

# ANÁLISE ECONOMÉTRICA DO IMPACTO DA OFERTA DE ENERGIA ELÉTRICA E PETRÓLEO NA INFLAÇÃO NO BRASIL



Estudo de Séries Temporais (2000-2024)

Autor: Diego Ribeiro Porto

MAIO DE 2025

"a econometria pode ser definida como a análise quantitativa dos fenômenos econômicos ocorridos com base no desenvolvimento paralelo da teoria e das observações e com o uso de métodos de inferência adequados."

— Samuelson, Koopmans, Stone

# ANÁLISE ECONOMÉTRICA DO IMPACTO DA OFERTA DE ENERGIA ELÉTRICA E PETRÓLEO NA INFLAÇÃO NO BRASIL

## 1. Introdução

A inflação é um fenômeno econômico que afeta diretamente o poder de compra da população e a estabilidade econômica de um país. No Brasil, a inflação tem sido uma preocupação constante, com variações significativas ao longo dos anos. Neste contexto, a oferta de energia elétrica e petróleo desempenha um papel crucial na formação dos preços, uma vez que esses insumos são fundamentais para a produção e o consumo de bens e serviços. Neste trabalho, buscamos analisar o impacto da oferta de energia elétrica e petróleo na inflação\* no Brasil, utilizando dados mensais de 2000 a 2024. A análise será realizada por meio de modelos econométricos, com o objetivo de identificar a relação entre essas variáveis e a inflação, bem como avaliar a magnitude e a significância desse impacto.

*"A inflação pode ser conceituada como um aumento contínuo e crítico e generalizado no nível de preços. Ou seja, os movimentos inflacionários representam elevações em todos os bens produzidos pela economia e não meramente o aumento de um determinado preço."* (LUQUE, Carlos; VASCONCELLOS, Marco. Considerações sobre o Problema da Inflação. In: PINHO, Diva; VASCONCELLOS, Marco (Orgs.). Manual de economia. 5. ed. São Paulo: Saraiva, 2006. p. 336.)

## 2. Metodologia

A metodologia adotada para a análise do impacto da oferta de energia elétrica armazenada como água nos reservatórios - ( ear ) e da produção de petróleo cru ( oleo\_bruto ) sobre a inflação brasileira medida pelo Índice Nacional de Preços ao Consumidor Amplo - ( ipca ) segue uma abordagem baseada em séries temporais mensais no período de 2000 a 2024, estruturada nas seguintes etapas:

### A. Coleta de Dados

As séries temporais mensais de ipca , ear e oleo\_bruto serão coletadas a partir de fontes oficiais, cobrindo o período de janeiro de 2000 a dezembro de 2024. Os dados foram obtidos no Sistema de Séries Temporais do Banco Central do Brasil (SGS) e no site do Operador Nacional do Sistema Elétrico (ONS).

### B. Análise Gráfica Exploratória

As três variáveis serão visualizadas graficamente para identificar padrões de tendência, sazonalidade, possíveis rupturas estruturais e presença de outliers. Serão utilizados histogramas, boxplots, QQ-plots, gráficos de densidade e violinos para apoiar essa caracterização.

### C. Testes de Estacionariedade

Serão aplicados testes de estacionariedade, em especial o **Teste de Dickey-Fuller Aumentado (ADF)**, para verificar se as séries possuem média e variância constantes ao longo do tempo. Em caso negativo, serão aplicadas transformações como a **diferenciação**, com o objetivo de torná-las estacionárias, requisito necessário para a modelagem econométrica.

### D. Avaliação de Defasagens e Correlações Temporais

A relação temporal entre as variáveis será analisada por meio de **correlação cruzada** (Cross-Correlation Function – CCF), a fim de identificar **defasagens significativas** entre as variáveis explicativas ( `ear` , `oleo_bruto` ) e a variável explicada ( `ipca` ).

