mês do ano.

A análise por mês revela a assinatura sazonal dominante da série: retração em fevereiro seguida de recuperação em março. A variação percentual média mês a mês mostra fevereiro com retração média de −9,15% e março com recuperação média de +9,10% em relação ao mês anterior (Figura 3.5). Essa dinâmica está ligada, em boa parte, ao menor número de dias em fevereiro (28–29 dias, contra os 31 dias dos demais meses adjacentes), que implica, mecanicamente, redução proporcional no volume mensal acumulado. Dezembro, agosto e outubro destacam-se como os meses de maior produção média; fevereiro e abril, como os de menor produção média. Esta última leitura — baseada em médias mensais calculadas sobre a série em nível, sem remoção da tendência de longo prazo — é parcialmente revista na Seção 3.2, onde a magnitude do efeito de calendário é quantificada diretamente.

Tabela 3.2: Estatísticas descritivas mensais da produção de petróleo no Brasil (m³), jan/1997 a dez/2025.

Mês	Média	DP	EP	Mediana	Assim.	Curtose	Mín.	Máx.
Janeiro	10.332.976	3.792.987	704.340	10.116.220	0,21	−1,05	3.953.347	17.344.377
Fevereiro	9.335.841	3.340.056	620.233	8.980.299	0,26	−0,95	4.070.983	15.898.965
Março	10.173.933	3.629.740	674.026	10.038.414	0,28	−0,88	4.210.639	17.845.551
Abril	9.893.449	3.498.043	649.570	9.641.935	0,23	−0,91	4.038.285	17.323.715
Maio	10.293.376	3.682.177	683.763	10.092.617	0,19	−0,90	4.028.209	18.132.705
Junho	10.066.885	3.654.977	678.712	9.784.466	0,28	−0,84	4.082.475	17.921.902
Julho	10.534.479	3.914.683	726.938	9.971.773	0,36	−0,75	4.214.599	19.511.945
Agosto	10.607.496	3.948.934	733.299	9.911.445	0,31	−0,91	4.120.922	19.202.227
Setembro	10.316.457	3.878.237	720.171	9.528.102	0,39	−0,78	3.940.067	18.671.815
Outubro	10.586.689	4.003.626	743.455	9.912.636	0,35	−0,67	3.629.331	19.861.174
Novembro	10.271.946	3.733.083	693.216	9.926.112	0,33	−0,83	3.836.152	17.993.617
Dezembro	10.868.083	3.888.169	722.015	10.395.815	0,38	−0,74	4.519.058	19.787.494

$n = 29$ para todos os meses.

Fonte: Elaborado pelo Autor (2026).

Figura 3.1: Série temporal da produção mensal de petróleo no Brasil (jan/1997 – dez/2025).

A linha tracejada dourada representa a tendência LOESS estimada com $\text{span} = 0,25$.

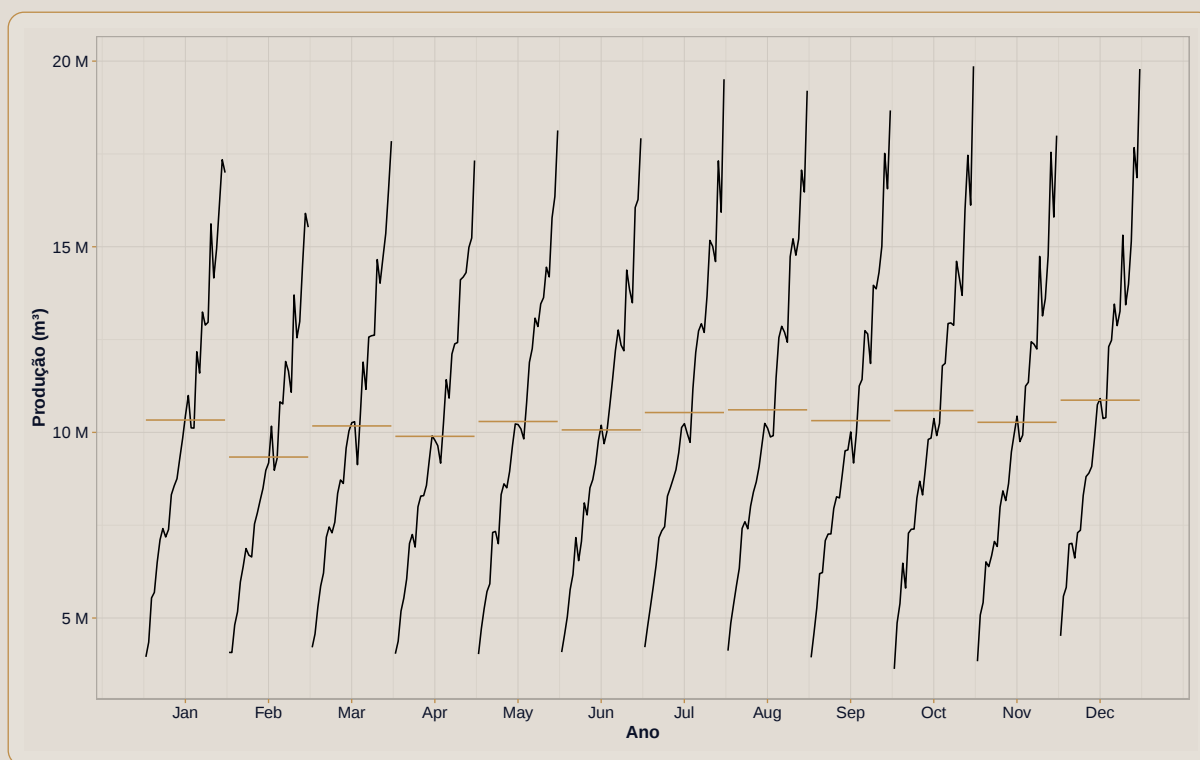


Fonte: Elaborado pelo Autor (2026).

A inspeção visual da Figura 3.1 permite identificar quatro grandes fases da evolução da produção petrolífera brasileira: (i) 1997–2008, crescimento relativamente linear, impulsionado pelos campos convencionais da Bacia de Campos; (ii) 2008–2015, descoberta e início de exploração do pré-sal, com relativa estabilidade de curto prazo decorrente dos longos prazos de desenvolvimento de FPSOs; (iii) 2015–2020, aceleração pronunciada do *ramp-up* dos grandes FPSOs do pré-sal, interrompida por uma desaceleração visível associada à crise da Operação Lava Jato (2014–2016); (iv) 2020–2025, marcada inicialmente pelo choque da pandemia de COVID-19 e, a partir de 2022–2023, por uma nova fase de crescimento acentuado e mais instável, com sucessivos recordes históricos intercalados por uma retração entre 2023 e 2024 — padrão discutido em detalhe nos Capítulos 6 e 8.

Figura 3.2: Gráfico de subséries sazonais da produção de petróleo (jan/1997–dez/2025).

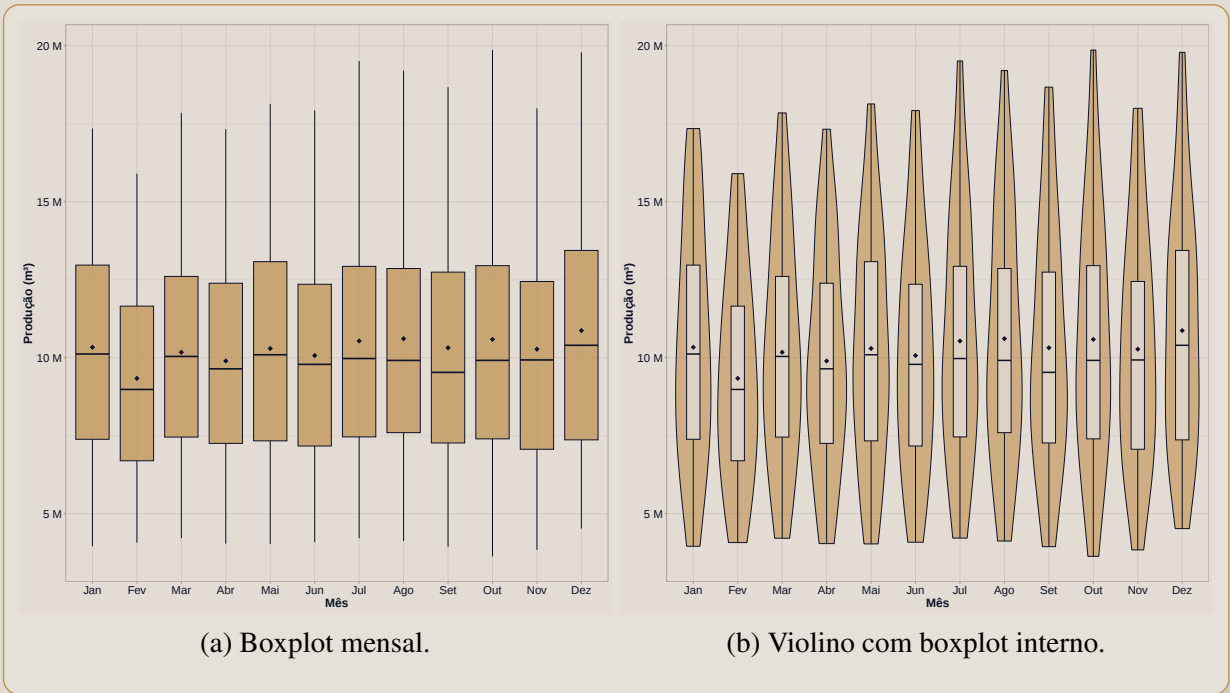
Cada painel mostra a evolução temporal das observações de um determinado mês ao longo dos anos; a linha horizontal representa a média histórica do respectivo mês.



Fonte: Elaborado pelo Autor (2026).

O gráfico de subséries (Figura 3.2) evidencia dois fatos estilizados: a tendência crescente é ubíqua (todos os doze painéis exibem trajetórias ascendentes), e o padrão intra-anual é estável na forma, mas crescente em amplitude absoluta — evidência direta de heterocedasticidade multiplicativa que motiva a transformação logarítmica (Seção 4.1).

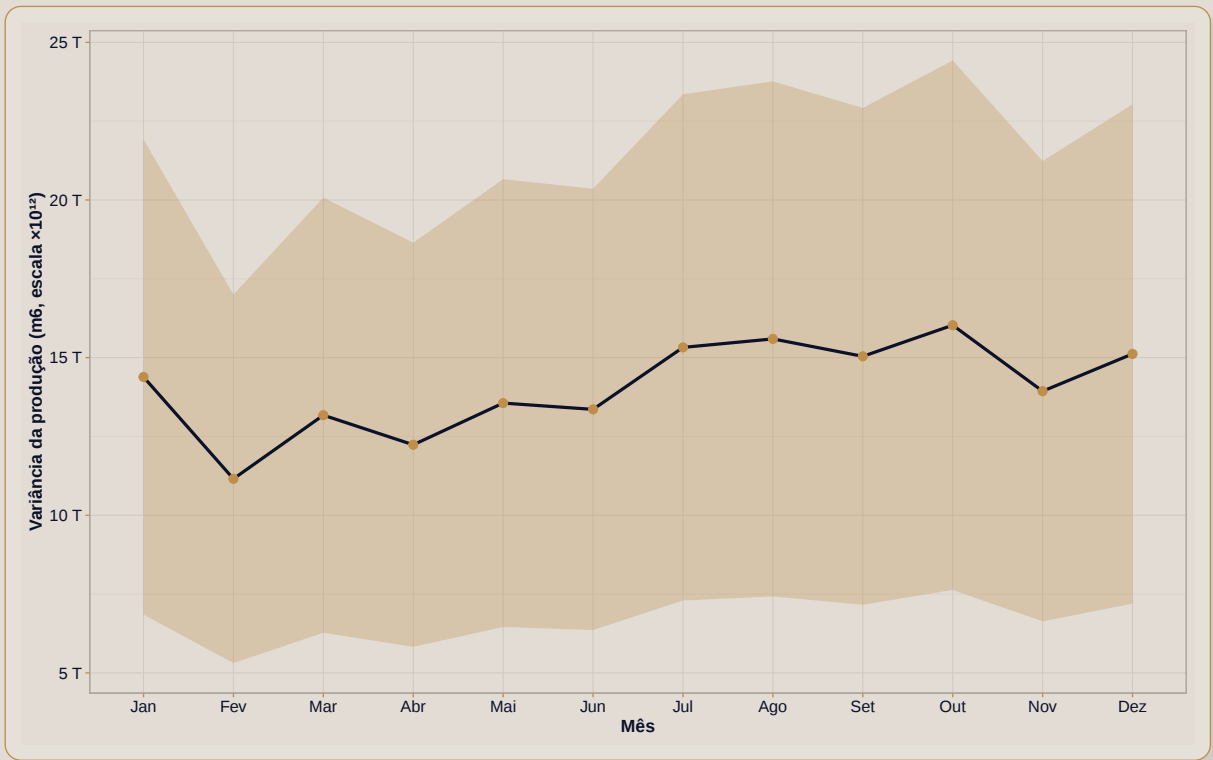
Figura 3.3: Distribuição mensal da produção de petróleo por mês do ano (1997–2025).



Fonte: Elaborado pelo Autor (2026).

Figura 3.4: Variância da produção mensal por mês do ano (1997–2025).

Intervalo de confiança de 95% (faixa dourada).

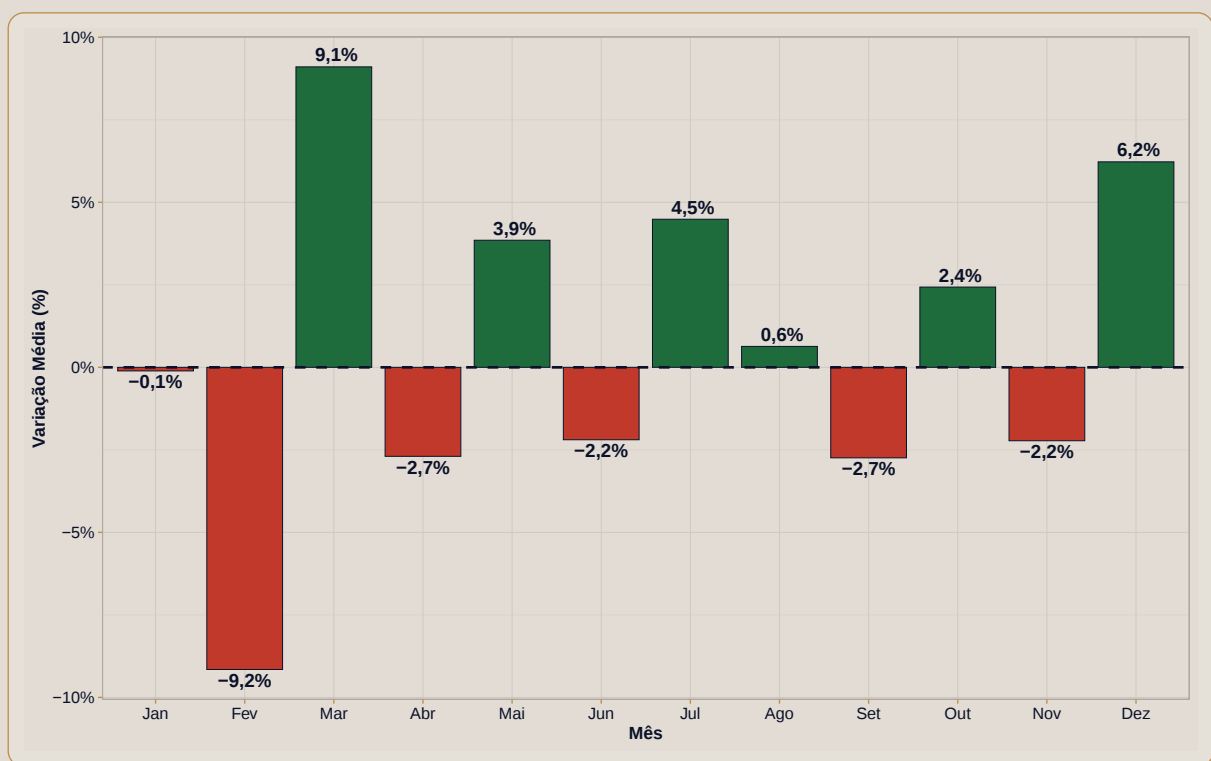


Fonte: Elaborado pelo Autor (2026).

Ressalva de leitura para a Figura 3.4: a variância de cada mês é calculada sobre a série *em nível*, sem remoção prévia da tendência de longo prazo. Como a série cresce de aproximadamente 4 milhões para 20 milhões de m³ ao longo do período analisado, parte substancial da variância mensal aqui reportada reflete a dispersão entre os 29 anos de observações de cada mês ao longo de uma trajetória ascendente — e não exclusivamente a dispersão intra-mês “pura” que o título da figura sugere. A leitura qualitativa (variância crescente ao longo do tempo, evidência adicional de heterocedasticidade multiplicativa) permanece válida e é corroborada por outras evidências independentes (Figura 3.7), mas o leitor deve ter em mente que esta figura específica mistura dispersão intra-anual com a própria tendência de crescimento.

Figura 3.5: Variação (%) média da produção de petróleo em relação ao mês anterior (1997–2025).

Verde: crescimento médio positivo; vermelho: retração média.



Fonte: Elaborado pelo Autor (2026).

O padrão revelado pela Figura 3.5 é dos mais informativos da análise exploratória: a retração de fevereiro e a recuperação de março, discutidas acima, são predominantemente um artefato determinístico da duração diferencial dos meses do calendário gregoriano, e não de um efeito sazonal de natureza econômica ou técnica. Meses com 30 dias (junho, setembro, novembro) também tendem a exibir retrações médias negativas frente aos meses de 31 dias que os antecedem, embora de magnitude menor (entre $-2,2\%$ e $-2,7\%$). Esta leitura qualitativa é testada quantitativamente a seguir.

3.2 Quantificação do Efeito de Calendário

A produção mensal é dividida pelo número de dias do respectivo mês e reescalada para um mês médio de $365,25/12 \approx 30,44$ dias, permitindo comparar a amplitude sazonal antes e depois de remover o efeito puramente mecânico do calendário.

Tabela 3.3: Desvio percentual de cada mês frente à média geral, antes e depois do ajuste por dias-no-mês.

Mês	Desvio bruto (%)	Desvio ajustado por dia (%)
Janeiro	0,58	−1,23
Fevereiro	−9,13	−2,09
Março	−0,97	−2,75
Abril	−3,70	−2,28
Maio	0,19	−1,61
Junho	−2,01	−0,57
Julho	2,54	0,70
Agosto	3,25	1,39
Setembro	0,42	1,90
Outubro	3,05	1,19
Novembro	−0,01	1,46
Dezembro	5,79	3,88

Nota: Desvio ajustado calculado sobre a produção diária média de cada mês (produção mensal / dias no mês), reescalada para um mês de 30,44 dias.

Fonte: Elaborado pelo Autor (2026).

O coeficiente de variação entre as médias mensais cai de 3,84% (série bruta) para 2,05% depois do ajuste por dia — uma redução de 46,6% na dispersão sazonal aparente, confirmando que praticamente metade da variação entre meses observada na Tabela 3.2 e na Figura 3.5 é, de fato, um artefato de calendário. O desvio de fevereiro encolhe de −9,13% para −2,09% depois do ajuste.

Um refinamento importante emerge, no entanto, para março. A Figura 3.5 e a leitura tradicional sugerem que março é um mês de “recuperação” sazonal, com crescimento médio de +9,1% em relação a fevereiro. A Tabela 3.3 mostra que isso é, em boa parte, ilusório: o desvio *ajustado por dia* de março frente à média geral é de −2,75% — mais negativo do que o de fevereiro ajustado (−2,09%). Em outras palavras, março não é um pico de produção genuíno: a recuperação observada na série bruta é majoritariamente o retorno a um nível normal depois de um fevereiro mecanicamente reduzido (28 ou 29 dias), não um efeito sazonal positivo próprio do mês. Já dezembro mantém um desvio ajustado claramente positivo (+3,88%), e abril, ao contrário do que sugeria a leitura bruta de produção “baixa”, tem um desvio ajustado (−2,28%)

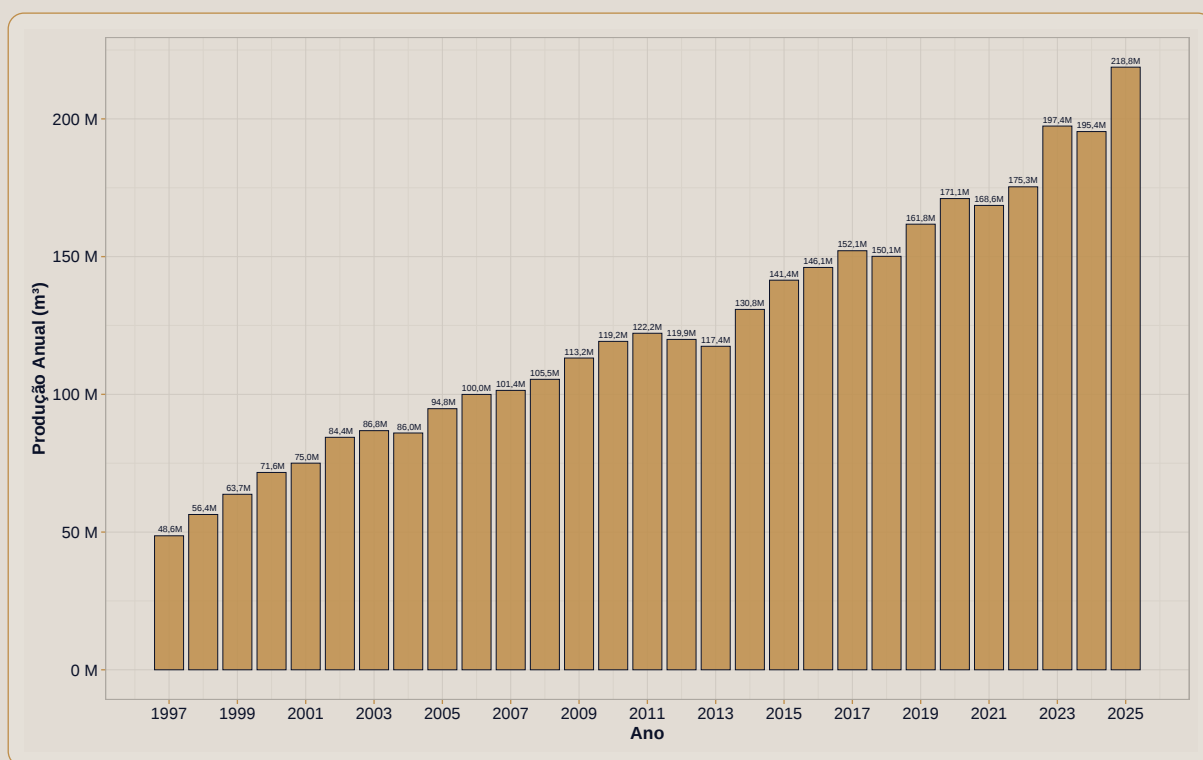
muito próximo do de fevereiro — ou seja, parte do padrão historicamente descrito como “vale de fevereiro, recuperação de março, novo vale em abril” é, na verdade, mais bem descrito como um patamar sazonalmente deprimido ao longo de todo o primeiro quadrimestre do ano, modulado por um efeito de calendário concentrado especificamente em fevereiro.

A fração remanescente da variação sazonal (os 2,05% de coeficiente de variação que sobrevivem ao ajuste) reflete sazonalidade genuína, de natureza operacional e/ou econômica — manutenções programadas de unidades de produção, sazonalidade de demanda, calendário de paradas técnicas — e é o que de fato compete pela energia espectral discutida no Capítulo 7.