### E. Modelagem Econométrica com Séries Temporais

Serão estimados modelos econôméticos apropriados para séries temporais, como:

- **ARIMAX**: quando `ipca` é modelado em função de suas próprias defasagens e variáveis exógenas;

### F. Avaliação do Modelo e Validação Fora da Amostra

A qualidade dos modelos será avaliada por meio da análise dos resíduos (autocorrelação, normalidade, homocedasticidade).

### G. Discussão e Conclusões

Os resultados obtidos serão interpretados à luz da teoria econômica e da conjuntura brasileira. Serão discutidas as implicações para políticas públicas e sugestões para estudos futuros.

## 3. Análise Descritiva

A análise descritiva das variáveis foi realizada com o objetivo de identificar tendências, sazonalidades e possíveis outliers. A seguir, apresentamos gráficos e estatísticas descritivas das variáveis consideradas na análise.

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import statsmodels.api as sm
from statsmodels.graphics.tsaplots import plot_ccf, plot_acf, plot_pacf
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
```

```
# Configurações para melhorar a visualização dos gráficos
plt.style.use('ggplot')
plt.rcParams['figure.figsize'] = (12, 8)
sns.set_palette('Set2')
sns.set_context('talk')
```

```
In [2]: df = pd.read_parquet('dados.parquet')
```

```
In [3]: df.head(5)
```

```
Out[3]:
```

	data	ear	u.m_ear	oleo_bruto	u.m_oleo	ipca	u.m_ipca
0	2000-01-01	41.344962	percentual	1155	Barris/dia (mil)	0.62	percentual
1	2000-02-01	52.108976	percentual	1121	Barris/dia (mil)	0.13	percentual
2	2000-03-01	62.794431	percentual	1189	Barris/dia (mil)	0.22	percentual
3	2000-04-01	66.530108	percentual	1161	Barris/dia (mil)	0.42	percentual
4	2000-05-01	61.894345	percentual	1159	Barris/dia (mil)	0.01	percentual

```
In [4]: df.tail(5)
```

```
Out[4]:
```

	data	ear	u.m_ear	oleo_bruto	u.m_oleo	ipca	u.m_ipca
295	2024-08-01	61.476806	percentual	3340	Barris/dia (mil)	-0.02	percentual
296	2024-09-01	50.736338	percentual	3470	Barris/dia (mil)	0.44	percentual
297	2024-10-01	47.900840	percentual	3269	Barris/dia (mil)	0.56	percentual
298	2024-11-01	48.537575	percentual	3310	Barris/dia (mil)	0.39	percentual
299	2024-12-01	52.692861	percentual	3419	Barris/dia (mil)	0.52	percentual

## Descrição dos Dados

Os dados fornecidos representam séries temporais mensais com as seguintes variáveis:

- `data` : Data de referência.
- `ear` : Energia elétrica armazenada como água nos reservatórios, medida em percentual.
- `u.m_ear` : Unidade de medida da energia elétrica armazenada.
- `oleo_bruto` : Produção de petróleo cru, medida em Barris/dia (mil).
- `u.m_oleo` : Unidade de medida da produção de petróleo.
- `ipca` : Índice Nacional de Preços ao Consumidor Amplo, representando a inflação brasileira, medida em percentual.
- `u.m_ipca` : Unidade de medida do índice de inflação.

```
In [5]: # Verificar informações básicas do DataFrame
print("Informações do DataFrame:")
print(f"Shape: {df.shape}")
print("\nInformações das colunas:")
df.info()
```

```
Informações do DataFrame:
Shape: (300, 7)

Informações das colunas:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   data        300 non-null    datetime64[ns]
 1   ear         300 non-null    float64 
 2   u.m_ear     300 non-null    object  
 3   oleo_bruto  300 non-null    int64   
 4   u.m_oleo    300 non-null    object  
 5   ipca        300 non-null    float64 
 6   u.m_ipca    300 non-null    object  
dtypes: datetime64[ns](1), float64(2), int64(1), object(3)
memory usage: 16.5+ KB
```