3.3 Visão Anual

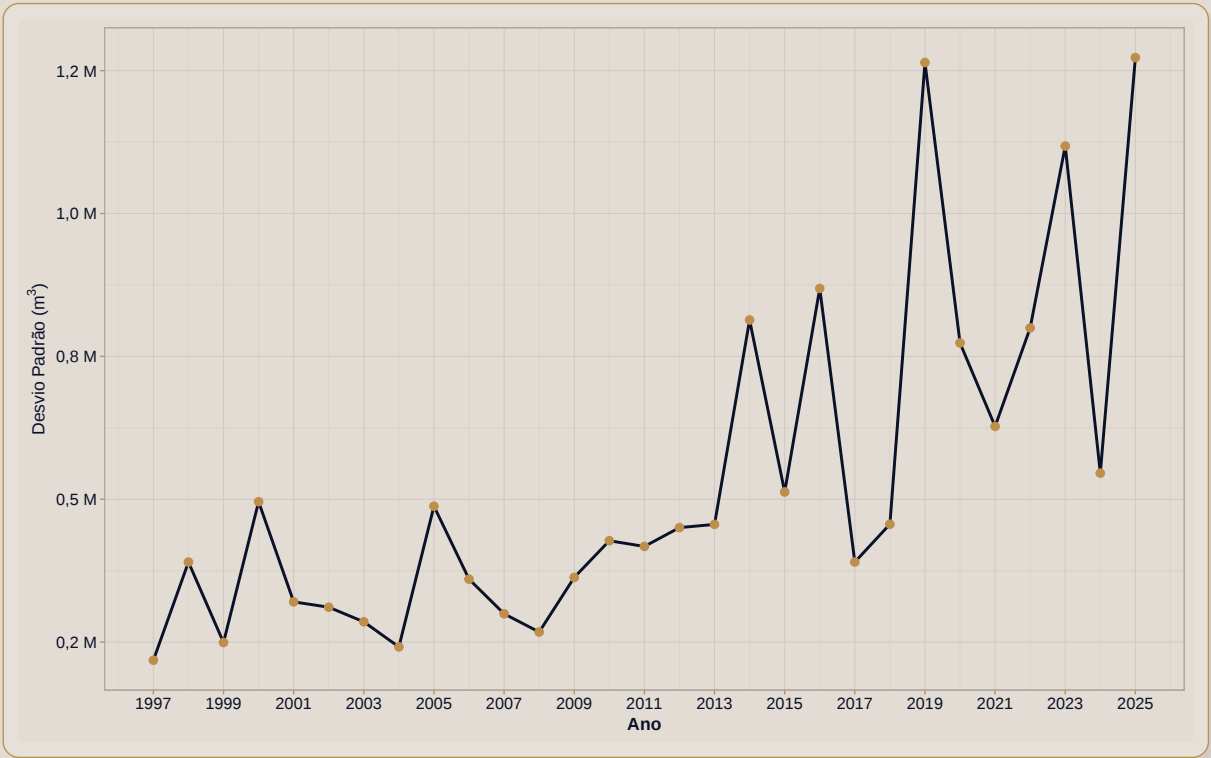
Figura 3.6: Volume Total Anual de Produção de Petróleo no Brasil (1997–2025).

Valores em milhões de m³.



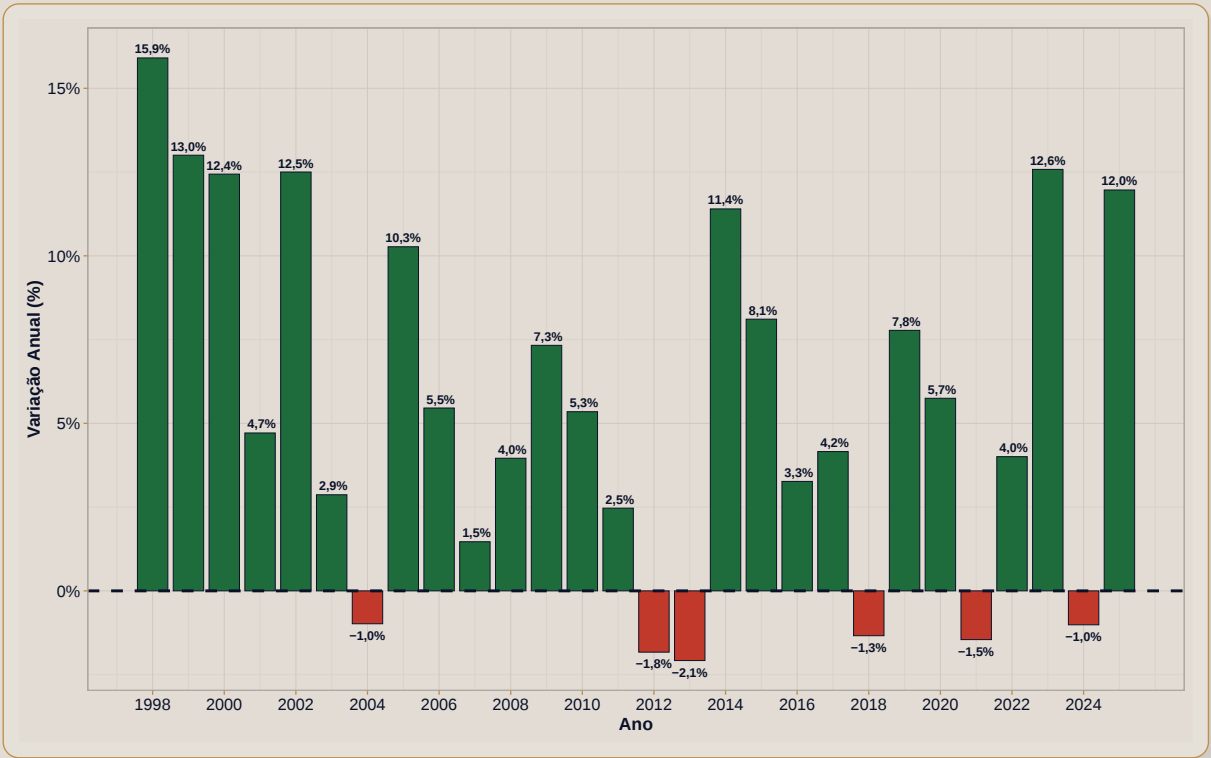
Fonte: Elaborado pelo Autor (2026).

Figura 3.7: Volatilidade Intra-Anual da Produção de Petróleo (1997–2025).
Desvio padrão dos valores mensais dentro de cada ano.



Fonte: Elaborado pelo Autor (2026).

Figura 3.8: Variação Percentual Anual da Produção de Petróleo (YoY).
Verde = crescimento; Vermelho = contração em relação ao ano anterior.

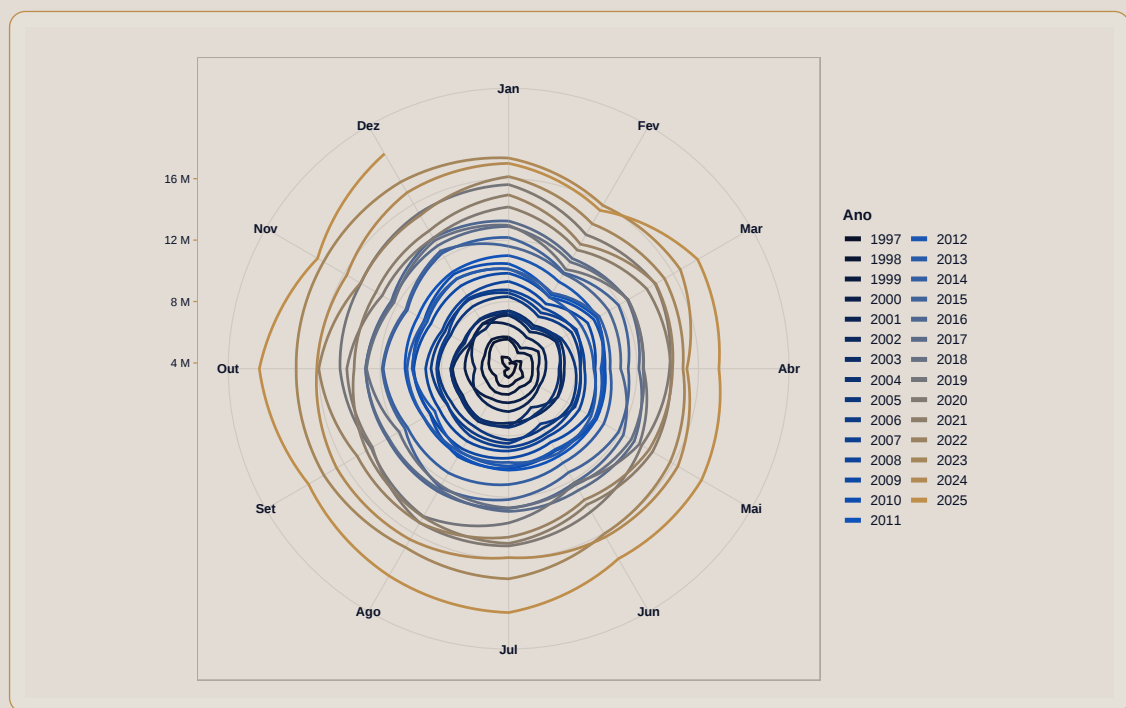


Fonte: Elaborado pelo Autor (2026).

A visão anual (Figura 3.6) confirma a natureza essencialmente crescente da produção agregada, com 2023 e 2025 nos níveis mais altos da série. O desvio padrão anual (Figura 3.7) cresce em paralelo ao nível da série — evidência adicional de heterocedasticidade multiplicativa, desta vez livre da ambiguidade apontada na ressalva sobre a Figura 3.4, já que aqui a dispersão é calculada *dentro* de cada ano, e não entre anos. A variação *year-over-year* (Figura 3.8) mostra que os anos de maior retração foram 2012 e 2013 (anos que antecederam à operação Lava Jato (2014–2016)); outras oscilações ocorreram em períodos espaçados: 2004, 2018, 2021 e 2024. Em particular a desaceleração do ano de 2024 em relação a 2023 seguida por forte recuperação em 2025 é assinatura típica de uma produção determinada por adições discretas de capacidade (entrada em operação de novas unidades de produção), e não por um processo suave de crescimento orgânico.

Figura 3.9: Gráfico Sazonal Polar — Ciclicidade Anual da Produção.

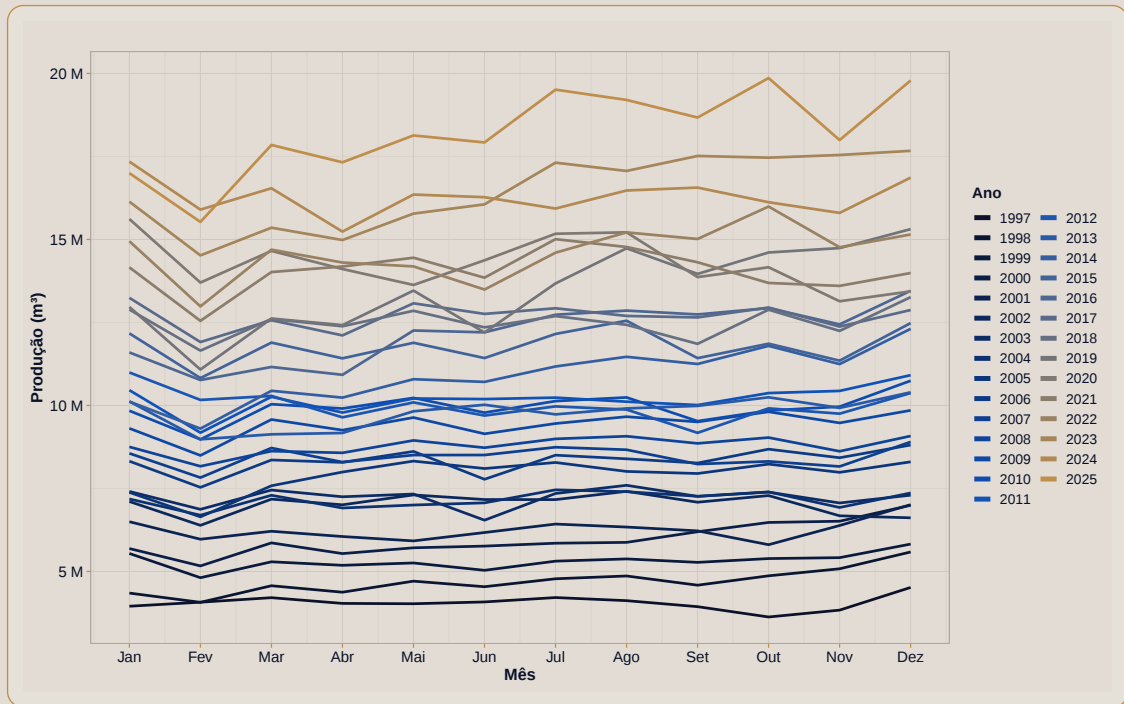
Cada cor representa um ano (1997–2025).



Fonte: Elaborado pelo Autor (2026).

Figura 3.10: Gráfico sazonal da produção de petróleo (1997–2025).

Cada linha representa um ano; a dispersão vertical em cada ponto indica a variabilidade inter-anual para aquele mês.



Fonte: Elaborado pelo Autor (2026).

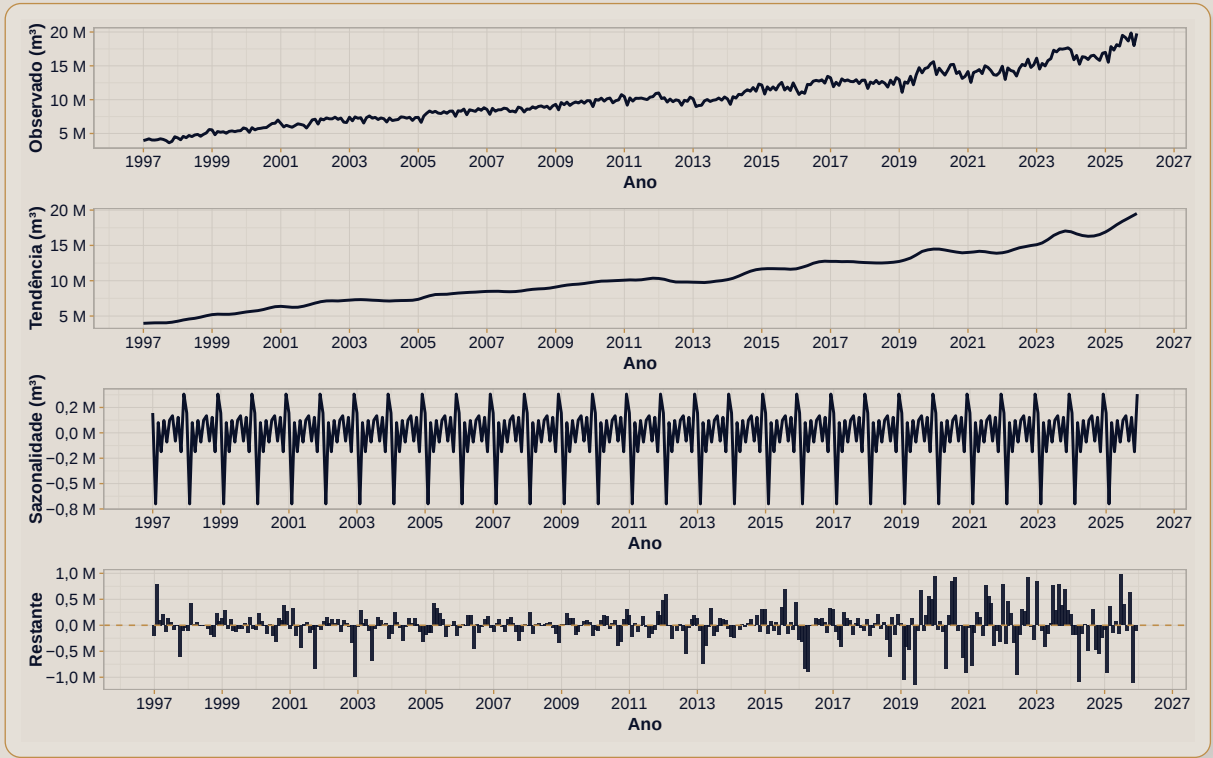
O gráfico sazonal (Figura 3.10) oferece a evidência visual mais direta para a pergunta central deste estudo. As linhas anuais exibem um padrão morfológico reconhecível, com o vale de fevereiro e os picos de março, maio, agosto e outubro repetindo-se ao longo de quase três décadas, progressivamente deslocado para cima — consistente com sazonalidade estável na forma e crescente em amplitude absoluta. Esta leitura visual é corroborada pelos testes formais da Seção 4.3, mas deve ser qualificada tanto pela quantificação do efeito de calendário (Seção 3.2, que mostra que o “pico de março” é majoritariamente aparente) quanto pela análise espectral do Capítulo 7, que mostra que a energia desse ciclo está distribuída por vários harmônicos, e não concentrada isoladamente na frequência fundamental de 12 meses.

3.4 Decomposição da Série

A escolha do modelo multiplicativo, $Y_t = T_t \times S_t \times \varepsilon_t$, é diretamente motivada pela evidência da Seção anterior: a amplitude das flutuações sazonais cresce proporcionalmente ao nível da série, de modo que $DP(Y_t | T_t) \propto T_t$ (equivalente, $\text{Var}(Y_t | T_t) \propto T_t^2$), violando o pressuposto de variância constante do modelo aditivo. O modelo multiplicativo é formalmente equivalente ao modelo aditivo aplicado ao logaritmo: $\ln(Y_t) = \ln(T_t) + \ln(S_t) + \ln(\varepsilon_t)$.

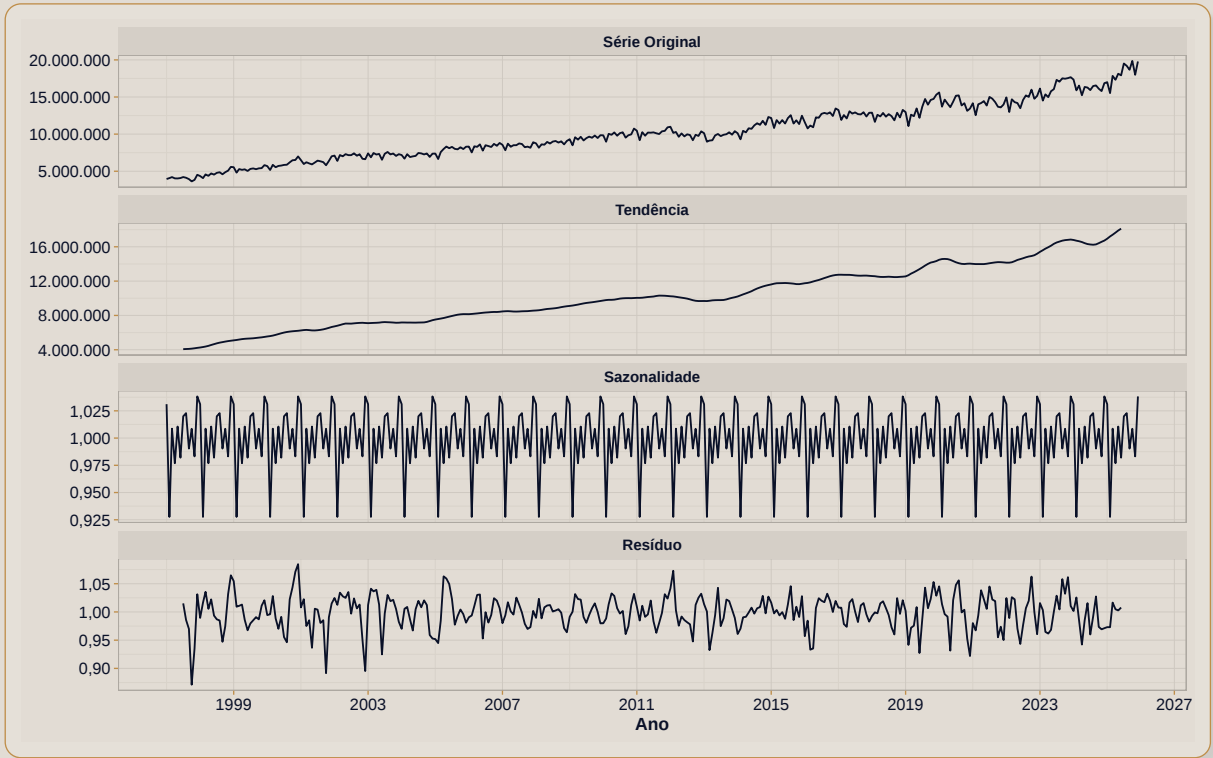
Além da decomposição clássica, aplica-se a decomposição STL Cleveland et al. (1990), que oferece robustez a *outliers*, permite modelar uma sazonalidade variante no tempo e ausência do problema de efeitos de borda.

Figura 3.11: Decomposição STL da série de produção de petróleo (1997–2025).
Parâmetros: s.window = "periodic", t.window = 13, robust = TRUE.



Fonte: Elaborado pelo Autor (2026).

Figura 3.12: Decomposição clássica multiplicativa da série de produção de petróleo (1997–2025).
Modelo: $Y_t = \text{Tendência} \times \text{Sazonalidade} \times \text{Resíduo}$.



Fonte: Elaborado pelo Autor (2026).

O componente de tendência revela uma trajetória de crescimento segmentada em fases com diferentes taxas, como já discutido. O componente residual da decomposição clássica não é ruído branco: picos de magnitude superior ao usual aparecem em períodos específicos que coincidem com os eventos de intervenção modelados formalmente no Capítulo 6 e com as datas de quebra estrutural identificadas pelos testes de CUSUM e Bai-Perron (Seção 6.3) — um dos argumentos mais fortes para a necessidade de uma análise de intervenção e de quebra estrutural formal, e não apenas descritiva.

CAPÍTULO 4

ANÁLISE DE ESTACIONARIEDADE E TRANSFORMAÇÕES

4.1 Transformação Logarítmica

A heterocedasticidade multiplicativa documentada no Capítulo 3 é tratada pela transformação logarítmica $Z_t = \ln(Y_t)$, que estabiliza a variância ao converter flutuações proporcionais ao nível em flutuações de amplitude aproximadamente constante. Esta escolha é corroborada visualmente pela comparação entre as Figuras 3.1 e 4.1.

Figura 4.1: Série temporal após transformação logarítmica, $\ln(Y_t)$.



Fonte: Elaborado pelo Autor (2026).

A escolha de $\lambda = 0$ (transformação log) acima é corroborada por evidência visual e também pela verificação quantitativa do parâmetro de Box-Cox; `forecast::BoxCox.lambda()` (Seção 2.2). O método de Guerrero — recomendado para séries com sazonalidade — estima $\hat{\lambda}_{\text{Guerrero}} = 0,001$, praticamente indistinguível de zero e corroborando fortemente a escolha de

$\lambda = 0$ já adotada. O método baseado em log-verossimilhança estima $\hat{\lambda}_{\loglik} = 0,5$, divergindo substancialmente — mas esse método é conhecidamente instável para séries com autocorrelação forte e tendência pronunciada, pois ele maximiza a verossimilhança de uma regressão linear simples no tempo sem considerar a estrutura de dependência temporal da série. Diante da convergência entre o método mais adequado ao caso (Guerrero), a evidência visual de heterocedasticidade multiplicativa (Capítulo 3) e a prática consolidada na literatura de séries de produção em unidades físicas, mantém-se $\lambda = 0$ em todos os modelos deste relatório.

4.2 Testes Formais de Raiz Unitária e de Tendência

Aplicam-se três testes complementares — *Augmented Dickey-Fuller* (ADF), *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) e *Phillips-Perron* (PP) — a quatro representações da série: nível original, log, primeira diferença e log com primeira diferença. A lógica de triangulação é: estacionariedade é declarada quando ADF e PP rejeitam a hipótese de raiz unitária e, simultaneamente, o KPSS não rejeita a hipótese de estacionariedade. A Tabela 4.1 reporta os resultados.

Tabela 4.1: Testes formais de raiz unitária e estacionariedade.

Série	ADF		KPSS		PP	
	Estatística	Valor- p	Estatística	Valor- p	Estatística	Valor- p
Original (Y_t)	-2,437	0,392	5,640	< 0,01	-110,2	< 0,01
Log ($\ln Y_t$)	-3,666	0,027	5,581	< 0,01	-73,44	< 0,01
1ª diferença (ΔY_t)	-8,293	< 0,01	0,096	> 0,10	-474,1	< 0,01
Log + 1ª diferença ($\Delta \ln Y_t$)	-8,315	< 0,01	0,064	> 0,10	-442,7	< 0,01

Fonte: Elaborado pelo Autor (2026).

A leitura conjunta da Tabela 4.1 aponta, de forma majoritária mas não unânime, para a não estacionariedade da série em nível. O teste KPSS rejeita a hipótese de estacionariedade em torno de nível tanto para a série original quanto para a série em log ($p < 0,01$ em ambos os casos) — o único dos três testes a fazê-lo de forma consistente nas duas representações. O ADF não rejeita a hipótese de raiz unitária para a série original ($p = 0,392$), mas rejeita-a para a série em log ($p = 0,027$); o teste PP, por sua vez, rejeita a hipótese de raiz unitária em ambas as representações ($p < 0,01$)¹.

Ou seja: nenhum dos três testes concorda integralmente entre si nas duas representações em nível — o ADF e o PP chegam a sugerir estacionariedade da série em log isoladamente, mas

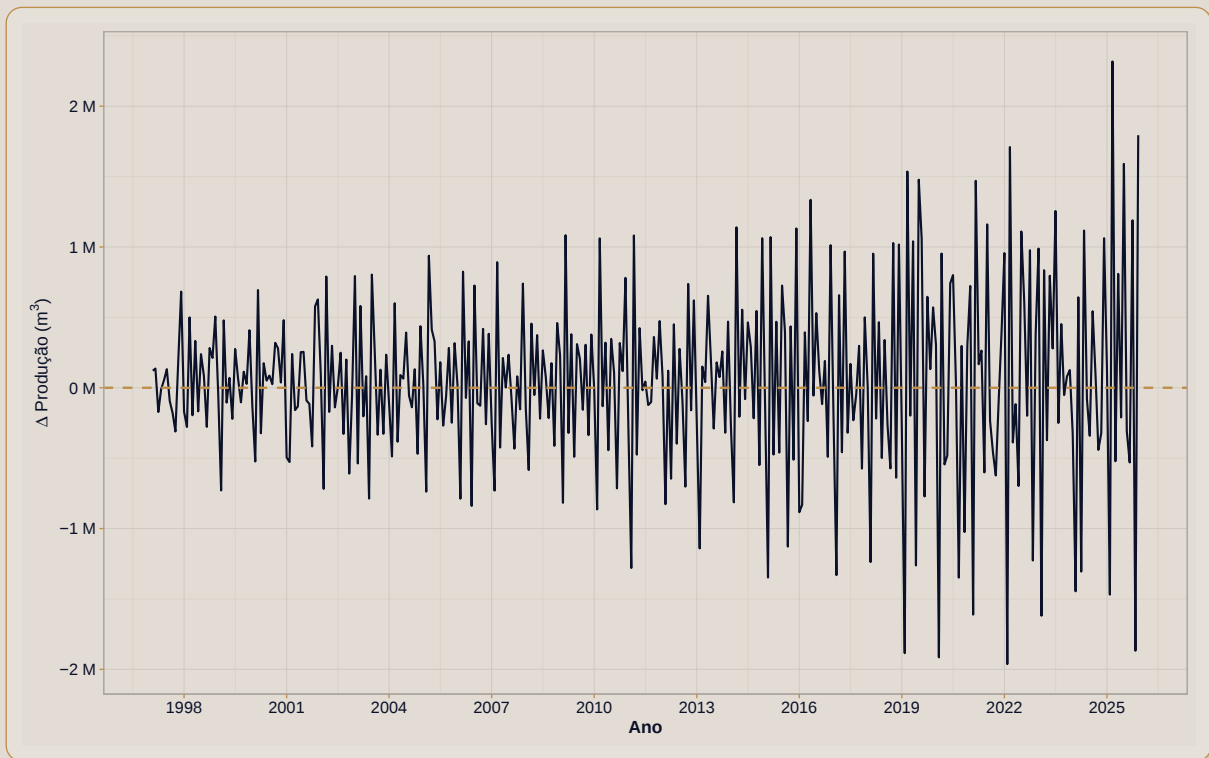
¹ As estatísticas de Phillips e Perron (1988) reportadas na Tabela 4.1 (-110,2 na série original; -73,44 em log) correspondem ao teste $Z(\alpha)$ — baseado no viés de segunda ordem do estimador OLS do coeficiente autorregressivo, com parâmetro de truncamento de defasagem igual a 5 (`tseries::pp.test()`) — e não devem ser comparadas, em magnitude, à estatística t do ADF: sua escala é estruturalmente maior por construção, não indício de um erro de cálculo ou de transcrição.

o KPSS nunca a confirma. Como o critério de estacionariedade adotado neste trabalho (Capítulo 2) exige a concordância simultânea dos três testes, e não uma maioria simples, essa divergência parcial entre ADF/PP e KPSS não é ignorada, mas resolvida pela própria regra de decisão: a ausência de confirmação do KPSS é suficiente para que nenhuma das duas representações em nível seja declarada estacionária. Já a primeira diferença, aplicada à série original ou à série logarítmica, satisfaz integralmente os três testes simultaneamente — ADF e PP rejeitam raiz unitária, e o KPSS não rejeita estacionariedade, com unanimidade total entre os três testes, tanto na diferença simples quanto na diferença do log. Conclui-se que $d = 1$ é suficiente para conferir estacionariedade na componente não sazonal, confirmando a integração de ordem 1, $Y_t \sim I(1)$.

Essa necessidade de diferenciação é corroborada analiticamente pela forte presença de inclinação de longo prazo na série. O teste de Mann-Kendall Mann (1945), Kendall (1975) confirma uma tendência monotônica altamente significativa na série original: $\tau = 0,908$ ($p < 0,001$), valor extremamente próximo do limite superior teórico ($\tau = 1$). Esse resultado reforça que praticamente toda a ordenação temporal da série é consistente com uma trajetória de crescimento — comportamento coerente com a quase ausência de reversões de longo prazo visível na Figura 3.1. Adicionalmente, o teste de Cox e Stuart (1955), incorporado como verificação complementar, valida essa dinâmica de forma independente ao registrar uma estatística de 174 ($p < 0,001$), rejeitando categoricamente a hipótese nula de ausência de tendência.

Figura 4.2: Série após primeira diferenciação.

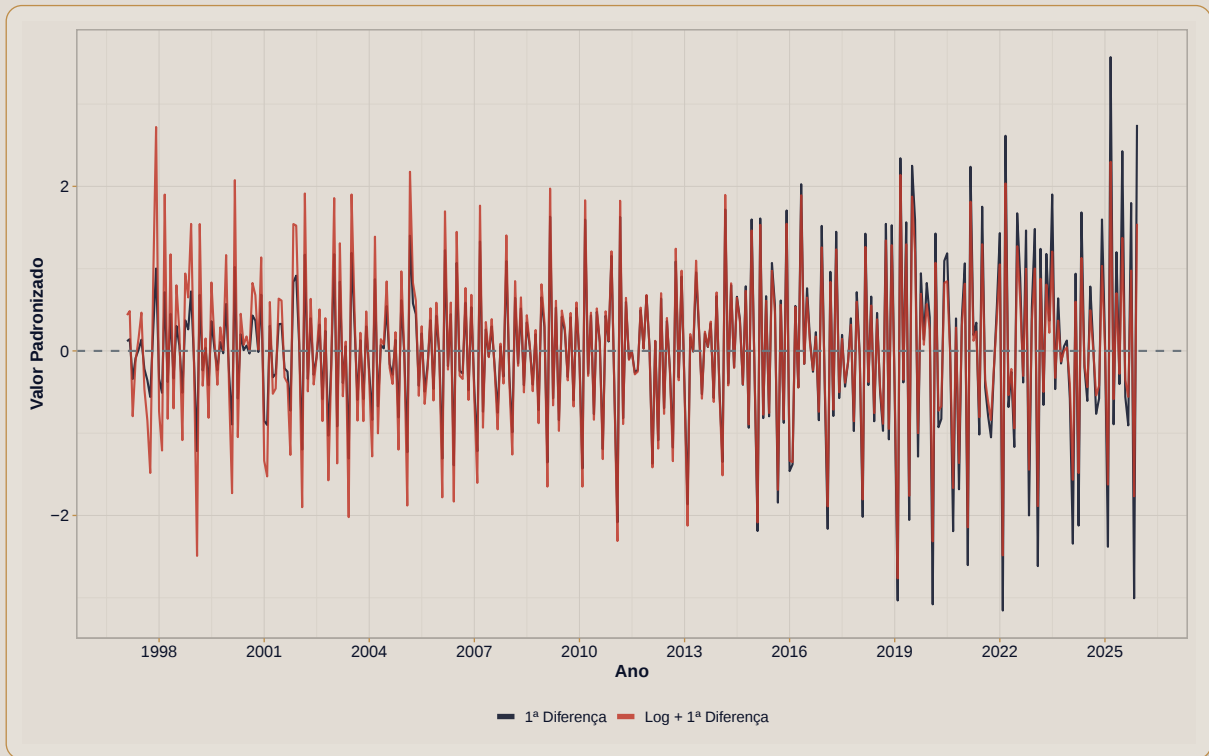
$$\Delta Y_t = Y_t - Y_{t-1}$$



Fonte: Elaborado pelo Autor (2026).

Figura 4.3: Comparação das transformações para estacionariedade.

Séries padronizadas (média 0, desvio-padrão 1) para comparação visual direta da estabilidade da variância.



Fonte: Elaborado pelo Autor (2026).

A Figura 4.3 sobrepõe a primeira diferença da série em nível e a primeira diferença da série em log, ambas padronizadas, e oferece a comparação visual mais direta da estabilização de variância obtida pela transformação logarítmica: a diferença em nível exibe amplitude visivelmente crescente ao longo do tempo, enquanto a diferença logarítmica mantém amplitude aproximadamente constante — a assinatura esperada de uma transformação que efetivamente neutraliza a heterocedasticidade multiplicativa documentada no Capítulo 3.

4.3 Diferenciação Sazonal

A presença de sazonalidade estável é avaliada formalmente pelo teste QS (*Quadratic Spectral*, `seastests::qs()`): estatística = 265,45, $p < 0,001$, rejeitando fortemente H_0 de ausência de sazonalidade estável (ciclo anual, frequência 12). Essa evidência motiva a aplicação adicional de diferenciação sazonal, $\Delta_{12}\Delta Y_t = (1 - B)(1 - B^{12})Y_t$, padrão na literatura para séries mensais com sazonalidade determinística forte e dependência temporal residual (BOX; JENKINS, 1970; HYNDMAN; ATHANASOPOULOS, 2018).

Verificação formal da ordem D . O teste QS acima estabelece a *presença* de sazonalidade estável, mas não testa, por si só, se ela deve ser tratada como uma raiz unitária sazonal ($D = 1$, diferenciação) ou como um padrão determinístico ($D = 0$, modelado via regressores — Capítulo 7). Esta distinção é avaliada formalmente com os testes OCSB (OSBORN et al., 1988) e de

Canova-Hansen (CANOVA; HANSEN, 1995) (`forecast::nsdiffs()`):

Tabela 4.2: Testes formais para a ordem de diferenciação sazonal D , e sua adesão pelos candidatos SARIMA.

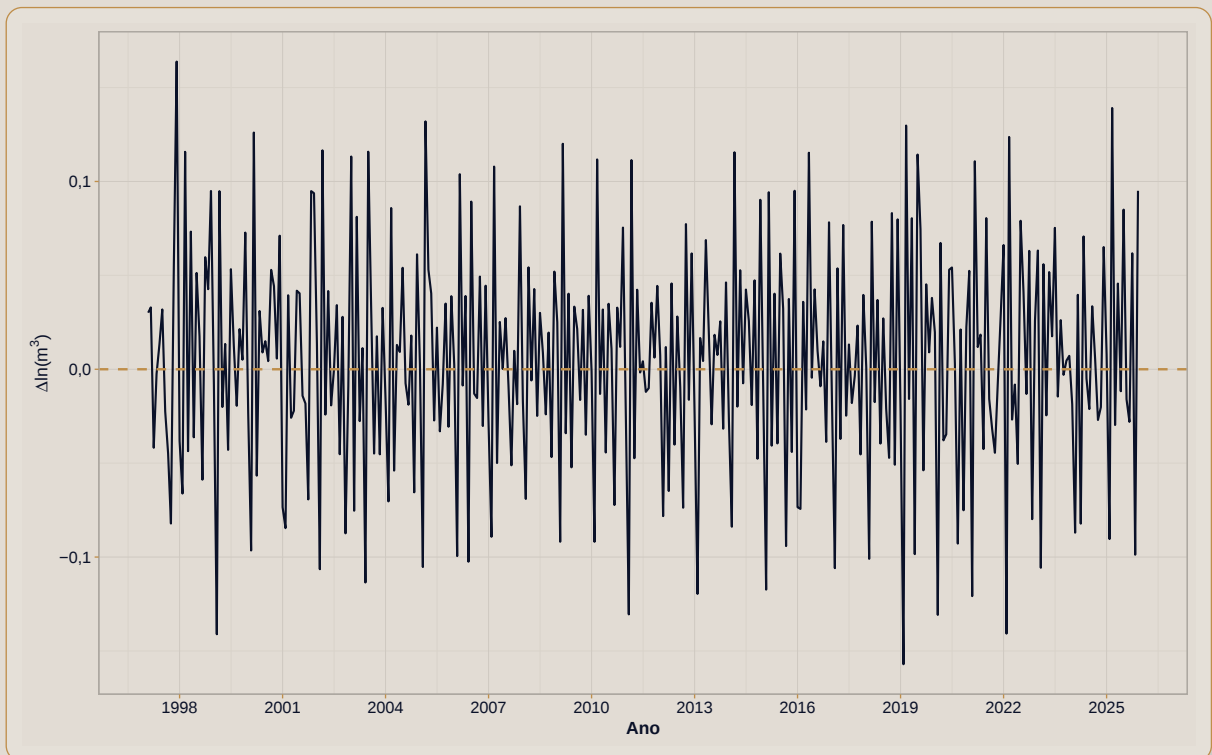
Teste	D recomendado
OCSB (OSBORN et al., 1988)	0
Canova-Hansen (CANOVA; HANSEN, 1995)	0
Candidatos SARIMA com $D = 1$ (Capítulo 5)	3 de 8
Candidatos SARIMA com $D = 0$ (Capítulo 5)	5 de 8

Fonte: Elaborado pelo Autor (2026).

Os testes formais recomendam $D = 0$. Diferentemente do que uma leitura apresada do conjunto de candidatos poderia sugerir, isso não representa uma contradição generalizada com os modelos SARIMA estimados neste trabalho: dos oito candidatos (Capítulo 5), cinco — SARIMA(1, 1, 1)(1, 0, 1)₁₂, SARIMA(8, 1, 1)(1, 0, 1)₁₂, SARIMA(0, 1, 1)(1, 0, 1)₁₂, SARIMA(1, 1, 2)(1, 0, 1)₁₂ e o modelo selecionado por `auto.arima()`, ARIMA(2, 1, 1)(2, 0, 0)₁₂ com deriva — já empregam $D = 0$, em linha direta com a recomendação dos testes OCSB e Canova-Hansen. Apenas três candidatos — SARIMA(1, 1, 1)(0, 1, 1)₁₂, SARIMA(8, 1, 1)(0, 1, 1)₁₂ e SARIMA(2, 1, 4)(0, 1, 1)₁₂ — mantêm $D = 1$, incluídos no conjunto para permitir comparação direta, por AICc/BIC e por acurácia preditiva (Capítulo 8), entre as duas famílias de especificação sazonal. O próprio resultado da competição por AICc (Capítulo 5) reforça essa leitura: os quatro modelos com menor AICc são justamente os quatro candidatos com $D = 0$, e o modelo campeão — SARIMA(8, 1, 1)(1, 0, 1)₁₂ — é um deles. Ou seja, a evidência formal (OCSB, Canova-Hansen) e a evidência empírica (ranking por AICc) convergem para a mesma conclusão: nesta série, a sazonalidade é mais bem tratada por diferenciação sazonal parcial ($D = 0$) combinada com termo autorregressivo sazonal e deriva determinística, do que por diferenciação sazonal completa ($D = 1$). Esse achado é consistente com a conclusão da Análise Espectral (Capítulo 7) de que a sazonalidade desta série é melhor descrita como um padrão determinístico (harmônico) do que como uma raiz unitária sazonal estocástica. Por essa razão, a especificação ARIMA + Fourier com deriva ($D = 0$), explorada no Capítulo 7 e formalmente comparada no Capítulo 8, é tratada neste relatório como concorrente de primeira classe da família SARIMA.

Figura 4.4: Série log-diferenciada $\Delta \ln(Y_t)$.

Taxa de crescimento mensal aproximada.



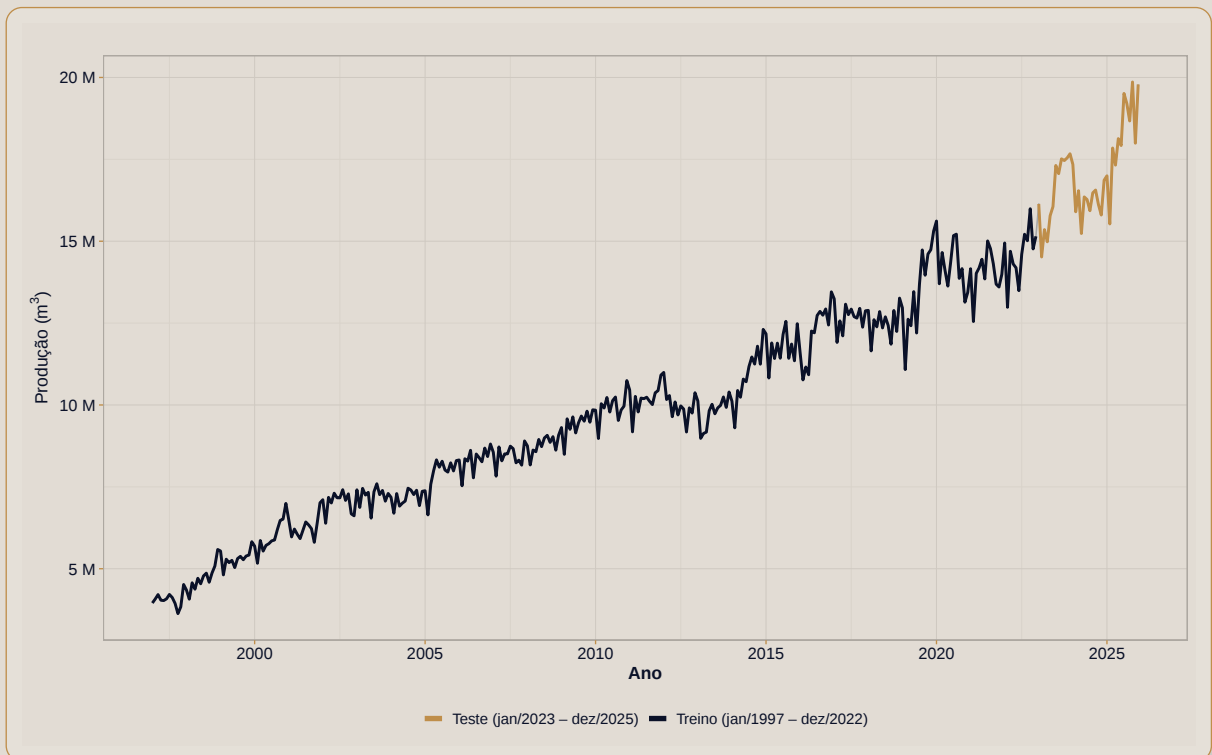
Fonte: Elaborado pelo Autor (2026).

4.4 Divisão Treino/Teste

A amostra foi particionada em um conjunto de treino, de janeiro de 1997 a dezembro de 2022, e um conjunto de teste, de janeiro de 2023 a dezembro de 2025, utilizado exclusivamente para avaliação *out-of-sample* no Capítulo 8.

Figura 4.5: Partição da Série em Conjuntos de Treino e Teste.

Treino: 312 observações (90%) | Teste: 36 observações (10%).



Fonte: Elaborado pelo Autor (2026).

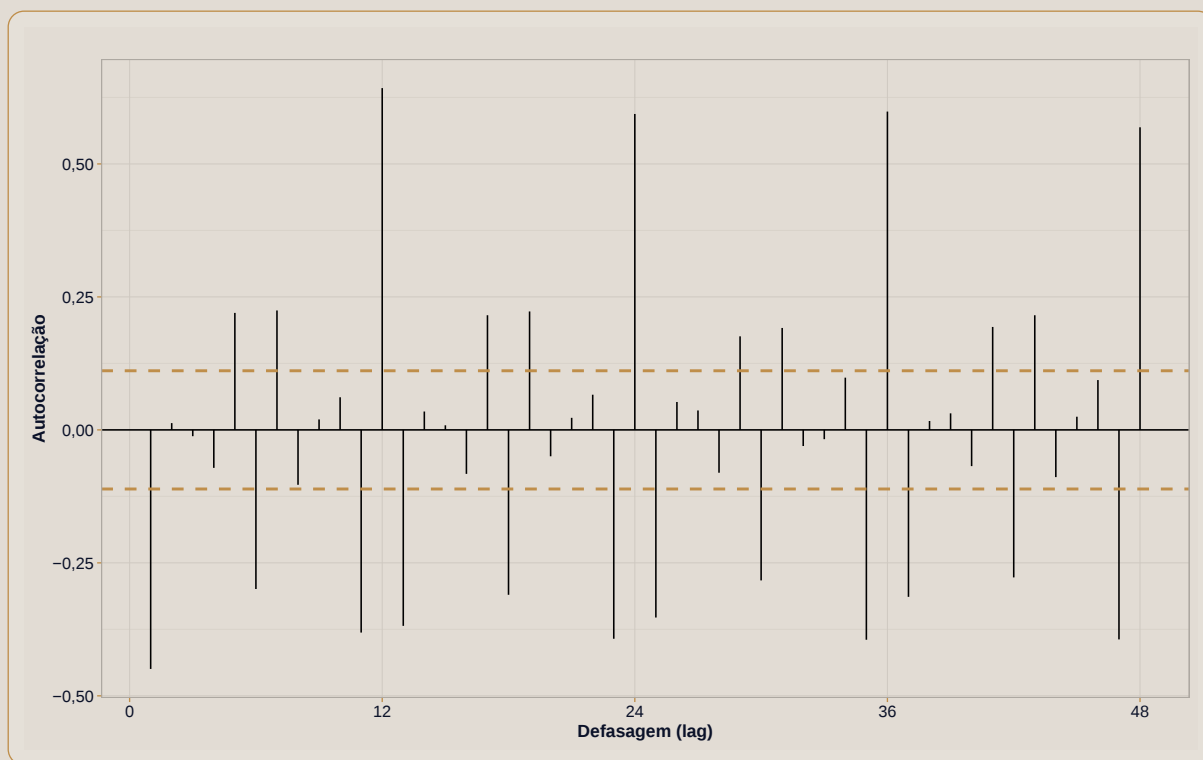
Uma ressalva metodológica importante, detalhada no Capítulo 8: o uso de um único corte de 36 meses para avaliação *out-of-sample* expõe a comparação de modelos a alta variância amostral, sobretudo quando — como discutido no Capítulo 6 — a fronteira do corte está próxima de uma região de instabilidade da série. A validação *rolling-origin* (Seção 2.7) foi adotada como verificação de robustez complementar a este holdout único.

CAPÍTULO 5

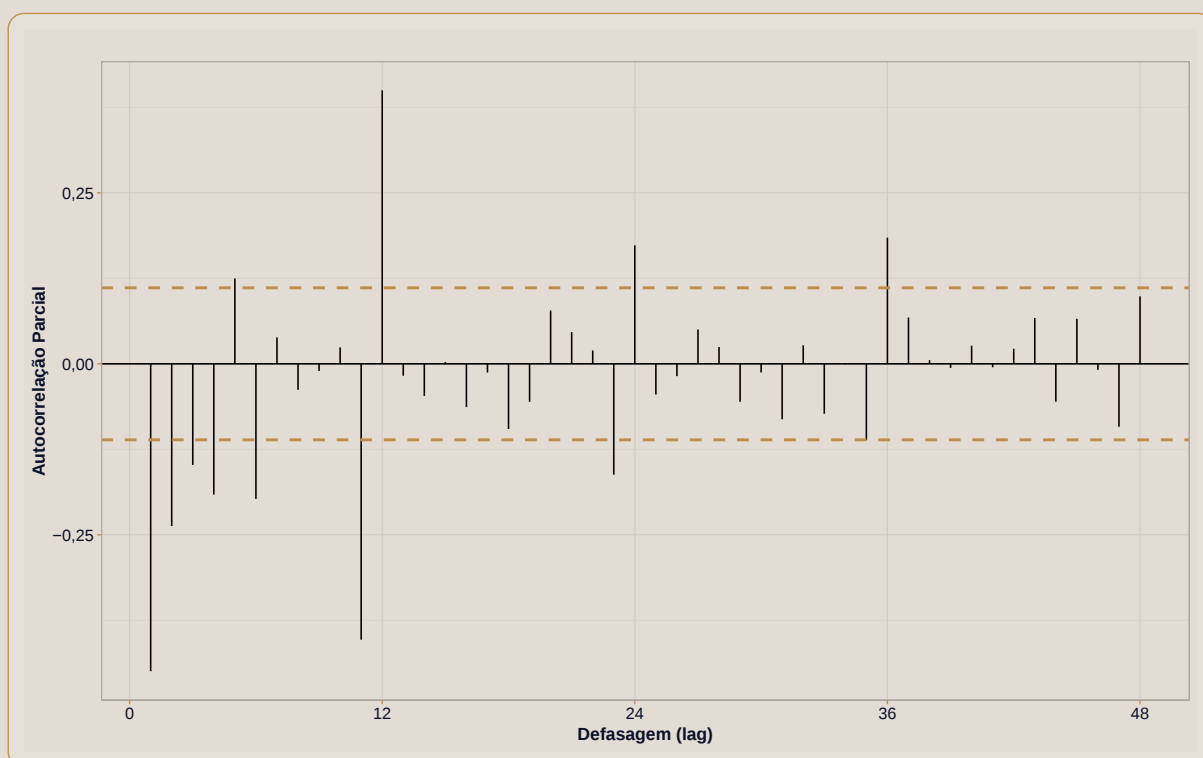
IDENTIFICAÇÃO E ESTIMAÇÃO DOS MODELOS SARIMA

5.1 Identificação via FAC, FACP e EACF

A função de autocorrelação (FAC) (Figura 5.1) e a função de autocorrelação parcial (FACP) (Figura 5.2) da série $\Delta \ln(Y_t)$ orientam a identificação preliminar das ordens (p, q) e (P, Q) . A FACP exibe decaimento lento nas defasagens múltiplas de 12, sugerindo $Q \geq 1$; picos isolados em defasagens não sazonais sugerem componentes AR de ordem baixa, possivelmente com contribuição residual em defasagens mais altas. A função EACF (*Extended ACF*, Tsay e Tiao (1984)), calculada via `TSA::eacf()`, indica uma região de vértice compatível com $(p, q) \in \{(0, 1), (1, 2), (2, 4)\}$. A especificação mais parcimoniosa $(1, 1)$, não está estritamente indicada pela EACF, mas é mantida como *benchmark* de referência; já as ordens autorregressivas elevadas ($p = 8$) não emergem diretamente da EACF (cuja grade testada não se estende além de $p = 5$), mas da evidência da FACP em defasagens mais altas, e são incorporadas de forma parcimoniosa, com os coeficientes intermediários $(\phi_2, \dots, \phi_7)$ fixados em zero.

Figura 5.1: FAC da série $\Delta \ln(Y_t)$ (treino, 1997–2022).

Fonte: Elaborado pelo Autor (2026).

Figura 5.2: FACP da série $\Delta \ln(Y_t)$ (treino, 1997–2022).

Fonte: Elaborado pelo Autor (2026).

5.2 Modelos Candidatos

Com base na identificação preliminar, oito especificações SARIMA foram ajustadas por máxima verossimilhança (`forecast::Arima()`, $\lambda = 0$) sobre o conjunto de treino, acompanhadas do candidato selecionado automaticamente por `auto.arima()` sob busca exaustiva (`stepwise = FALSE`, `approximation = FALSE`). Conforme a distinção estabelecida na Seção 2.2, três candidatos — com diferenciação sazonal completa ($D = 1$) — foram estimados com `include.drift = FALSE`¹, enquanto os cinco candidatos com $D = 0$ (incluindo o selecionado por `auto.arima()`) foram estimados com `include.drift = TRUE`. A Tabela 5.1 resume os critérios AICc e BIC de cada candidato.

Tabela 5.1: Comparação dos modelos SARIMA candidatos por AICc e BIC (conjunto de treino, 1997–2022).

Modelo	AICc	BIC
SARIMA(8,1,1)(1,0,1) ₁₂	−1.174 [‡]	−1.148
SARIMA(1,1,1)(1,0,1) ₁₂	−1.173	−1.151 [†]
SARIMA(1,1,2)(1,0,1) ₁₂	−1.171	−1.145
SARIMA(0,1,1)(1,0,1) ₁₂	−1.165	−1.146
SARIMA(8,1,1)(0,1,1) ₁₂	−1.136	−1.117
SARIMA(1,1,1)(0,1,1) ₁₂	−1.135	−1.120
SARIMA(2,1,4)(0,1,1) ₁₂	−1.132	−1.103
ARIMA(2,1,1)(2,0,0) ₁₂ c/ deriva (= <code>auto.arima</code>)	−1.120	−1.095

Nota: [‡] Menor AICc. [†] Menor BIC.

Fonte: Elaborado pelo Autor (2026).

O modelo SARIMA(8,1,1)(1,0,1)₁₂ apresenta o menor AICc entre os candidatos. Nele, os coeficientes autorregressivos de ordem 2 a 7 foram restritos a zero (`fixed = c(NA, 0, 0, 0, 0, 0, 0, NA, NA, NA, NA, NA)`), preservando ϕ_1 , ϕ_8 , θ_1 , Φ_1 (componente sazonal autorregressiva), Θ_1 e a deriva como parâmetros livres — seis parâmetros ao todo, não os onze que a notação (8,1,1)(1,0,1)₁₂ sugeriria sob uma leitura literal e não restrita da ordem AR de 8. Esse ponto é retomado na discussão de sobreajuste do Capítulo 8.

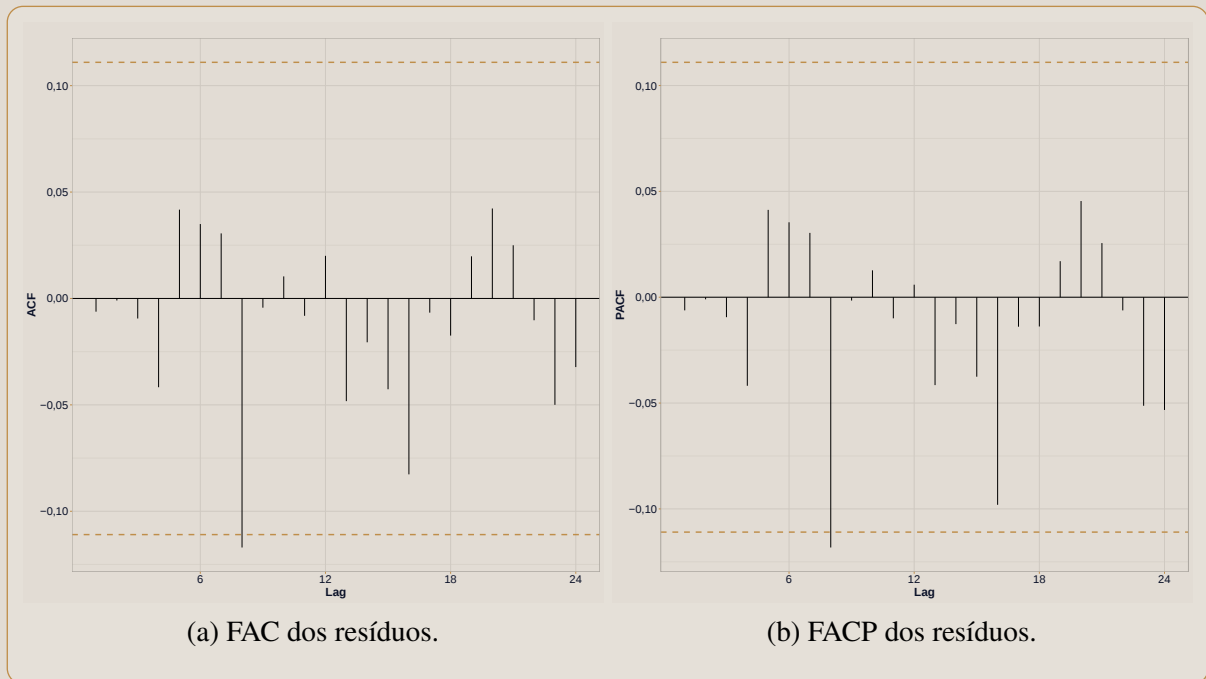
A margem de vitória do SARIMA(8,1,1)(1,0,1)₁₂ sobre o segundo colocado por AICc, SARIMA(1,1,1)(1,0,1)₁₂, é de apenas 1 ponto, abaixo do limiar de 2 pontos frequentemente citado como diferença substancial (BURNHAM; ANDERSON, 2002). A coluna de BIC confirma que essa fragilidade se traduz em uma reversão: como o BIC pune mais fortemente a complexidade ($k \ln n$ em vez de $2k$), o parâmetro livre adicional do SARIMA(8,1,1) (a componente ϕ_8 , frente à especificação mais simples ϕ_1 apenas) deixa de compensar o ganho de

¹ Com $D = 1$, um termo de deriva não é identificável de forma não degenerada, de modo que `include.drift=FALSE` é a escolha consistente, e não uma omissão.

verossimilhança, e o $\text{SARIMA}(1,1,1)(1,0,1)_{12}$ vence por BIC, com vantagem de 3 pontos sobre o próprio campeão por AICc. Esta reversão é, em si, uma evidência adicional — já dentro do próprio capítulo de modelagem, antes de qualquer consideração preditiva — de que a noção de “melhor modelo” depende do critério adotado, tema que retorna com ainda mais força no Capítulo 8. Note-se que a reversão AICc/BIC ocorre inteiramente dentro da família $D = 0$ (ambos os modelos envolvidos, campeão e vice-campeão, usam diferenciação sazonal parcial com deriva) — os quatro candidatos com $D = 0$ dominam integralmente o ranking por ambos os critérios, muito à frente dos candidatos com $D = 1$ e do próprio `auto.arima()` exaustivo.

Diante do exposto, este relatório mantém o $\text{SARIMA}(8,1,1)(1,0,1)_{12}$ como modelo de referência da família SARIMA (por sua margem sobre o segundo colocado, embora pequena, ainda favorecê-lo sob o critério originalmente adotado, AICc), mas ambos — o campeão por AICc e o campeão por BIC, $\text{SARIMA}(1,1,1)(1,0,1)_{12}$ — são levados à validação preditiva do Capítulo 8, de modo que a escolha de especificação não dependa de uma decisão tomada unicamente dentro da amostra de treino. Vale notar que os quatro primeiros colocados do ranking, tanto por AICc quanto por BIC, são justamente os modelos SARIMA manuais com $D = 0$. O modelo do `auto.arima()` exaustivo — que também assume $D = 0$, mas com busca automatizada na parte ARMA — ocupa a última posição entre as oito opções (Tabela 5.1). Isso evidencia que a superioridade dos quatro primeiros colocados decorre da especificação manual da parte ARMA, e não estritamente da escolha de $D = 0$ em si.

Figura 5.3: FAC e FACP dos resíduos do modelo $\text{SARIMA}(1,1,1)(1,0,1)_{12}$ (campeão por BIC).



Fonte: Elaborado pelo Autor (2026).

O diagnóstico residual do candidato por BIC (Figura 5.3) é, em termos agregados, satisfatório: o teste de Ljung-Box não rejeita a ausência de autocorrelação ($\text{SARIMA}(1,1,1)(1,0,1)_{12}$):

$\chi^2 = 12,51$, $df = 19$, $p = 0,863$). Um exame mais granular da FAC e da FACP individuais desses mesmos resíduos, no entanto, revela uma ultrapassagem pontual e leve da banda de referência assintótica exatamente na defasagem 8 — pequena o suficiente para não comprometer o resultado agregado do teste conjunto, mas que fornece uma segunda linha de evidência, agora a partir do diagnóstico de resíduos do candidato mais parcimonioso, e não apenas da FACP da série diferenciada já discutida na Seção 5.1, para a inclusão do termo ϕ_8 no candidato SARIMA(8,1,1)(1,0,1)₁₂: a estrutura de dependência que o modelo mais simples deixa marginalmente descoberta na defasagem 8 é precisamente a que o candidato por AICc absorve ao preservar ϕ_8 como parâmetro livre, em vez de fixá-lo em zero junto com $\phi_2, \dots, \phi_7$. Isso reforça que a escolha entre os dois candidatos, nesta etapa, é genuinamente indeterminada pelos critérios *in-sample* — ainda que essa ultrapassagem residual pontual do candidato por BIC ofereça um argumento técnico adicional, embora modesto, a favor da especificação mais complexa; a evidência decisiva quanto ao desempenho preditivo é detalhada no Capítulo 8.

A Tabela 5.2 apresenta os coeficientes estimados para o modelo de referência SARIMA(8,1,1)(1,0,1)₁₂. Três aspectos numéricos dessa estimação merecem destaque. Primeiro, o termo de deriva, embora estruturalmente necessário pela especificação $D = 0$ (Seção 2.2), não se mostrou estatisticamente significativo isoladamente ($p = 0,375$); seu papel na tendência de extrapolação de longo prazo é discutido no Capítulo 8. Segundo, o coeficiente sazonal autorregressivo estimado ($\Phi_1 = 0,9934$) encontra-se criticamente próximo do limite teórico de não estacionariedade ($\Phi_1 = 1$) — uma assinatura numérica que reflete a sutil fronteira entre a escolha por $D = 0$ com forte persistência e a imposição de uma raiz unitária sazonal via $D = 1$, já discutida no Capítulo 4. Terceiro, o próprio ϕ_8 não atinge significância convencional isoladamente ($p = 0,082$) — um resultado que, sem contexto adicional, poderia sugerir um parâmetro superespecificado; a evidência residual do candidato por BIC discutida a seguir, no entanto, mostra que sua inclusão responde a uma estrutura de dependência genuína e localizada na defasagem 8, e não a um ajuste espúrio à amostra de treino.

Tabela 5.2: Coeficientes estimados do modelo SARIMA(8,1,1)(1,0,1)₁₂.

Parâmetro	Estimativa	Erro Padrão	Valor-z	Valor-p
ϕ_1	0,4636	0,1236	3,752	< 0,001 ***
ϕ_8	-0,0952	0,0548	-1,738	0,082
θ_1	-0,7213	0,0988	-7,300	< 0,001 ***
Φ_1	0,9934	0,0049	203,180	< 0,001 ***
Θ_1	-0,8605	0,0478	-18,015	< 0,001 ***
Deriva	0,0047	0,0053	0,888	0,375

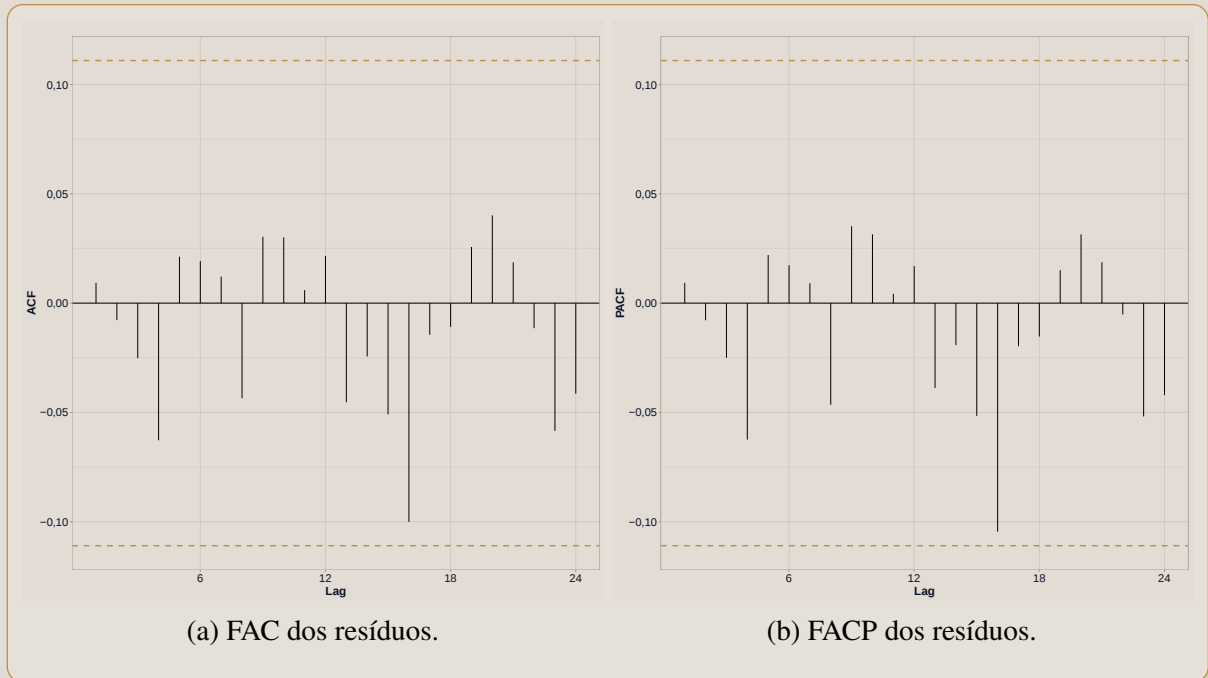
Nota: Método de estimação: Máxima Verossimilhança (MV); $\hat{\sigma}_a^2 = 0,001219$; Log-verossimilhança = 594,13.

Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.

Fonte: Elaborado pelo Autor (2026).

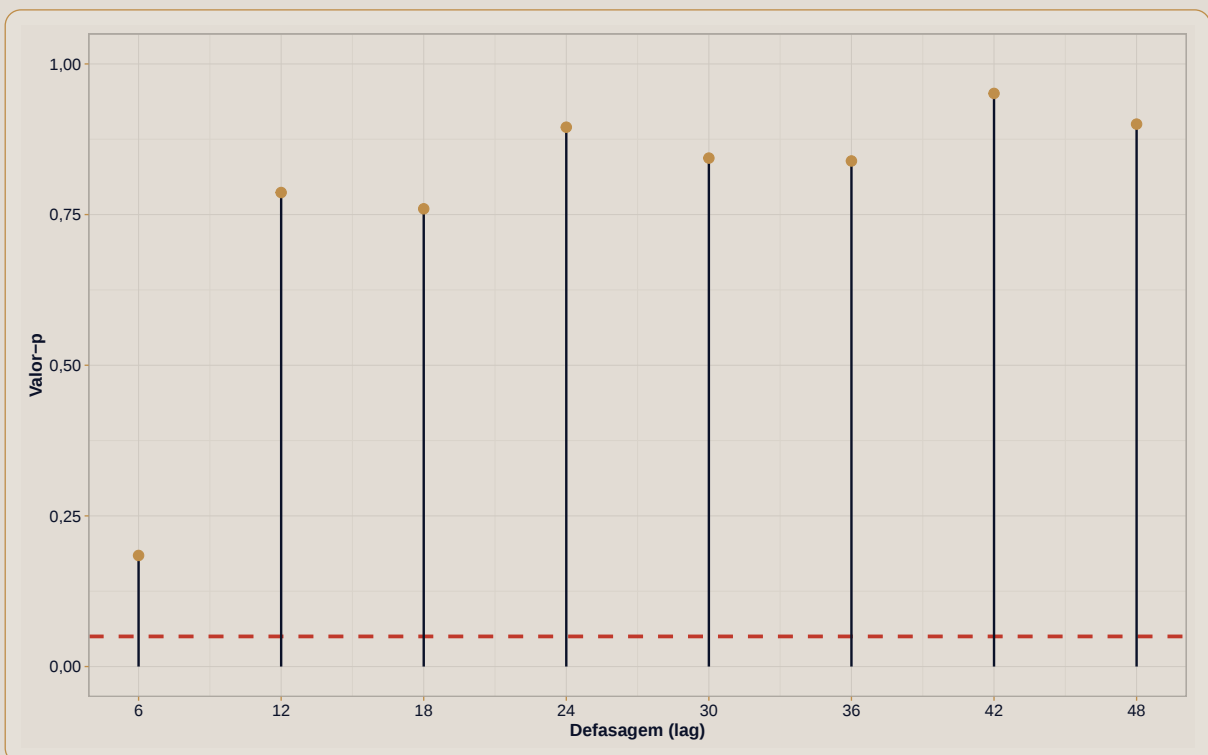
5.3 Diagnóstico de Resíduos do Modelo Campeão

Figura 5.4: FAC e FACP dos resíduos do modelo SARIMA(8,1,1)(1,0,1)₁₂.



Fonte: Elaborado pelo Autor (2026).

Figura 5.5: Valores- p do teste de Ljung-Box dos resíduos — modelo SARIMA(8,1,1)(1,0,1)₁₂. Defasagens de 6 a 48, com linha de referência em $\alpha = 0,05$.



Fonte: Elaborado pelo Autor (2026).

Tabela 5.3: Testes de diagnóstico dos resíduos do modelo SARIMA(8,1,1)(1,0,1)₁₂.

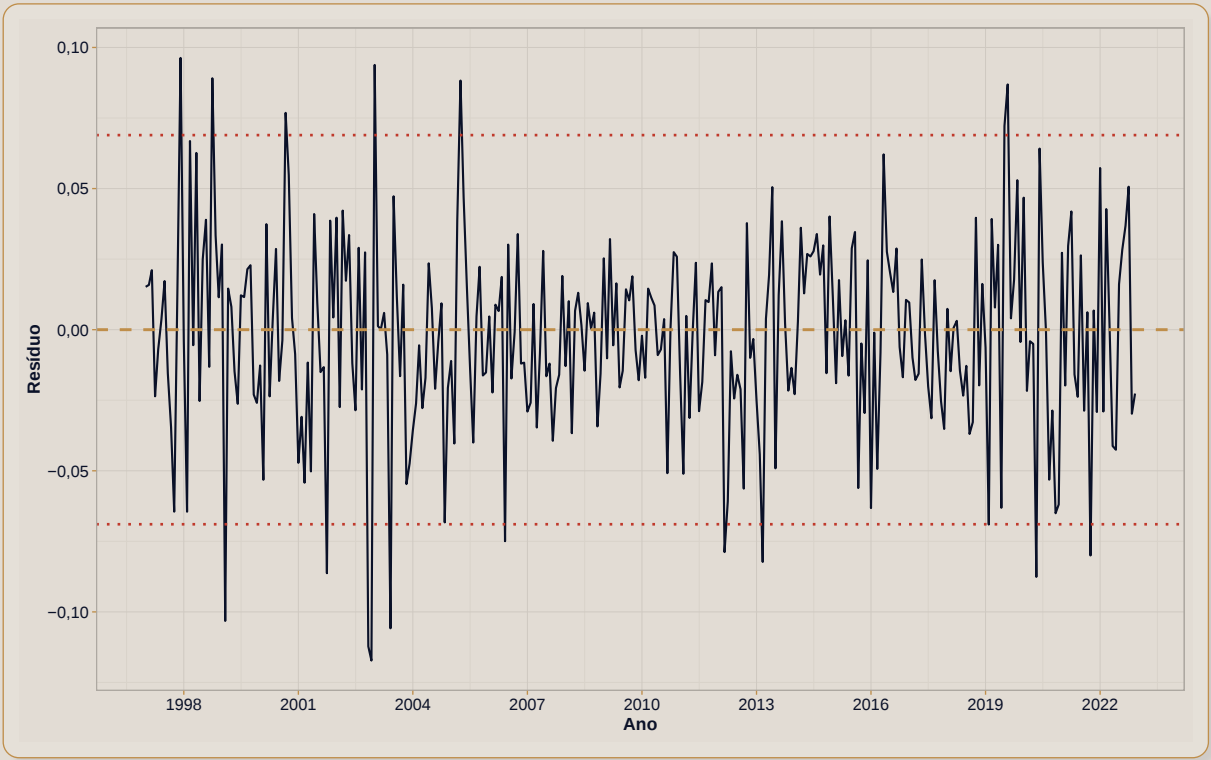
Teste	Hipótese nula (H_0)	Estatística	Valor- p
Ljung–Box (lag = 24)	Resíduos não autocorrelacionados	$\chi^2 = 10,98$ ($df = 18$)	0,895
Shapiro–Wilk	Resíduos seguem distribuição normal	$W = 0,985$	0,003***
Jarque–Bera	Resíduos seguem distribuição normal	$\chi^2 = 13,07$ ($df = 2$)	0,001***

Nota: Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.

Fonte: Elaborado pelo Autor (2026).

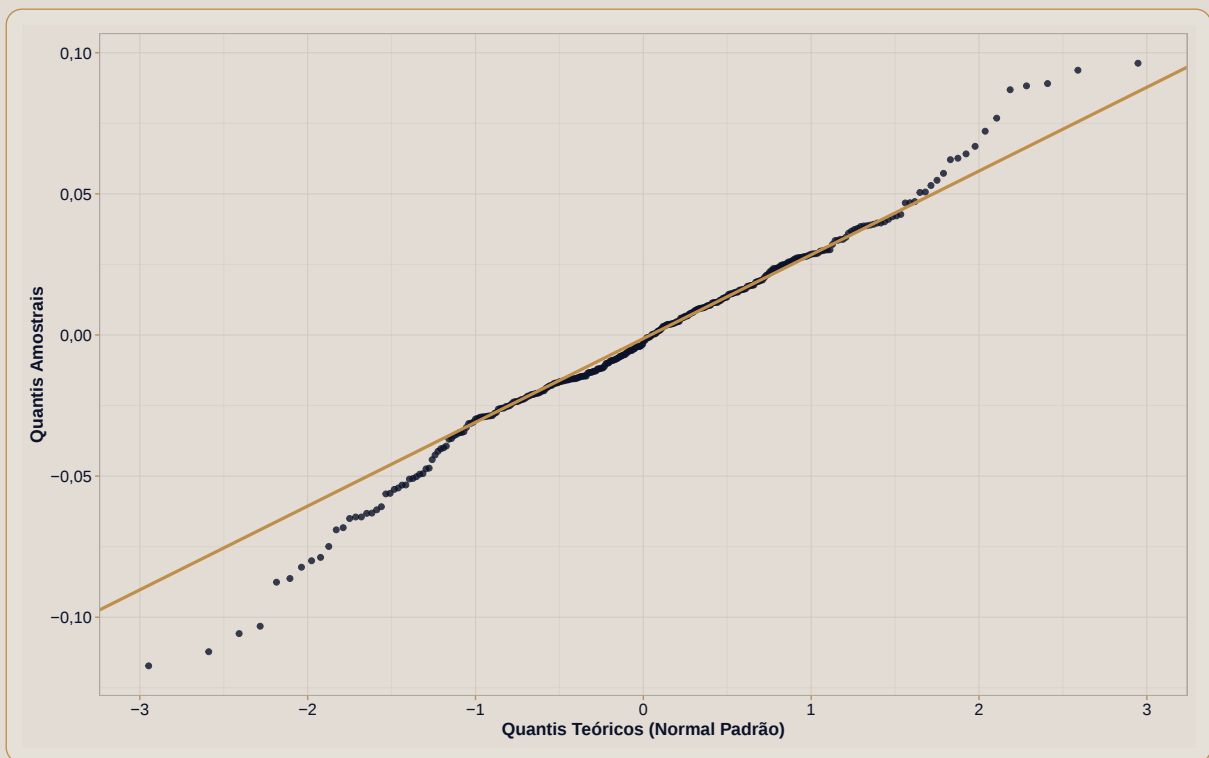
A ausência de autocorrelação residual (Ljung-Box, $p = 0,895$) é a condição central exigida pela metodologia de Box-Jenkins e é satisfeita: o modelo captura adequadamente a estrutura de dependência linear da série. A rejeição de normalidade (Shapiro-Wilk e Jarque-Bera, ambos $p < 0,01$) indica caudas mais pesadas que a normal nos resíduos — usual em séries de produção sujeitas a choques discretos (entrada/saída de unidades de produção) — e tem como principal implicação prática que os intervalos de previsão baseados na normalidade assintótica (Capítulo 8) devem ser interpretados com cautela adicional, sobretudo nas caudas.

Figura 5.6: Resíduos do modelo campeão ao longo do tempo.



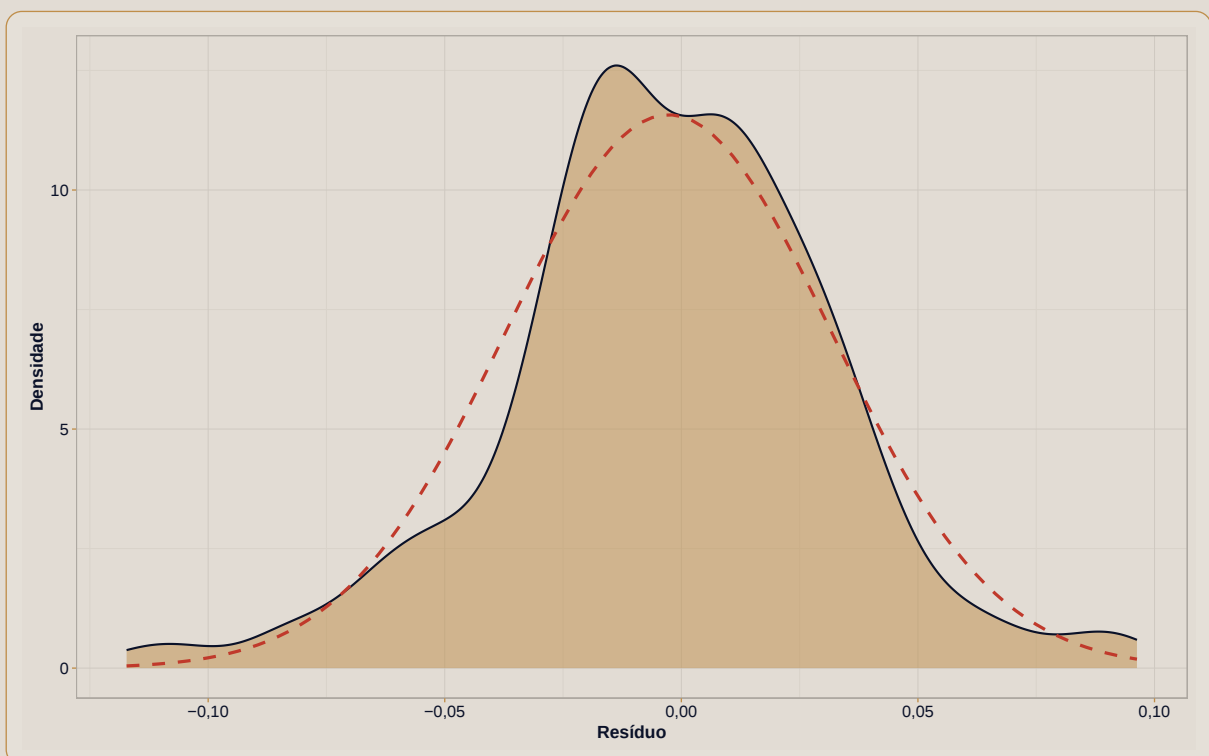
Fonte: Elaborado pelo Autor (2026).

Figura 5.7: Gráfico Q-Q dos resíduos do modelo campeão.



Fonte: Elaborado pelo Autor (2026).

Figura 5.8: Densidade empírica dos resíduos do modelo campeão.



Fonte: Elaborado pelo Autor (2026).

5.4 Heterocedasticidade Condicional dos Resíduos (ARCH)

O teste de Ljung-Box foi aplicado aos resíduos *ao quadrado* do modelo campeão, como *proxy* operacional do teste ARCH-LM (ENGLE, 1982).

Tabela 5.4: Teste de heterocedasticidade condicional (ARCH) nos resíduos do modelo SARIMA(8,1,1)(1,0,1)₁₂.

Defasagem	Estatística	Valor- <i>p</i>
6	27,46	< 0,001***
12	33,60	< 0,001***
24	40,56	0,0186*

Nota: Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.
Fonte: Elaborado pelo Autor (2026).

O resultado é inequívoco: há heterocedasticidade condicional fortemente significativa nos resíduos do modelo campeão em todas as defasagens testadas. Isso significa que, embora a variância *incondicional* dos resíduos seja adequadamente estabilizada pela transformação logarítmica (Capítulo 4), a variância *condicional* — a magnitude esperada do erro de previsão dado o passado recente — não é constante: os erros de previsão tendem a se concentrar em períodos de alta e de baixa volatilidade, em vez de se distribuírem uniformemente no tempo. Esta é uma camada de evidência adicional, e quantitativa, para a mesma cautela que a rejeição de normalidade (Tabela 5.3) já indicava qualitativamente: os intervalos de previsão assintóticos baseados em variância constante (Capítulo 8) tendem a ser mal calibrados, e essa expectativa é diretamente verificada por cobertura empírica naquele capítulo.

É importante enfatizar — e este ponto é retomado extensamente no Capítulo 8 — que “resíduos satisfatórios” segundo os critérios desta seção (Ljung-Box em nível) atestam apenas que o modelo capturou adequadamente a estrutura de autocorrelação *em média* dentro da amostra de treino. Isso não implica, e não deve ser lido como implicando, qualquer garantia sobre o desempenho preditivo fora da amostra, que depende também da estabilidade da relação estimada no futuro — precisamente o que os testes de quebra estrutural do Capítulo 6 colocam em questão, e o que a heterocedasticidade condicional aqui documentada já antecipa em relação à confiabilidade dos intervalos de previsão.

CAPÍTULO 6

ANÁLISE DE DADOS ATÍPICOS, INTERVENÇÃO E QUEBRA ESTRUTURAL

6.1 Detecção Automática de Outliers

A função `tsoutliers::tso()` foi aplicada ao logaritmo da série restrita ao conjunto de treino, utilizando como modelo subjacente um ARIMA selecionado automaticamente (`tsmethod = "auto.arima"`), com busca conjunta dos quatro tipos de dados atípicos (AO, LS, TC, IO) e até 15 iterações do laço externo de detecção.

A chamada combinada de `tso()` com múltiplos tipos e mais de uma iteração invoca, em seu laço interno, uma otimização numérica por diferenças finitas (BFGS) que pode falhar ("`non-finite finite-difference value`") dependendo da versão exata do R, do pacote `forecast` e das bibliotecas BLAS/LAPACK do sistema operacional — com o mesmo código-fonte. O *script* implementa um mecanismo de novas tentativas que substitui o otimizador por um método livre de gradiente (Nelder-Mead) caso o padrão (BFGS) falhe. A troca de otimizador não apenas evita a falha, mas pode alterar qual conjunto de *outliers* é identificado, porque o próprio modelo ARIMA de base selecionado automaticamente a cada iteração já diverge sob caminhos de otimização distintos. Os resultados a seguir são os obtidos, de forma reprodutível, no ambiente de execução R 4.4.2 (ver Apêndice A) e podem diferir marginalmente de uma reexecução em outro ambiente — uma limitação de reprodutibilidade do método em si, não deste relatório, e que este trabalho torna explícito em vez de ocultar.

Tabela 6.1: *Outliers* detectados na série de produção de petróleo (escala logarítmica, treino).

Tipo	Data	Coefficiente	Estatística t
TC	out/1997	-0,143	-4,59
AO	out/2001	-0,108	-4,60
AO	dez/2002	-0,116	-4,97
AO	jun/2003	-0,105	-4,53
LS	abr/2005	0,099	3,79
LS	ago/2019	0,197	4,50

Nota: AO = Additive Outlier; TC = Transient Change; LS = Level Shift. Modelo de base selecionado automaticamente — ARIMA(1,1,3)(2,0,0)₁₂.

Fonte: Elaborado pelo Autor (2026).

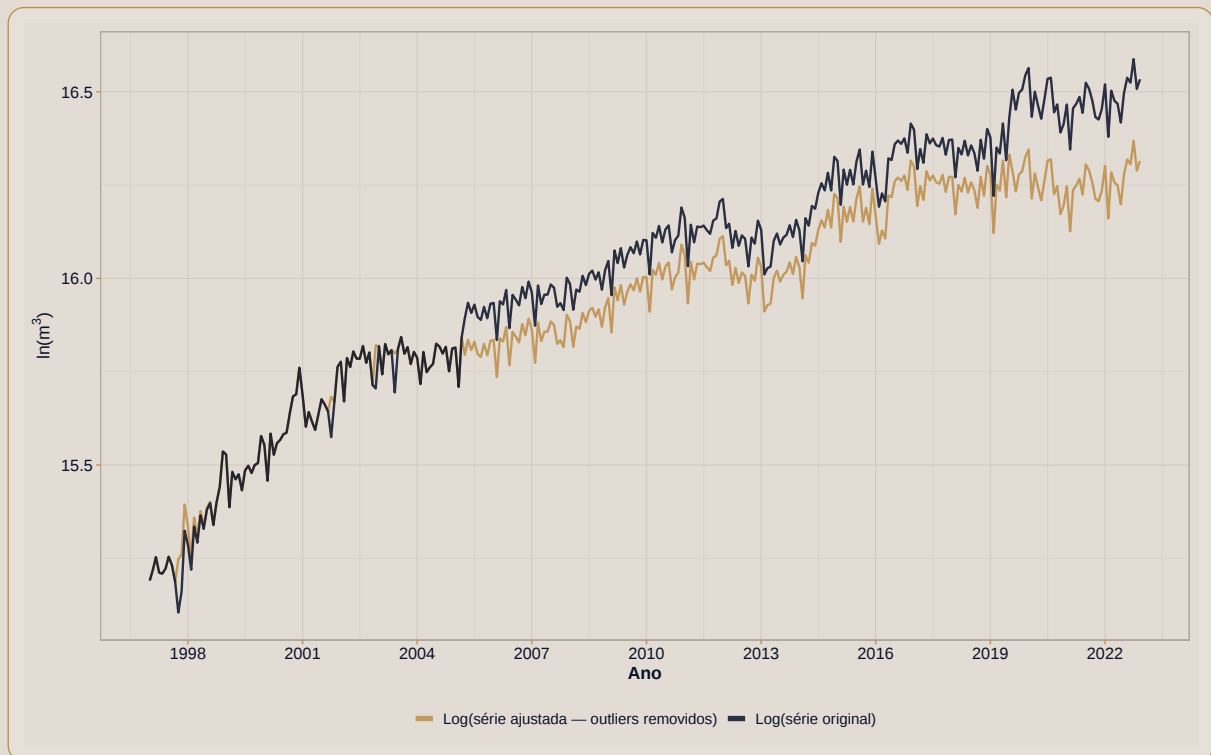
Figura 6.1: Outliers identificados automaticamente pelo procedimento `tsO()`.



Fonte: Elaborado pelo Autor (2026).

Figura 6.2: Efeito da remoção dos outliers detectados.

Linha escura: série original (em Log) | Linha dourada: série ajustada sem *outliers* (em Log).



Fonte: Elaborado pelo Autor (2026).

A Figura 6.2 sobrepõe a série em log original e a série ajustada pela remoção dos efeitos estimados dos seis outliers da Tabela 6.1 (`deteccao_outliers$yadj`), tornando visualmente explícito o peso de cada intervenção sobre o nível da série.

6.2 Construção dos Regressores de Intervenção

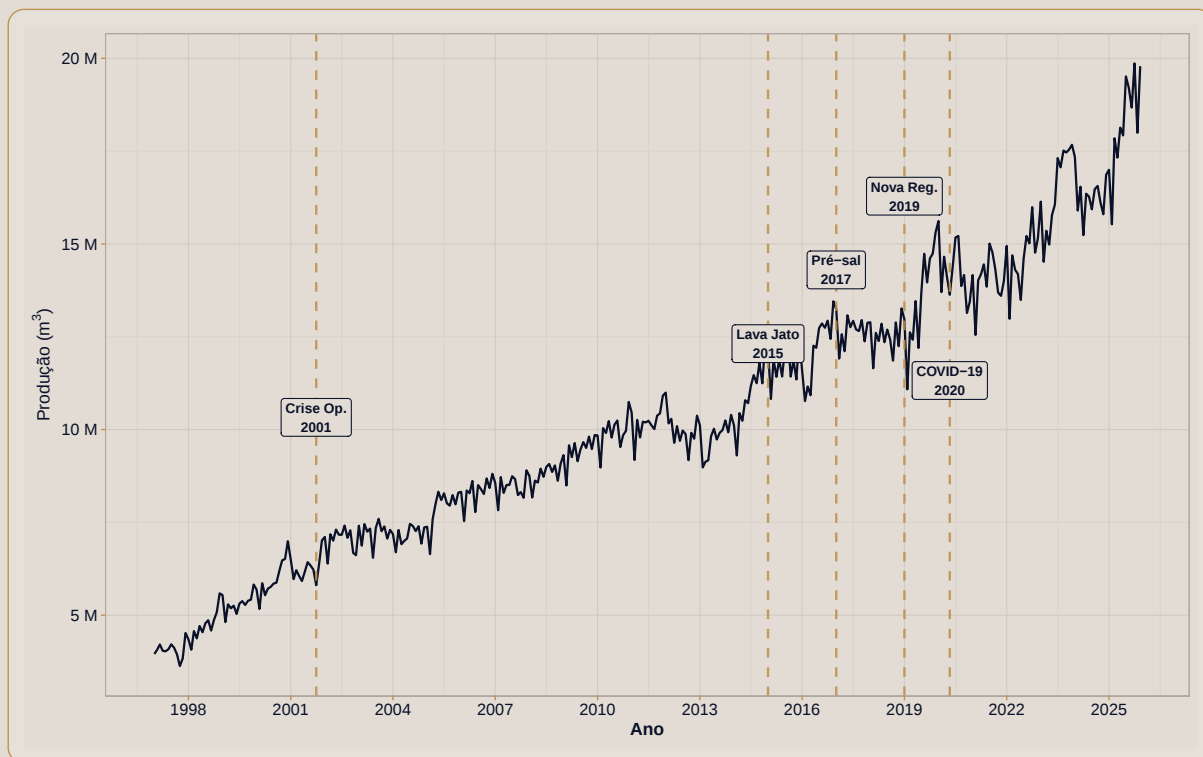
Com base no conhecimento de domínio sobre a história do setor petrolífero brasileiro, cinco variáveis de intervenção manuais foram construídas:

- **AO (choque de 2001):** pulso aditivo para a crise do apagão energético e instabilidade operacional de 2001.
- **LS (Operação Lava Jato):** mudança de nível a partir de 2015, associada à retração de investimentos durante a Operação Lava Jato.
- **LS (início do Pré-Sal):** mudança de nível a partir da entrada em produção comercial dos grandes campos do pré-sal.
- **LS (marco regulatório):** mudança de nível associada a alterações no marco regulatório do setor.

- **AO (pandemia de COVID-19):** pulso aditivo para o choque da pandemia de COVID-19 (2020).

Figura 6.3: Pontos de intervenção manual sobre a série observada de produção de petróleo.

Linhas tracejadas douradas: datas dos cinco eventos históricos modelados nesta seção.



Fonte: Elaborado pelo Autor (2026).

Note-se que o *outlier* de nível (LS) identificado de forma inteiramente automática pelo procedimento $\text{tsO}()$ em agosto de 2019 (Tabela 6.1) recai muito próximo da data historicamente atribuída às alterações do marco regulatório de 2019 modeladas manualmente nesta seção — uma convergência entre detecção estatística não supervisionada e conhecimento de domínio que reforça a plausibilidade de ambas as fontes de evidência, de forma independente uma da outra.

Essas cinco variáveis manuais são complementadas por variáveis de quebra estrutural estimadas estatisticamente pelo procedimento de Bai-Perron — tratado em detalhe na seção seguinte.

6.3 Testes Formais de Quebra Estrutural

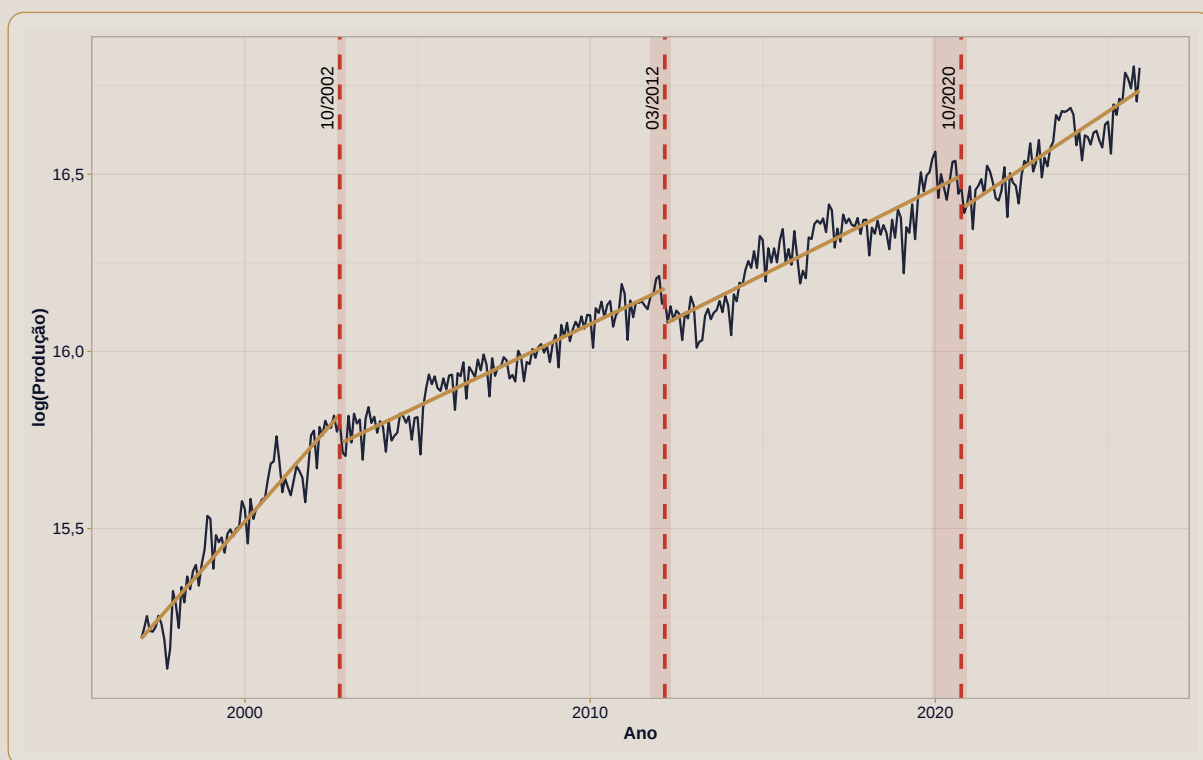
O teste OLS-CUSUM de Brown, Durbin e Evans (1975), aplicado à regressão de $\ln(Y_t)$ contra uma tendência linear determinística, rejeita fortemente a hipótese de constância de parâmetros ($p \approx 0$), confirmando instabilidade temporal na relação tendência-tempo ao longo da amostra completa.

6.3.1 Caracterização Descritiva: Bai-Perron sobre o treino

O procedimento de Bai-Perron (BAI; PERRON, 2003) foi aplicado sobre $\ln(Y_t)$ de duas formas complementares: (i) de forma descritiva, sobre a série completa (jan/1997–dez/2025); e (ii) de forma restrita ao treino (jan/1997–dez/2022, $n = 312$), esta última utilizada para a construção, sem vazamento de informação futura, dos regressores de intervenção estatísticos empregados na Seção 6.5.

Figura 6.4: Série $\ln(Y_t)$ com três quebras estruturais estimadas pelo método de Bai–Perron (série completa).

Linhas tracejadas vermelhas: estimativas das datas de quebra (IC de 95%); linha dourada: tendências lineares ajustadas em cada regime.



Fonte: Elaborado pelo Autor (2026).

Sobre a série completa, o procedimento identifica quebras em outubro de 2002, março de 2012 e outubro de 2020 (Figura 6.4). Já quando o mesmo procedimento é restrito exclusivamente ao treino — a especificação relevante para os regressores construídos sem vazamento de informação futura —, as datas estimadas deslocam-se para janeiro de 2001, fevereiro de 2010 e abril de 2014 (Tabela 6.2):

Tabela 6.2: Datas de quebra estrutural identificadas pelo procedimento de Bai-Perron restrito ao conjunto de treino (jan/1997–dez/2022).

Quebra	Data estimada
1	Janeiro de 2001
2	Fevereiro de 2010
3	Abril de 2014

Fonte: Elaborado pelo Autor (2026).

Essa sensibilidade das datas estimadas à janela de estimação — um deslocamento de um a oito anos, dependendo do ponto de quebra, entre a estimativa na série completa e a estimativa restrita ao treino — não é uma inconsistência do método, mas seu comportamento esperado: o algoritmo de Bai-Perron busca partições ótimas dentro da janela de dados fornecida, e a inclusão ou exclusão do período mais recente e mais instável da série (2023–2025, Capítulo 3) desloca a localização ótima das quebras nos regimes anteriores. Esta é, precisamente, a razão metodológica pela qual apenas as datas restritas ao treino — e não as datas obtidas com a série completa — são utilizadas para construir os regressores de intervenção estatísticos empregados na modelagem preditiva (Seção 6.5): usar as datas da série completa introduziria vazamento de informação do próprio conjunto de teste na construção dos regressores.

As três quebras restritas ao treino têm correspondência histórica plausível, embora não estritamente simultânea, com eventos documentados do setor: janeiro de 2001 antecede em alguns meses o início formal do racionamento de energia elétrica (1º de junho de 2001) que caracterizou a crise do apagão (ALMEIDA et al., 2022) — uma antecipação consistente com o fato de que os sinais de crise hídrica subjacente já eram de conhecimento público desde o final de 2000, e que o AO_choque2001 manual, datado de outubro daquele mesmo ano, captura separadamente o período em que o racionamento formal (junho de 2001 a fevereiro de 2002) já estava plenamente em curso; abril de 2014 corresponde ao início público da Operação Lava Jato (a operação foi deflagrada em março de 2014, com desdobramentos sobre a Petrobras emergindo nos meses seguintes); fevereiro de 2010 situa-se no período de desenvolvimento inicial intensivo do pré-sal. As quebras identificadas na série completa (2002, 2012, 2020) são, por sua vez, também historicamente plausíveis — 2002 próxima ao fim do ciclo identificado pelo AO_choque2001, 2012 no auge do desenvolvimento pré-sal, e 2020 coincidente com o choque da pandemia de COVID-19 — e reforçam, de forma independente, a conclusão central desta seção: a série exibe instabilidade estrutural genuína e estatisticamente confirmada (CUSUM, e Bai-Perron sob ambas as janelas de estimação).

6.4 Diagnóstico de Multicolinearidade

Como múltiplos regressores do tipo degrau (LS) são incluídos simultaneamente — cinco manuais mais até três quebras estruturais —, o fator de inflação de variância (VIF, Seção 2.3) foi calculado para cada especificação.

Tabela 6.3: Fator de Inflação de Variância (VIF) dos regressores de intervenção.

Especificação	Regressor	VIF
5 manuais	AO_choque2001	1,00
	LS_lava_jato	3,08
	LS_presal	4,62
	LS_regulatorio	2,56
	AO_covid	1,02
5 manuais + 3 quebras (treino)	AO_choque2001	1,01
	LS_lava_jato	10,15
	LS_presal	4,62
	LS_regulatorio	2,56
	AO_covid	1,02
	LS_quebra1_bp (jan/2001)	1,22
	LS_quebra2_bp (fev/2010)	2,28
	LS_quebra3_bp (abr/2014)	9,13

Fonte: Elaborado pelo Autor (2026).

No modelo de 5 regressores manuais, a colinearidade é moderada (LS_presal, VIF = 4,62, é o caso mais próximo do limiar convencional de atenção, $VIF > 5$). Ao incluir as três quebras estruturais estimadas no treino, dois regressores cruzam o limiar de colinearidade severa ($VIF > 10$ ou muito próximo dele): LS_lava_jato (VIF = 10,15) e LS_quebra3_bp (VIF = 9,13, a quebra de abril/2014) — ambos situados a poucos meses um do outro no calendário, o que é exatamente o tipo de proximidade temporal que infla a colinearidade entre regressores do tipo degrau. Este diagnóstico é um argumento adicional — ao lado da comparação de AICc e MAPE da seção seguinte — para preferir a especificação mais parcimoniosa.

6.5 Os Três Modelos ARIMAX

Tabela 6.4: Comparação por AICc: modelo sem intervenção vs. os três candidatos ARIMAX (treino).

Modelo	AICc
Sem intervenção (campeão SARIMA, Capítulo 5)	−1.174
ARIMAX (5 manuais)	−1.179
ARIMAX (5 manuais + quebras estruturais)	−1.178
ARIMAX (parcimonioso)	−1.188

Fonte: Elaborado pelo Autor (2026).

Todas as três especificações de intervenção melhoram o AICc em relação ao modelo sem intervenção — evidência de que os regressores capturam variação sistemática não explicada pela componente ARMA pura. Entre os dois candidatos com conjunto completo de regressores, o modelo de 5 regressores manuais tem o melhor AICc (−1.179 contra −1.178): mais simples e, ainda assim, com ajuste levemente superior ao modelo de 8 regressores. Isso já é uma primeira indicação de que as três quebras estruturais (Tabela 6.2) agregam pouco poder explicativo adicional ao conjunto de 5 manuais.

Essa indicação é confirmada de forma mais direta por um terceiro candidato, construído a partir dos regressores individualmente significativos ($p < 0,05$) na especificação de 8 regressores: AO_choque2001, AO_covid e LS_quebra1_bp (a única das três quebras estruturais estatisticamente significativa — ver discussão abaixo). Esse ARIMAX parcimonioso, com apenas três regressores, obtém o melhor AICc entre todos os candidatos de intervenção, e por margem considerável: 9 pontos à frente do segundo colocado (5 manuais) e 10 pontos à frente do modelo de 8 regressores. Este resultado é a confirmação mais direta, dentro da própria amostra de treino, do padrão que a análise de VIF (Seção anterior) já sugeria: a maior parte do poder explicativo dos regressores de intervenção está concentrada em três eventos específicos, e a adição dos demais regressores (LS_lava_jato, LS_presal, LS_regulatorio, LS_quebra2_bp, LS_quebra3_bp) apenas acrescenta parâmetros sem ganho de verossimilhança suficiente para compensar sua penalização no AICc.

Tabela 6.5: Efeitos estimados — ARIMAX (5 manuais).

Variável	Efeito estimado	Valor-z	Valor-p
AO (choque de 2001)	−9,62%	−3,95	0,0001***
LS (Operação Lava Jato)	−0,57%	−0,18	0,854
LS (início do Pré-Sal)	−0,19%	−0,06	0,953
LS (marco regulatório)	−2,08%	−0,66	0,507
AO (pandemia de COVID-19)	−7,80%	−3,17	0,0015**

Nota: Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.

Fonte: Elaborado pelo Autor (2026).

Tabela 6.6: Efeitos estimados — ARIMAX (5 manuais + 3 quebras estruturais sem vazamento).

Variável	Origem	Efeito estimado	Valor-z	Valor-p
AO (choque de 2001)	Manual	−9,40%	−3,89	0,0001***
LS (Lava Jato)	Manual	−0,69%	−0,23	0,821
LS (início do Pré-Sal)	Manual	−0,30%	−0,10	0,922
LS (marco regulatório)	Manual	−2,08%	−0,67	0,504
AO (COVID-19)	Manual	−7,86%	−3,22	0,0013**
LS quebra 1 (jan/2001)	Bai-Perron (treino)	−6,39%	−2,12	0,0337*
LS quebra 2 (fev/2010)	Bai-Perron (treino)	0,42%	0,14	0,888
LS quebra 3 (abr/2014)	Bai-Perron (treino)	2,91%	0,94	0,348

Nota: Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.

Fonte: Elaborado pelo Autor (2026).

Tabela 6.7: Efeitos estimados — ARIMAX parcimonioso (3 regressores).

Variável	Efeito estimado	Valor-z	Valor-p
AO (choque de 2001)	−9,40%	−3,89	0,0001***
AO (pandemia de COVID-19)	−7,84%	−3,21	0,0013**
LS quebra 1 (jan/2001, Bai-Perron)	−6,39%	−2,11	0,0345*

Nota: Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.

Fonte: Elaborado pelo Autor (2026).

Das três quebras estruturais de Bai-Perron, apenas a primeira (janeiro de 2001) é estatisticamente significativa ($p = 0,034$), e mesmo essa, marginalmente; as outras duas não atingem significância convencional. Combinada à comparação de AICc da Tabela 6.4 e ao diagnóstico de colinearidade severa da seção anterior — precisamente nos dois regressores mais próximos temporalmente (LS_lava_jato e a quebra de abril/2014) —, a evidência converge de forma consistente: as quebras estruturais de fevereiro/2010 e abril/2014 não agregam valor explicativo robusto sobre a especificação de 5 regressores manuais, e apenas a quebra de janeiro/2001 carrega sinal estatístico independente — coerentemente reincorporada no modelo parcimonioso. Os efeitos manuais, por sua vez, são qualitativamente estáveis entre as três especificações (AO de 2001 e AO de COVID-19 seguem sendo os únicos efeitos manuais individualmente significativos, com magnitudes muito próximas nas três tabelas) — o que reforça a robustez desses efeitos específicos, em contraste com a fragilidade da maioria dos regressores de quebra.

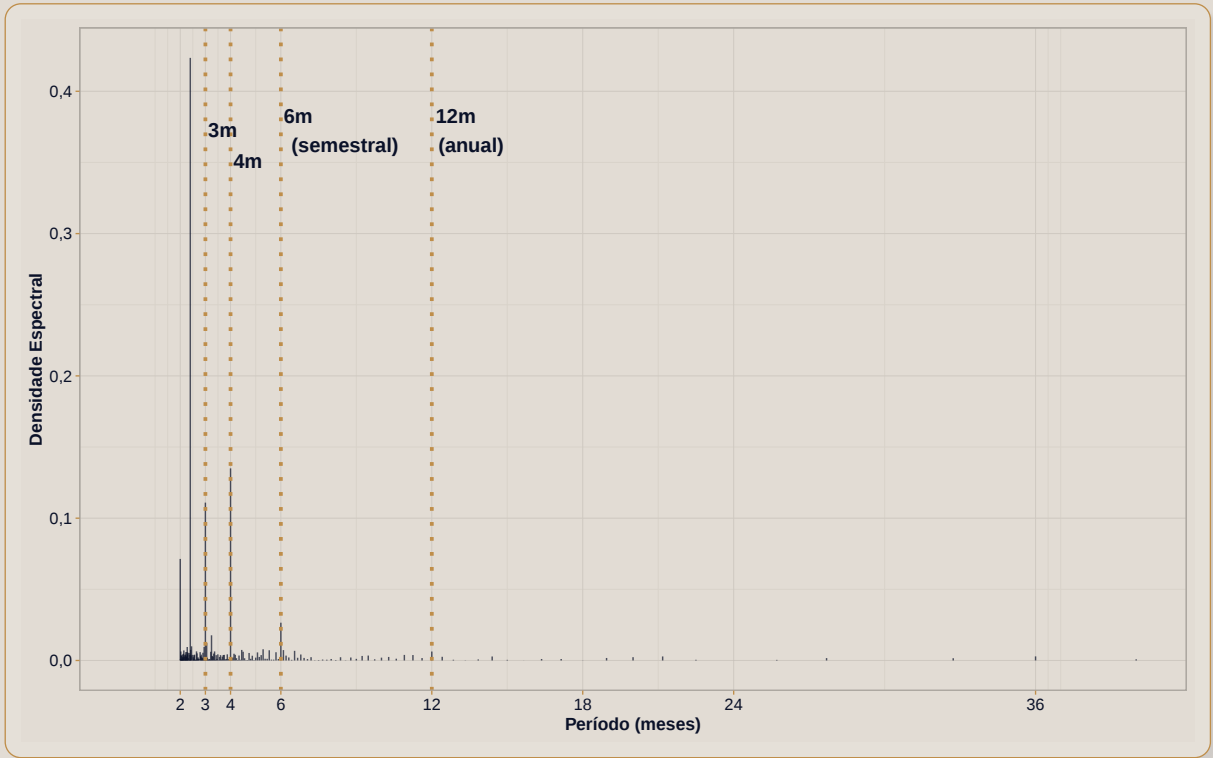
Os três modelos ARIMAX são levados ao Capítulo 8 para avaliação *out-of-sample*, permitindo que a evidência preditiva, e não apenas o AICc dentro da amostra, informe qual especificação é preferível na prática.

ANÁLISE ESPECTRAL

7.1 Periodograma

O periodograma bruto da série $\Delta \ln(Y_t)$ (Figura 7.1) decompõe a variância da série em contribuições associadas a cada período de oscilação. A Tabela 7.1 lista os cinco períodos de maior densidade espectral.

Figura 7.1: Periodograma bruto da série $\Delta \ln(Y_t)$, em função do período (meses).



Fonte: Elaborado pelo Autor (2026).

Tabela 7.1: Cinco maiores picos de densidade espectral do periodograma bruto.

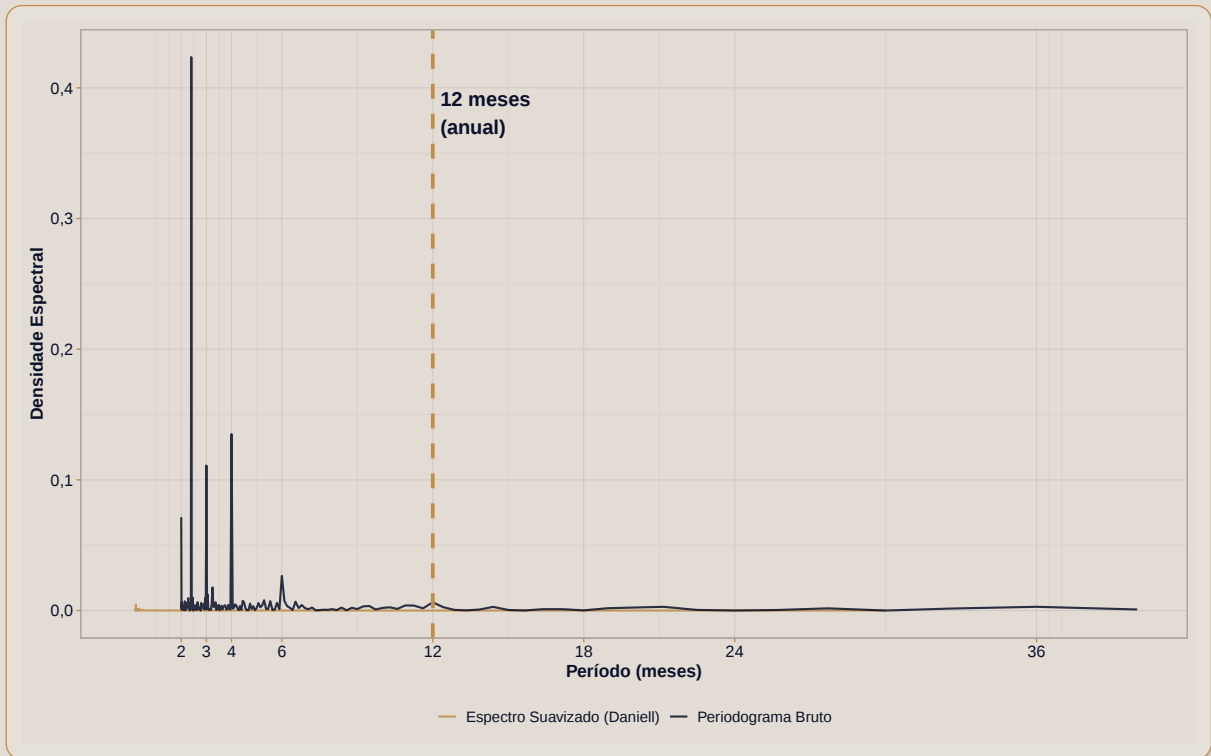
Posição	Período (meses)	Interpretação em relação ao ciclo anual
1º	$\approx 2,4$	5º harmônico de 12 meses
2º	4,0	3º harmônico de 12 meses
3º	3,0	4º harmônico de 12 meses
4º	2,0	6º harmônico de 12 meses
5º	6,0	2º harmônico de 12 meses

Fonte: Elaborado pelo Autor (2026).

A leitura correta deste resultado não é que “não existe ciclo anual” — os períodos de 2,4, 3,0, 4,0 e 6,0 meses são, todos, submúltiplos exatos de 12 meses, isto é, harmônicos do ciclo anual ($12/5$, $12/4$, $12/3$ e $12/2$, respectivamente). O que esses resultados revelam é que a energia espectral do ciclo sazonal está distribuída por seus harmônicos, e não concentrada na frequência fundamental. Esse padrão é a assinatura espectral esperada de um ciclo sazonal fortemente não senoidal (assimétrico): uma função periódica perfeitamente senoidal concentraria praticamente toda sua energia na frequência fundamental; uma função com formato assimétrico — como o padrão documentado no Capítulo 3 — exige a soma de múltiplos harmônicos para ser reproduzida, e esses harmônicos competem com a fundamental pela energia espectral total.

O periodograma da Figura 7.1 é calculado sobre a série diferenciada $\Delta \ln(Y_t)$, não sobre $\ln(Y_t)$ diretamente. O filtro $(1 - B)$ tem ganho $|1 - e^{-i\omega}|^2 = 2(1 - \cos \omega)$, estritamente crescente com a frequência ω — isto é, diferenciar amplifica mecanicamente a energia aparente de períodos curtos (frequências altas) em relação a períodos longos (Seção 2.4). Isso significa que parte da concentração de energia nos harmônicos de período curto documentada na Tabela 7.1 pode refletir esse ganho do filtro, e não exclusivamente a assimetria genuína do ciclo sazonal. Esta ressalva não reverte a conclusão qualitativa (a sazonalidade desta série é, de fato, não senoidal — evidência convergente e independente disso vem da quantificação do efeito de calendário no Capítulo 3 e da própria forma da regressão de Fourier na Seção 7.3), mas indica que a magnitude exata da dispersão harmônica relatada nesta seção deve ser interpretada com essa fonte adicional de cautela.

Figura 7.2: Espectro suavizado (kernel de Daniell) versus periodograma bruto.

Linha dourada: Espectro suavizado com $\text{span} = (3, 5)$.

Fonte: Elaborado pelo Autor (2026).

7.2 Teste de Fisher para Significância Espectral

O teste de Fisher (FISHER, 1929) avalia formalmente se o maior pico do periodograma é estatisticamente distinguível de ruído branco, sob H_0 : ausência de componente periódica determinística. A estatística observada é $g = 0,3448$ ($p < 0,001$), rejeitando fortemente H_0 e confirmando a presença de pelo menos uma componente periódica determinística genuína na série — coerente com a evidência do teste QS (Seção 4.3) e da decomposição STL/clássica (Capítulo 3).

7.3 Regressão Harmônica de Fourier

Uma regressão harmônica de Fourier (Equação 2.7) foi ajustada como alternativa à diferenciação sazonal, com o número de pares harmônicos K selecionado por minimização do AICc dentre $K = 1, \dots, 6$.

Tabela 7.2: AICc por número de harmônicos K — seleção restrita ao treino, busca exaustiva.

K	AICc
1	−988,0
2	−997,6
3	−1.050,7
4	−1.099,0
5	−1.178,7
6	−1.205,4

Fonte: Elaborado pelo Autor (2026).

O K ótimo é $K = 6$, exatamente o limite de Nyquist para $s = 12$ (o número máximo de pares harmônicos admissível), e o AICc estava ainda estritamente decrescente nesse limite — isto é, a busca não encontrou um mínimo interior dentro da grade $K = 1, \dots, 6$ testada. Isso significa que a regressão de Fourier está convergindo para uma representação saturada da sazonalidade: em $K = 6$, o termo seno associado à frequência de Nyquist, $\sin(2\pi \cdot 6t/12) = \sin(\pi t)$, é identicamente nulo para qualquer t inteiro e não contribui como parâmetro livre, de modo que o modelo saturado retém exatamente μ (em média), 6 termos cosseno e 5 termos seno genuinamente não degenerados — 12 parâmetros ao todo, estatisticamente equivalente, em graus de liberdade, ao uso de $s - 1 = 11$ indicadores mensais (*dummies*) mais um intercepto. Esse resultado é, em si, uma confirmação adicional e independente da conclusão da Seção 3.2: a forma sazonal desta série é genuinamente complexa e não se resume a um único ciclo suave — embora, como discutido naquela seção, parte dessa complexidade aparente seja também atribuível ao efeito de calendário, e não apenas a uma dinâmica operacional intrinsecamente rica.

Figura 7.3: Ajuste da regressão harmônica de Fourier ($K = 6$) sobre a série completa.

Linha escura: série original | Linha dourada: regressão de Fourier.



Fonte: Elaborado pelo Autor (2026).

Quanto à sua relevância para o Capítulo 8: diferentemente da diferenciação sazonal ($D = 1$), que impõe uma raiz unitária sazonal e — combinada a $d = 1$ — torna não identificável qualquer termo de deriva (*drift*) (Seção 2.2), a regressão de Fourier modela a sazonalidade como puramente determinística, preservando a possibilidade de um termo de deriva linear não degenerado mesmo com $d = 1$. Esta propriedade, combinada à recomendação dos testes formais de D do Capítulo 4 (que apontam $D = 0$), motiva tratar esta especificação como concorrente de primeira classe da família SARIMA na avaliação preditiva do próximo capítulo.

PREVISÃO E AVALIAÇÃO DE MODELOS

8.1 Desenho de Avaliação

Os modelos candidatos — as oito especificações SARIMA do Capítulo 5 (sete manuais mais `auto.arima`), os três modelos com regressores de intervenção do Capítulo 6, ARIMA com regressores de Fourier (Capítulo 7), Holt-Winters, um modelo estrutural de espaço de estados (BSM, Harvey (1989)), TBATS, NNAR e Prophet — são avaliados sob dois desenhos de validação distintos: (i) um *holdout* estático único, com origem em dezembro de 2022 e horizonte de 36 meses (jan/2023–dez/2025); e (ii) uma validação *rolling-origin*, com sete origens sucessivas e horizonte fixo de 12 meses. Os dois desenhos são apresentados separadamente porque — como ficará evidente — eles não concordam sobre qual modelo é superior, e essa divergência é, por si só, um resultado substantivo deste estudo.

8.2 Resultado sob o Holdout Estático (36 meses)

Tabela 8.1: Acurácia preditiva no conjunto de teste (jan/2023–dez/2025), origem única, horizonte de 36 meses.

Modelo	MAE	RMSE	MAPE (%)	MASE	U-Theil	Elegível
Holt-Winters	830.657	1.008.788	4,742	1,350	1,260	Sim
<code>auto.arima</code> [ARIMA(2,1,1)(2,0,0) ₁₂ c/ deriva]	922.925	1.127.748	5,254	1,500	1,409	Sim
ARIMA + Fourier (parcimonioso)	1.000.498	1.229.305	5,684	1,626	1,535	Sim
ARIMAX (5 manuais)	1.082.475	1.334.219	6,121	1,759	1,666	Sim
ARIMAX (5 manuais + quebras estruturais)	1.092.171	1.350.086	6,170	1,774	1,686	Sim
SARIMA(2,1,4)(0,1,1) ₁₂	1.094.191	1.331.606	6,199	1,778	1,663	Sim
ARIMAX (parcimonioso)	1.121.984	1.381.857	6,340	1,823	1,726	Sim
SARIMA(0,1,1)(1,0,1) ₁₂	1.120.174	1.360.310	6,351	1,820	1,699	Sim
SARIMA(8,1,1)(1,0,1) ₁₂ (campeão AICc, Cap. 5)	1.156.409	1.413.640	6,542	1,879	1,766	Sim
SARIMA(1,1,1)(1,0,1) ₁₂ (campeão BIC, Cap. 5)	1.163.846	1.412.913	6,593	1,891	1,765	Sim
SARIMA(1,1,2)(1,0,1) ₁₂	1.165.889	1.415.060	6,605	1,894	1,767	Sim
SARIMA(8,1,1)(0,1,1) ₁₂	1.189.579	1.446.484	6,737	1,933	1,807	Sim
SARIMA(1,1,1)(0,1,1) ₁₂	1.198.065	1.447.100	6,794	1,947	1,807	Sim
Prophet	1.333.382	1.622.580	7,530	2,166	2,027	Sim
BSM	1.727.964	2.073.130	9,758	2,808	2,589	Sim
NNAR	1.972.473	2.316.409	11,143	3,205	2,893	Sim
TBATS	2.045.106	2.356.559	11,616	3,323	2,943	Sim
Naive Sazonal	2.376.508	2.651.313	13,646	3,861	3,312	Sim

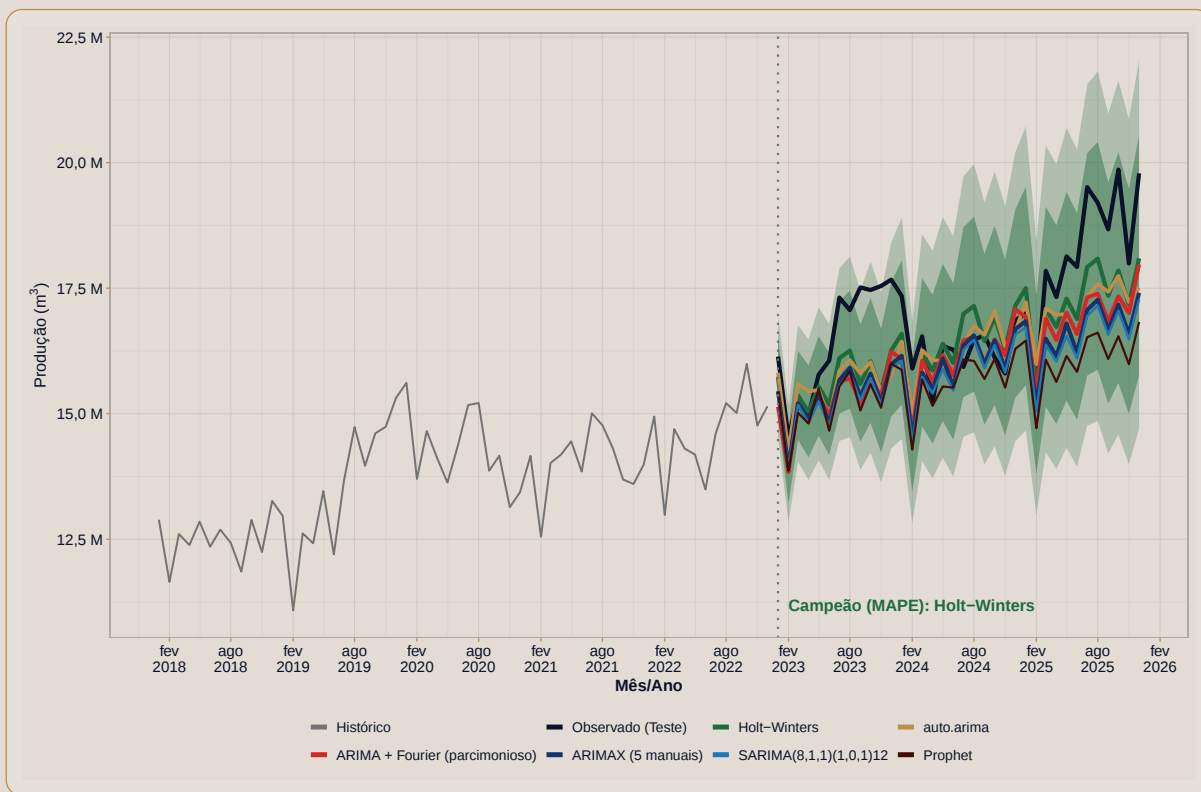
Fonte: Elaborado pelo Autor (2026).

A leitura é sóbria: a aceleração de produção observada em 2023–2025 foi suficientemente atípica em relação ao padrão histórico do treino para que nenhum dos dezoito modelos comparados batesse, em erro absoluto médio, o referencial de dificuldade histórica de previsão sazonal embutido no denominador do MASE — o mesmo previsor que simplesmente repete o valor do mesmo mês do ano anterior, calculado dentro do próprio treino (Capítulo 2, Equação 2.8).

É digno de nota que o modelo aqui tratado como referência da família SARIMA (Capítulo 5) — selecionado por vencer pelo critério AICc dentro da amostra de treinamento — é superado, fora da amostra, pelo único candidato SARIMA que perdeu essa mesma competição por AICc/BIC: o `auto.arima()` em sua busca exaustiva, que havia ficado em último lugar entre os oito candidatos (Tabela 5.1). Essa inversão completa do ranking entre os desempenhos *in-sample* e *out-of-sample* constitui, provavelmente, a ilustração mais direta, neste relatório, de que a seleção de modelos baseada em critérios de informação e a validação preditiva respondem a perguntas distintas.

Figura 8.1: Comparação dos principais modelos no conjunto de teste (jan/2023–dez/2025).

Faixas verdes: intervalos de confiança de 80% e 95% do modelo Holt-Winters.



Fonte: Elaborado pelo Autor (2026).

Os candidatos SARIMA ajustados manualmente — independentemente da ordem (p, q) — permanecem visivelmente abaixo da série observada em boa parte dos meses de 2023 e 2025, formando um feixe relativamente compacto, enquanto Holt-Winters acompanha mais de perto a aceleração da produção observada nesses anos. Vale notar uma exceção pontual: em 2024, o viés do Holt-Winters se inverte — passa a superestimar discretamente a produção observada (viés médio de +111 mil m^3), em vez de subestimá-la — e sua vantagem de acurácia pontual sobre os candidatos SARIMA praticamente desaparece: o MAE do Holt-Winters naquele ano (551 mil m^3) cai dentro da própria faixa observada entre os sete candidatos SARIMA (542 a 568 mil m^3), ainda que estes continuem subestimando a produção de forma mais acentuada (viés entre -327 e -464 mil m^3). Essa aproximação é, no entanto, pontual e não se sustenta nos demais anos do teste (Tabela 8.1). O modelo Holt-Winters mantém um estado de tendência que é extrapolado linearmente (no espaço log) para todo o horizonte futuro. Os três modelos ARIMAX melhoram sensivelmente sobre os candidatos SARIMA manuais puros (MAPE de 6,12–6,34% contra 6,20–6,79%) — evidência de que os regressores de intervenção (em particular, os impulsos de 2001 e COVID-19, ambos altamente significativos, Capítulo 6) absorvem parte do erro de nível que, de outra forma, se propagaria para a previsão.

Um resultado adicional, que qualifica a leitura acima, é o desempenho do `auto.arima`: embora pertença à mesma família SARIMA e não inclua nenhum regressor de intervenção,

sua especificação — mais simples que a de todos os sete candidatos manuais, e com $D = 0$ e deriva incluída — obtém MAPE de 5,25%, o segundo melhor resultado entre os dezoito modelos comparados, superando inclusive o ARIMA+Fourier e todos os três modelos ARIMAX, e ficando atrás apenas do Holt-Winters. Isso reforça, sob uma quarta linha de evidência independente (Capítulos 4 e 5 já haviam apontado nessa direção pelos testes formais de D e pelo ranking de AICc/BIC), que a presença de um termo de deriva determinística ($D = 0$) tem, nesta série, mais influência sobre o desempenho preditivo de longo prazo do que a complexidade da parte autorregressiva/média-móvel do modelo.

8.3 Validação Rolling-Origin

A implementação desta validação é restrita a seis modelos: Naive Sazonal (piso de referência), Holt-Winters, o SARIMA campeão por AICc (SARIMA(8,1,1)(1,0,1)₁₂, Capítulo 5), e os três modelos não lineares/não paramétricos (NNAR, TBATS, Prophet). Os modelos com regressores de intervenção (os três ARIMAX) e o ARIMA+Fourier ficam fora deste desenho por uma razão metodológica específica: seus regressores foram construídos a partir de datas de quebra estrutural estimadas uma única vez, sobre todo o treino (Capítulo 6); recalculá-los a cada uma das sete origens exigiria repetir a própria detecção de outliers/quebras a cada corte, o que deixaria de medir a robustez da previsão e passaria a medir a estabilidade da detecção de intervenção — um objeto de teste diferente. O BSM é excluído pela mesma razão de escopo controlado que orienta a lista de modelos reajustados nas sete origens.

Tabela 8.2: MAPE médio (%) obtido na validação rolling-origin, sete origens (dez/2022 a dez/2024, passo de 4 meses), horizonte de 12 meses.

Modelo	MAPE médio (%)
Prophet	5,67
SARIMA(8,1,1)(1,0,1)₁₂ (campeão AICc)	6,00
Holt-Winters	6,32
TBATS	7,28
NNAR	8,13
Naive Sazonal	8,77

Fonte: Elaborado pelo Autor (2026).

O SARIMA campeão por AICc supera o Holt-Winters em horizonte de 12 meses — o oposto exato do padrão observado no *holdout* de 36 meses, no qual o mesmo modelo obtém MAPE de 6,542% contra 4,742% do Holt-Winters (Tabela 8.1). A tese de que “a família SARIMA sofre de uma limitação geral, independentemente do desenho de validação” não se sustenta: o que se sustenta, de forma mais precisa, é que essa limitação é específica de horizontes

longos — precisamente o que a lógica por trás da ausência de deriva já predizia (Capítulo 5, Seção 2.2) para os candidatos SARIMA com $D = 1$. Note-se, porém, que o próprio SARIMA campeão por AICc tem $D = 0$ e deriva incluída (embora estatisticamente não significativa, $p = 0,375$, Capítulo 5) — de modo que a explicação mecânica exata é discutida com mais cuidado a seguir.

8.4 Discussão: Por que os Desenhos Discordam, e o que Isso Significa

A divergência entre as Seções anteriores não é um erro de cálculo — é a manifestação esperada de uma fragilidade conhecida na literatura de previsão de séries temporais (HYNDMAN; ATHANASOPOULOS, 2018; TASHMAN, 2000): um único *holdout*, sobretudo com horizonte longo (36 meses), produz uma estimativa de erro de previsão com variância elevada, sensível a um único evento de grande magnitude situado próximo à fronteira do corte. Mas a explicação mecanicamente mais precisa vai além disso:

1. O horizonte de previsão, não apenas o desenho de validação, determina o desempenho relativo do SARIMA campeão. Em $h = 12$, a dinâmica transitória dos termos AR/MA ainda carrega informação útil sobre a trajetória recente da série; em $h = 36$, essa dinâmica já se dissipou quase completamente, e a extrapolação de longo prazo passa a depender quase inteiramente do termo de deriva — que, no modelo campeão SARIMA(8,1,1)(1,0,1)₁₂ ($D = 0$), é numericamente pequeno e estatisticamente não significativo ($\hat{\delta} = 0,0047$ por mês, $p = 0,375$; Capítulo 5). Ainda que não seja nulo, esse componente determinístico fixo não acompanha a magnitude da aceleração observada em 2023–2025 tão de perto quanto o estado de tendência de Holt-Winters, que é reestimado localmente a cada novo dado e pode se ajustar de forma mais responsiva a uma mudança de regime recente. A validação *rolling-origin* desta revisão, ao testar exatamente o mesmo SARIMA campeão sob $h = 12$, é consistente com essa leitura: sob um horizonte em que a dinâmica transitória ainda domina sobre a extrapolação de longo prazo, o mesmo modelo que perde de forma acentuada sob $h = 36$ supera o Holt-Winters sob $h = 12$.
2. Os testes de quebra estrutural do Capítulo 6 (Seção 6.3) mostram que a fronteira treino/teste (dezembro de 2022) não coincide com nenhuma quebra detectável dentro do treino. Isso não elimina a possibilidade de que a aceleração de 2023–2025 seja, ela mesma, um regime estruturalmente distinto (ainda não detectável a partir de dados anteriores a ela).

A implicação prática para este relatório é que a afirmação “Holt-Winters é o melhor modelo para esta série” — ou, simetricamente, “a família SARIMA é inadequada para esta

série” — não pode ser sustentada sem qualificação por horizonte. O que pode ser afirmado com segurança, à luz de toda a evidência reunida neste capítulo, é: (i) sob o desenho de *holdout* único de 36 meses, Holt-Winters obteve o menor erro entre os dezoito modelos comparados, e nenhum modelo obteve MASE inferior a 1 — ou seja, nenhum superou, em erro absoluto médio, a dificuldade histórica de previsão sazonal já observada dentro do próprio treino (o referencial embutido no denominador do MASE, Capítulo 2). Isso não deve ser confundido com o desempenho do candidato Naive Sazonal listado na Tabela 8.1: aquele modelo, por simplesmente repetir os doze valores do último ano do treino ao longo de todo o horizonte de teste, sem nenhum componente de tendência, obtém o pior resultado entre os dezoito candidatos (MASE = 3,861) — o referencial do MASE mede a dificuldade histórica de um passo sazonal dentro do treino, não o desempenho de um modelo naive efetivamente extrapolado para o futuro; (ii) sob um desenho *rolling-origin* de horizonte mais curto (12 meses), Prophet e o SARIMA campeão por AICc superam o Holt-Winters, em média; (iii) a causa estrutural mais provável para essa inversão é a interação entre o horizonte de previsão e a magnitude prática do componente de tendência determinística do SARIMA campeão — pequena e estatisticamente não significativa mesmo estando formalmente presente ($D = 0$, Capítulo 5) — frente ao estado de tendência adaptativo e localmente reestimado do Holt-Winters: uma limitação de horizonte longo, não uma ausência categórica de componente de crescimento nem uma limitação geral da família SARIMA.

8.5 Cobertura Empírica dos Intervalos de Previsão

Esta seção verifica a proporção de observações do conjunto de teste efetivamente contidas dentro dos intervalos de previsão nominais de 80% e 95% de cada modelo elegível — uma checagem quantitativa de calibração, apesar da rejeição de normalidade e da heterocedasticidade condicional já documentadas nos resíduos do modelo campeão (Capítulo 5).

Tabela 8.3: Cobertura empírica dos intervalos de previsão nominais (holdout, 36 meses).

Modelo	Cobertura emp. (nominal 80%)	Cobertura emp. (nominal 95%)
Holt-Winters	86,1%	94,4%
ARIMA + Fourier (parcimonioso)	80,6%	97,2%
<code>auto.arima</code>	86,1%	100,0%
SARIMA(2,1,4)(0,1,1) ₁₂	83,3%	97,2%
SARIMA(0,1,1)(1,0,1) ₁₂	88,9%	100,0%
SARIMA(1,1,1)(1,0,1) ₁₂ , SARIMA(8,1,1)(1,0,1) ₁₂ , SARIMA(1,1,2)(1,0,1) ₁₂ , SARIMA(8,1,1)(0,1,1) ₁₂	72,2%	94,4%
SARIMA(1,1,1)(0,1,1) ₁₂	69,4%	94,4%
ARIMAX (5 manuais)	80,6%	94,4%
ARIMAX (5 manuais + quebras estruturais)	83,3%	94,4%
ARIMAX (parcimonioso)	80,6%	94,4%
BSM	44,4%	63,9%
TBATS	44,4%	88,9%
NNAR	22,2%	44,4%
Prophet	27,8%	47,2%

Fonte: Elaborado pelo Autor (2026).

O achado mais importante desta seção é que os intervalos de previsão de Prophet e de NNAR — os dois modelos de natureza mais “orientada a dados”/black-box entre todos os testados — estão severamente mal calibrados. O intervalo de 95% do Prophet contém a observação real em apenas 47,2% dos casos, e o de 80% em apenas 27,8%; o NNAR é ainda pior (44,4% e 22,2%, respectivamente). Em contraste, Holt-Winters, ARIMA+Fourier e os três ARIMAX exibem cobertura de 80% consistentemente igual ou acima do nível nominal (entre 80,6% e 88,9%); a cobertura de 95%, por sua vez, situa-se ligeiramente abaixo do nominal na maioria desses modelos (94,4%), com exceção do ARIMA+Fourier, do `auto.arima` e do SARIMA(0,1,1)(1,0,1)₁₂, cuja cobertura de 95% iguala ou excede o nível nominal (97,2% a 100,0%).

Esta descoberta qualifica de forma decisiva o resultado da Seção anterior. O Prophet venceu a validação *rolling-origin* em acurácia pontual (MAPE médio de 5,67%, Tabela 8.2) — mas seus intervalos de incerteza permanecem dramaticamente mais estreitos do que deveriam. Uma interpretação plausível é que o modelo de incerteza do Prophet — baseado na variação histórica das taxas de mudança de tendência (*changepoints*) observadas *dentro* do próprio histórico de treino — subestima sistematicamente a possibilidade de uma aceleração de magnitude tão atípica quanto a observada em 2023–2025, precisamente o tipo de evento que os IC deveriam, por definição, acomodar com a frequência nominal declarada. A recomendação prática que se segue é que a boa acurácia pontual do Prophet não deve ser interpretada como evidência de boa quantificação de incerteza: para fins de comunicação de risco e construção de cenários (Seção 8.7), Holt-Winters e ARIMA+Fourier oferecem intervalos mais confiáveis.

8.6 Diagnóstico de Sobreajuste do Modelo Campeão SARIMA

Uma pergunta natural, dado o desempenho relativamente mais fraco da família SARIMA sob horizonte longo, é se o modelo campeão por AICc — SARIMA(8,1,1)(1,0,1)₁₂ — estaria sofrendo de sobreajuste. A evidência reunida neste relatório aponta para uma resposta mais específica do que um simples “sim” ou “não”: como discutido no Capítulo 5, o modelo tem apenas seis parâmetros livres (os coeficientes AR de ordem 2 a 7 foram fixados em zero, restando ϕ_1 , ϕ_8 , θ_1 , a componente sazonal autorregressiva e de média móvel, e a deriva), o que torna improvável o sobreajuste clássico por excesso de parâmetros. O enfraquecimento do desempenho *out-of-sample* em horizonte longo é mais bem descrito como uma limitação estrutural de subextrapolação de tendência: mesmo com deriva formalmente incluída ($D = 0$), sua magnitude estimada é pequena e estatisticamente não significativa ($p = 0,375$, Capítulo 5), de modo que a extrapolação de longo prazo do modelo se aproxima, na prática, de uma trajetória sem crescimento determinístico robusto — uma limitação que, em maior ou menor grau, é compartilhada pelos demais candidatos SARIMA testados (Tabela 8.1), e não uma característica exclusiva da ordem (8, 1, 1), e que, como demonstrado na Seção 8.4, se manifesta especificamente em horizontes longos.

8.7 Previsão para 2026

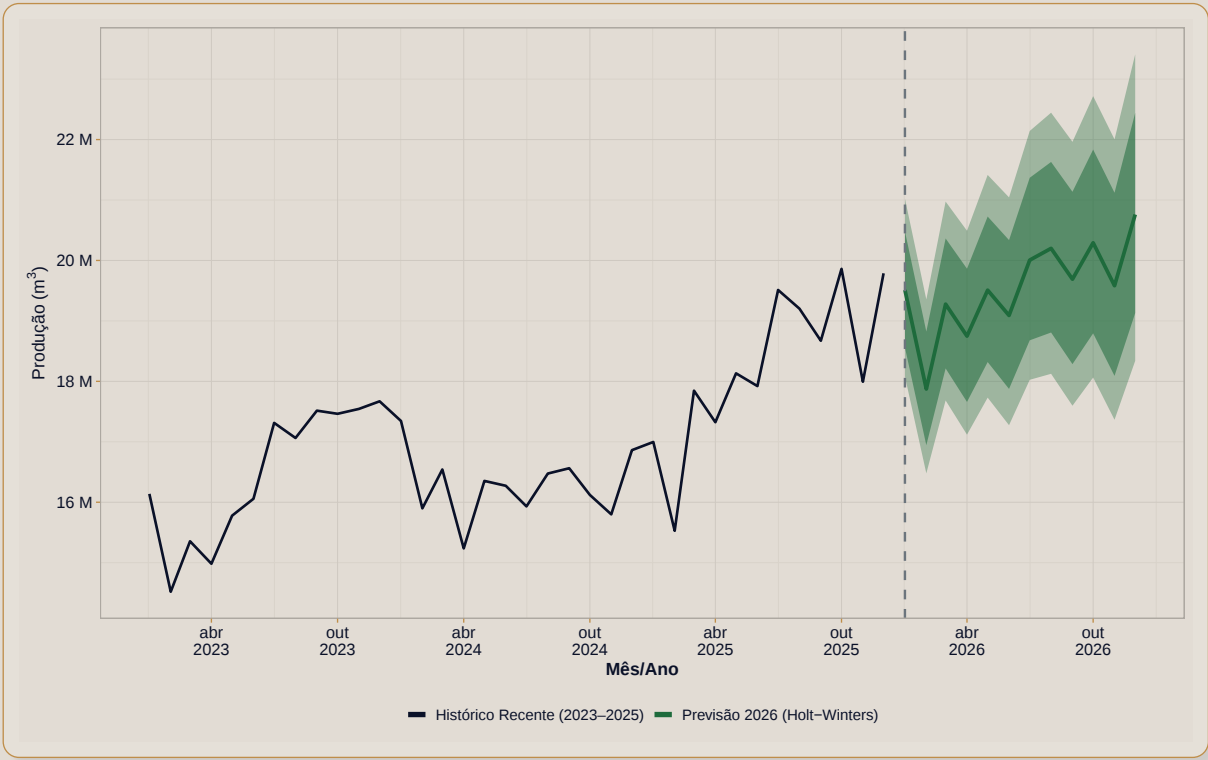
Tabela 8.4: Cenários de previsão mensal para 2026 (m³), por modelo.

Mês/2026	Holt-Winters	ARIMA + Fourier (parc.)	SARIMA(8,1,1)(1,0,1) ₁₂	Prophet	ARIMAX (parc.)
Janeiro	19.511.394	19.558.253	19.859.521	18.348.739	21.341.420
Fevereiro	17.873.531	17.779.174	17.844.214	16.491.707	19.073.367
Março	19.277.260	19.482.163	19.351.269	17.904.236	20.717.470
Abril	18.748.571	18.946.251	18.859.164	17.511.381	20.199.715
Maio	19.508.463	19.607.189	19.628.612	18.269.709	21.075.394
Junho	19.090.336	19.159.019	19.125.410	17.741.214	20.513.398
Julho	20.005.660	20.001.969	20.280.665	18.726.506	21.713.107
Agosto	20.199.513	20.061.604	20.292.527	18.849.004	21.748.031
Setembro	19.690.214	19.498.324	19.775.759	18.265.199	21.186.430
Outubro	20.292.192	20.069.502	20.289.872	18.762.125	21.773.148
Novembro	19.582.604	19.656.214	19.629.256	18.179.834	21.018.672
Dezembro	20.760.087	20.740.545	20.691.990	19.150.053	22.145.622

Fonte: Elaborado pelo Autor (2026).

Figura 8.2: Previsão da produção de petróleo no Brasil para 2026 — Holt-Winters.

Intervalos de confiança de 80% e 95%.

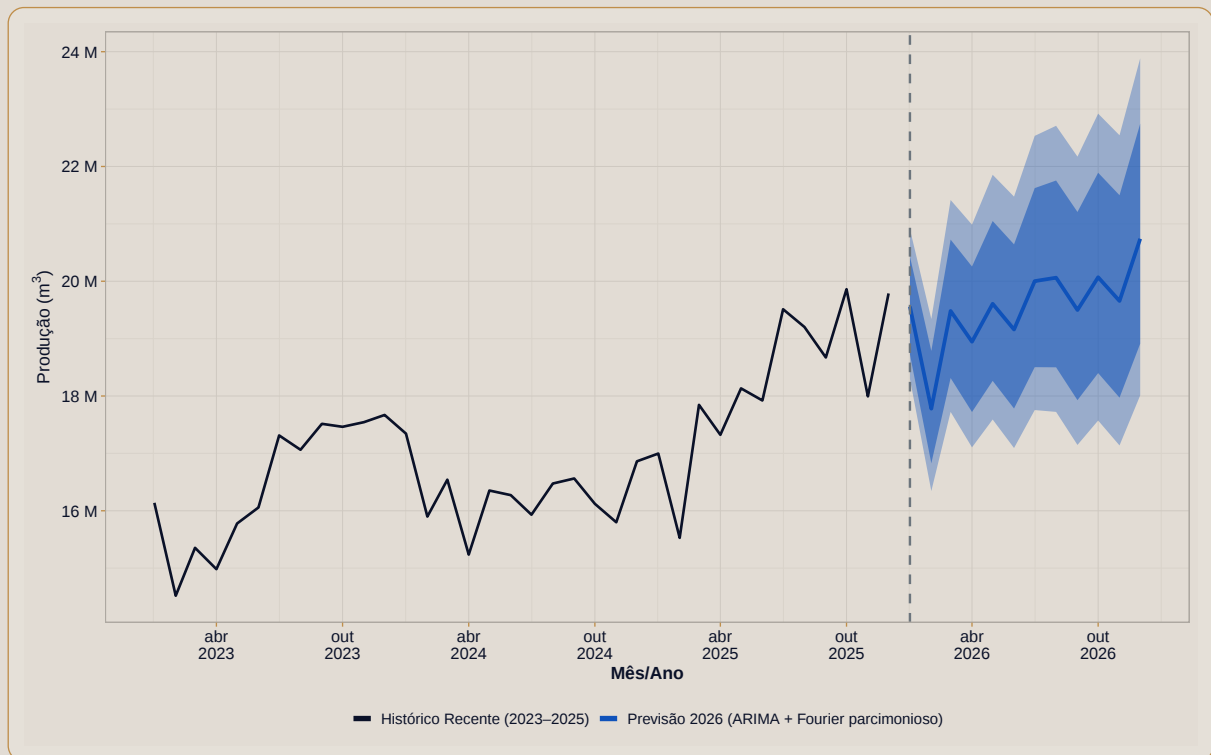


Fonte: Elaborado pelo Autor (2026).

A média mensal projetada para 2026 é de aproximadamente 19,54 milhões de m³ (Holt-Winters, total anual $\approx 234,5$ milhões de m³), 19,55 milhões de m³ (ARIMA + Fourier parcimonioso, $\approx 234,6$ milhões de m³), 19,64 milhões de m³ (SARIMA campeão por AICc, $\approx 235,6$ milhões de m³), 18,18 milhões de m³ (Prophet, $\approx 218,2$ milhões de m³) e 21,04 milhões de m³ (ARIMAX parcimonioso, $\approx 252,5$ milhões de m³). Holt-Winters, ARIMA+Fourier e o SARIMA campeão por AICc — três modelos com cobertura de IC próxima do nível nominal na Seção 8.5 — convergem de forma notável para um cenário de crescimento moderado em relação ao total de 2025 (218,8 milhões de m³, Capítulo 3), entre 7% e 8% ao ano; o cenário do Prophet — cuja boa acurácia pontual histórica vem acompanhada de intervalos de incerteza não confiáveis — projeta produção praticamente estável em relação a 2025.

O ARIMAX parcimonioso destoa visivelmente dos demais, projetando uma média $\approx 1,4$ milhão de m³ acima do SARIMA campeão (uma diferença relativa de $\approx 7,2\%$) e um crescimento anual de $\approx 15\%$ acima de 2025 — mais que o dobro do crescimento projetado pelos três modelos convergentes discutidos acima. A explicação é mecânica, não uma falha de ajuste: os três regressores do modelo (AO_choque2001, AO_covid e LS_quebral_bp, Capítulo 6) são, por construção, identicamente nulos no período de previsão — de modo que a previsão de 2026 reflete a trajetória ARMA+deriva *livre dos três choques negativos historicamente absorvidos pelos regressores*, e não a trajetória “como observada”. Como esses regressores estimam efeitos negativos ($-9,4\%$, $-7,8\%$ e $-6,4\%$, respectivamente, Capítulo 6) sobre uma base ARMA que, de outra forma, precisaria absorver esses choques dentro de sua própria dinâmica, a extrapolação “limpa” do ARIMAX tende a ser mais otimista do que a de modelos que nunca separaram explicitamente esses choques da tendência de fundo. Isso não torna o cenário do ARIMAX “errado” — apenas responde a uma pergunta contrafactual ligeiramente diferente (“qual seria a trajetória na ausência de novos choques da mesma natureza dos já modelados”).

Figura 8.3: Previsão da produção de petróleo no Brasil para 2026 — ARIMA + Fourier (parc.).
Intervalos de confiança de 80% e 95%.



Fonte: Elaborado pelo Autor (2026).

Mês a mês, a dispersão entre os cinco cenários é ampla: o valor mensal mínimo ocorre em fevereiro/Prophet (16,49 milhões de m^3) e o máximo em dezembro/ARIMAX parcimonioso (22,15 milhões de m^3). A recomendação é tratar o intervalo entre os cinco cenários (aproximadamente 16,5 a 22,1 milhões de m^3 /mês, dependendo do mês e do modelo) como a faixa de incerteza substantiva da previsão para 2026, com a média simples dos cinco modelos (19,59 milhões de m^3 /mês, $\approx 235,1$ milhões de m^3 no ano) como estimativa pontual de consenso, em vez de um único valor pontual de um modelo isolado.

CONCLUSÃO

9.1 Síntese dos Achados

Este estudo investigou a existência de um ciclo sazonal anual na produção mensal de petróleo no Brasil (1997–2025) e comparou um conjunto amplo de modelos preditivos. Três conclusões emergem:

1. Existe sazonalidade determinística estatisticamente confirmada, mas sua estrutura é mais complexa do que uma onda senoidal anual simples — e parte dessa complexidade tem origem mecânica, não operacional. A energia espectral está distribuída por harmônicos de período curto (Capítulo 7), refletindo a forte assimetria do padrão sazonal. A quantificação direta do efeito de calendário (Capítulo 3) mostra que aproximadamente 47% do coeficiente de variação sazonal aparente desaparece quando a produção é ajustada por dias-no-mês — e que o suposto “pico de março” é, em boa parte, ilusório: o desvio ajustado por dia de março é negativo, e mais próximo do desvio (também ajustado) de fevereiro do que a leitura bruta sugeria. A sazonalidade genuína remanescente, de natureza operacional, é o que de fato compete pela energia espectral discutida no Capítulo 7.
2. A série exibe instabilidade estrutural genuína e estatisticamente confirmada, mas a datação específica dessa instabilidade depende criticamente de como — e sobre qual janela — ela é estimada. Os testes CUSUM confirmam instabilidade sob qualquer janela de estimação. O procedimento de Bai-Perron, no entanto, produz datas de quebra substancialmente diferentes quando restrito ao treino (jan/2001, fev/2010, abr/2014) em comparação à série completa (out/2002, mar/2012, out/2020) — e, em particular, a quebra de outubro/2020, embora situada dentro do próprio período de treino, só emerge como estatisticamente detectável quando o procedimento tem acesso a observações posteriores a 2022 para estimar com confiança a inclinação do novo regime — exatamente o tipo de dependência de janela de estimação que motiva restringir ao treino os regressores de intervenção efetivamente usados na modelagem preditiva (Capítulo 6).
3. A escolha do modelo preditivo “superior” depende criticamente do horizonte de previsão — mais do que do desenho de validação em si. Sob um *holdout* estático de 36 meses, Holt-Winters supera SARIMA, ARIMAX, Prophet e os demais candidatos, e — sob a

definição correta de MASE/U-Theil (Capítulo 2) — nenhum modelo elegível supera um previsor naive sazonal de referência. Sob uma validação *rolling-origin* de horizonte mais curto (12 meses), o quadro se inverte de forma mais matizada do que uma simples “divergência entre desenhos”: o SARIMA campeão por AICc supera o Holt-Winters, evidenciando que a limitação estrutural de subextrapolação de tendência — decorrente não da ausência categórica de deriva, mas de sua magnitude pequena e estatisticamente não significativa mesmo quando formalmente incluída ($D = 0$, Capítulo 5) — penaliza especificamente horizontes *longos*, e não constitui uma fragilidade geral da família SARIMA. A boa acurácia pontual do Prophet na validação *rolling-origin* (campeão, MAPE médio de 5,67%), por fim, vem acompanhada de intervalos de previsão severamente mal calibrados (cobertura empírica de 27,8% a 47,2% contra níveis nominais de 80%/95% — Capítulo 8).

9.2 Limitações do Estudo

- A avaliação *out-of-sample* reportada neste relatório baseia-se, em sua forma final, em um único *holdout* para a maior parte das comparações detalhadas (Tabela 8.1); a validação *rolling-origin* (Tabela 8.2) mitiga parcialmente esse problema e, passou a ser efetivamente implementada, incluindo o candidato SARIMA campeão por AICc ao lado de Holt-Winters, NNAR, TBATS, Prophet e Naive Sazonal — mas ainda cobre apenas sete origens, um único horizonte (12 meses), e deixa de fora os candidatos ARIMAX e ARIMA+Fourier (cujos regressores de intervenção, fixados sobre todo o treino, não se prestam a um redesenho consistente em múltiplas origens sem repetir a própria detecção de quebras a cada corte — Capítulo 8). Um desenho mais robusto, com mais origens, múltiplos horizontes (6, 12, 24 meses) e uma estratégia de regressores de intervenção compatível com múltiplas origens, continua sendo desejável.
- Não foram incorporadas variáveis exógenas com poder preditivo genuíno sobre o período de teste (capacidade instalada, cronograma de novas unidades de produção, indicadores de investimento) — as variáveis de intervenção utilizadas (Capítulo 6) são, por construção, identicamente nulas no período de previsão.
- A não normalidade e a heterocedasticidade condicional dos resíduos do modelo campeão SARIMA (Capítulo 5) implicam que os intervalos de previsão baseados em normalidade assintótica devem ser interpretados com cautela adicional; a verificação de cobertura empírica (Capítulo 8) já fornece evidência quantitativa direta dessa cautela, mostrando que essa preocupação é especialmente severa para Prophet e NNAR.
- O procedimento automático de detecção de outliers ($t_{so}()$) exibe sensibilidade não trivial ao otimizador numérico e à versão dos pacotes utilizados (Capítulo 6) — uma

limitação de reprodutibilidade do método, que recomenda a fixação de um *lockfile* de pacotes (ver Apêndice A) para qualquer reexecução futura deste *script*.

9.3 Distinção Frente às Projeções Estruturais Oficiais

Cabe distinguir o exercício conduzido neste trabalho da previsão oficial de produção de petróleo no Brasil, publicada anualmente pela Empresa de Pesquisa Energética (EPE) no Caderno de Previsão da Produção de Petróleo e Gás Natural, parte do Plano Decenal de Expansão de Energia Empresa de Pesquisa Energética (2024) — e cujas premissas de novos projetos alinham-se, no horizonte de médio prazo, com as metas de produção do Plano de Negócios da Petrobras Petrobras (2025). Aquela previsão é construída de forma estrutural e ascendente (*bottom-up*), agregando o cronograma de entrada em operação de cada unidade estacionária de produção, a taxa de declínio esperada de campos maduros e o grau de maturidade de cada categoria de recurso (reserva, contingente, não descoberto) — um paradigma de engenharia de reservatório e planejamento de projetos, não de modelagem estatística da série histórica agregada.

Os dois exercícios não são substitutos um do outro: a abordagem estrutural da EPE é superior para capturar mudanças de patamar já contratadas (a exemplo da própria expansão do pré-sal, tratada neste trabalho via regressores de intervenção — Capítulo 6), ao passo que a abordagem estatística aqui conduzida — com validação preditiva explícita, intervalos de previsão calibrados empiricamente e comparação sistemática entre famílias de modelos — responde a uma pergunta que os documentos oficiais, por desenho, não respondem: qual é o erro de previsão esperado, fora da amostra, de um exercício puramente estatístico sobre a série agregada, e como esse erro se compara entre diferentes classes de modelo.

9.4 Recomendações e Trabalhos Futuros

1. Modelar explicitamente o efeito de calendário (dias-no-mês) como regressor ou como ajuste de pré-processamento em todos os modelos subsequentes, em vez de apenas documentá-lo qualitativamente (Capítulo 3) — esta revisão quantifica o efeito, mas nenhum modelo de previsão o incorpora diretamente; isso pode reduzir a riqueza harmônica espúria identificada no Capítulo 7 e simplificar a sazonalidade residual genuína a ser modelada.
2. Versionar as dependências de pacotes R deste projeto com `renv` (Apêndice A), de modo a mitigar a sensibilidade de reprodutibilidade documentada para o procedimento `tso()` e, potencialmente, para outros procedimentos numericamente iterativos do *pipeline*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, D. et al. A crise no fornecimento e distribuição de energia elétrica no brasil em 2001: uma análise panorâmica com foco em na prevenção de eventos análogos futuros. *Revista de Gestão Ambiental e Sustentabilidade*, v. 11, p. 1–21, 03 2022.
- BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, v. 18, n. 1, p. 1–22, 2003.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.
- BOX, G. E. P.; TIAO, G. C. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, v. 70, n. 349, p. 70–79, 1975.
- BROCKWELL, P. J.; DAVIS, R. A. *Time Series: Theory and Methods*. 2nd. ed. New York: Springer, 1991.
- BROWN, R. L.; DURBIN, J.; EVANS, J. M. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B*, v. 37, n. 2, p. 149–192, 1975.
- BURNHAM, K. P.; ANDERSON, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd. ed. New York: Springer, 2002.
- CANOVA, F.; HANSEN, B. E. Are seasonal patterns constant over time? a test for seasonal stability. *Journal of Business & Economic Statistics*, v. 13, n. 3, p. 237–252, 1995.
- CHEN, C.; LIU, L.-M. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, v. 88, n. 421, p. 284–297, 1993.
- CLEVELAND, R. B. et al. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, v. 6, n. 1, p. 3–73, 1990.
- COX, D. R.; STUART, A. Some quick sign tests for trend in location and dispersion. *Biometrika*, v. 42, n. 1–2, p. 80–95, 1955.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, v. 74, n. 366, p. 427–431, 1979.
- Empresa de Pesquisa Energética. *Previsão da Produção de Petróleo e Gás Natural — Plano Decenal de Expansão de Energia 2034*. Brasília, 2024.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, v. 50, n. 4, p. 987–1007, 1982.

- FISHER, R. A. Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A*, v. 125, n. 796, p. 54–59, 1929.
- GUERRERO, V. M. Time-series analysis supported by power transformations. *Journal of Forecasting*, v. 12, n. 1, p. 37–48, 1993.
- HAMILTON, J. D. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.
- HARVEY, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press, 1989.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2nd. ed. Melbourne, Australia: OTexts, 2018.
- HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, v. 22, n. 4, p. 679–688, 2006.
- KENDALL, M. G. *Rank Correlation Methods*. 4th. ed. London: Charles Griffin, 1975.
- KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, v. 54, n. 1–3, p. 159–178, 1992.
- MANN, H. B. Nonparametric tests against trend. *Econometrica*, v. 13, n. 3, p. 245–259, 1945.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. 2a. ed. São Paulo: Editora Blucher, 2006.
- OSBORN, D. R. et al. Seasonality and the order of integration for consumption. *Oxford Bulletin of Economics and Statistics*, v. 50, n. 4, p. 361–377, 1988.
- Petrobras. *Plano Estratégico e Plano de Negócios 2026-2030*. Rio de Janeiro, 2025.
- PHILLIPS, P. C. B.; PERRON, P. Testing for a unit root in time series regression. *Biometrika*, v. 75, n. 2, p. 335–346, 1988.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2024. Disponível em: <<https://www.R-project.org/>>.
- SHUMWAY, R. H.; STOFFER, D. S. *Time Series Analysis and Its Applications: With R Examples*. 4th. ed. New York: Springer, 2017.
- TASHMAN, L. J. Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, v. 16, n. 4, p. 437–450, 2000.
- TSAY, R. S.; TIAO, G. C. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*, v. 79, n. 385, p. 84–96, 1984.
- WEI, W. W. S. *Time Series Analysis: Univariate and Multivariate Methods*. 2nd. ed. Boston: Pearson, 2006.

APÊNDICE DE REPRODUTIBILIDADE

Este apêndice documenta o ambiente de execução e os mecanismos de controle de aleatoriedade e de falha numérica do script definitivo, de modo que os resultados numéricos apresentados neste relatório possam ser reproduzidos ou auditados de forma independente.

A.1 Ambiente de Execução

- **R:** versão 4.4.2 (2024-10-31 ucrt), plataforma `x86_64-w64-mingw32/x64`, sistema operacional Windows 11 x64 (*build* 26200), *locale* `Portuguese_Brazil.utf8`, fuso horário `America/Fortaleza`.
- **Pacotes-chave e versões:** `forecast` 9.0.2, `tseries` 0.10-61, `tsoutliers` 0.6-10, `seastests` 0.15.4, `strucchange` 1.5-4, `car` 3.1-3, `prophet` 1.1.7, `nanoparquet` 0.5.1, `Kendall` 2.2.2, `randtests` 1.0.2, `TSA` 1.3.1, `dplyr` 1.2.1, `ggplot2` 4.0.3, entre os 26 pacotes carregados pelo script (Capítulo 2).

A.2 Controle de Aleatoriedade

Todos os procedimentos que dependem de inicialização aleatória (`nnetar()`, `prophet()`, `tbats()`) são precedidos, de forma imediata, pela fixação de uma semente global única (`SEMENTE_GLOBAL`) via uma função auxiliar (`with_seed()`), tornando cada ajuste individualmente reproduzível. Como esses três procedimentos são reajustados tanto na avaliação única (Capítulo 8) quanto nas sete origens da validação *rolling-origin* (oito ajustes cada, no total), a fixação da semente imediatamente antes de cada chamada — e não uma única vez no início do script — é o que garante que a ordem de execução não afete os resultados.

A.3 Sensibilidade Numérica do Procedimento `tso()`

Conforme documentado no Capítulo 6, a chamada combinada de `tsoutliers::tso()` com múltiplos tipos de outlier e mais de uma iteração invoca, em seu laço interno, uma otimização por diferenças finitas (BFGS) que pode falhar ("*non-finite finite-difference value*") dependendo da versão de R, do pacote `forecast` e das bibliotecas BLAS/LAPACK do sistema — com o mesmo código-fonte e os mesmos dados de entrada. O script implementa

uma função de nova tentativa (`tso_robusto()`) que substitui o otimizador padrão por um método livre de gradiente (Nelder-Mead) caso o BFGS falhe. Esse comportamento é específico ao ambiente numérico e pode não se repetir de forma idêntica em outro ambiente, ainda que o mecanismo de *fallback* garanta que o *script* sempre produza uma saída válida.

A.4 Recomendação para Reexecução Futura

Para que resultados numéricos futuros permaneçam diretamente comparáveis aos deste relatório, recomenda-se fixar as versões exatas de R e de todos os pacotes listados nesta seção — por exemplo, via um *lockfile* gerenciado pelo pacote `renv` — antes de qualquer reexecução do script definitivo, especialmente para fins de auditoria dos procedimentos numericamente iterativos identificados (`tso()`), e os ajustes de `nnetar()`/`tbats()`/`prophet()` nas sete origens da validação *rolling-origin*.



EMAIL

douglascmoura21@gmail.com



LINKEDIN

linkedin.com/in/douglas-chaves-moura



GITHUB

github.com/douglascmoura



WEBSITE

douglas-moura-portfolio.pages.dev