```
In [6]: # Estatísticas descritivas do DataFrame
print("Estatísticas descritivas:")
display(df.describe(exclude=['datetime64[ns]', 'object']).T)
```

Estatísticas descritivas:

	count	mean	std	min	25%	50%	75%	max
<b>ear</b>	300.0	60.751847	13.644838	30.028935	50.560811	60.566563	71.783052	87.47971
<b>oleo_bruto</b>	300.0	2195.160000	631.169277	1121.000000	1695.000000	2077.500000	2636.500000	3678.00000
<b>ipca</b>	300.0	0.501133	0.394905	-0.680000	0.260000	0.450000	0.675000	3.02000

```
In [7]: # Lista todas as colunas no DataFrame exceto a data (se houver)
numeric_columns = df.select_dtypes(include=['int64', 'float64']).columns.tolist()

# Exibir a lista de variáveis que serão analisadas
print(f"Variáveis numéricas disponíveis para análise: {numeric_columns}")
```

Variáveis numéricas disponíveis para análise: ['ear', 'oleo\_bruto', 'ipca']

```
In [8]: def plot_variable_distribution(dataframe, variable):
    """
        Função para criar múltiplos gráficos de distribuição para uma variável
    """
    fig, axes = plt.subplots(2, 2, figsize=(16, 12))
    fig.suptitle(f'Análise de Distribuição: {variable}', fontsize=16)

    # Histograma com curva de densidade
    sns.histplot(dataframe[variable], kde=True, ax=axes[0, 0])
    axes[0, 0].set_title('Histograma com Curva de Densidade')
    axes[0, 0].set_xlabel(variable)
    axes[0, 0].grid(True)

    # Boxplot
    sns.boxplot(y=dataframe[variable], ax=axes[0, 1])
    axes[0, 1].set_title('Boxplot')
    axes[0, 1].set_ylabel(variable)
    axes[0, 1].grid(True)

    # QQ Plot para verificar normalidade
    from scipy import stats
    qq = stats.probplot(dataframe[variable].dropna(), dist="norm", plot=axes[1, 0])
    axes[1, 0].set_title('QQ Plot (Verificação de Normalidade)')
```

```

axes[1, 0].grid(True)

# Gráfico de violino
sns.violinplot(y=dataframe[variable], ax=axes[1, 1])
axes[1, 1].set_title('Violin Plot')
axes[1, 1].set_ylabel(variable)
axes[1, 1].grid(True)

plt.tight_layout(rect=[0, 0, 1, 0.96])
plt.show()

# Estatísticas descritivas específicas da variável
print(f"\nEstatísticas descritivas para {variable}:")
stats_df = pd.DataFrame({
    'Média': [dataframe[variable].mean()],
    'Mediana': [dataframe[variable].median()],
    'Desvio Padrão': [dataframe[variable].std()],
    'Variância': [dataframe[variable].var()],
    'Mínimo': [dataframe[variable].min()],
    'Máximo': [dataframe[variable].max()],
    'Assimetria': [dataframe[variable].skew()],
    'Curtose': [dataframe[variable].kurtosis()]
})
display(stats_df.T)

# Verificar normalidade com teste estatístico
print(f"\nTeste de Normalidade (Shapiro-Wilk) para {variable}:")
shapiro_test = stats.shapiro(dataframe[variable].dropna())
print(f"Estatística de teste: {shapiro_test[0]:.4f}")
print(f"Valor p: {shapiro_test[1]:.4f}")
if shapiro_test[1] < 0.05:
    print("Conclusão: A distribuição não parece ser normal (rejeita hipótese nula).")
else:
    print("Conclusão: A distribuição pode ser considerada normal (não rejeita hipótese nula).")

```

In [9]: # Loop para criar gráficos de distribuição para cada variável numérica

```

for column in numeric_columns:
    print(f"\n'*80")
    print(f"Análise de Distribuição para: {column}")

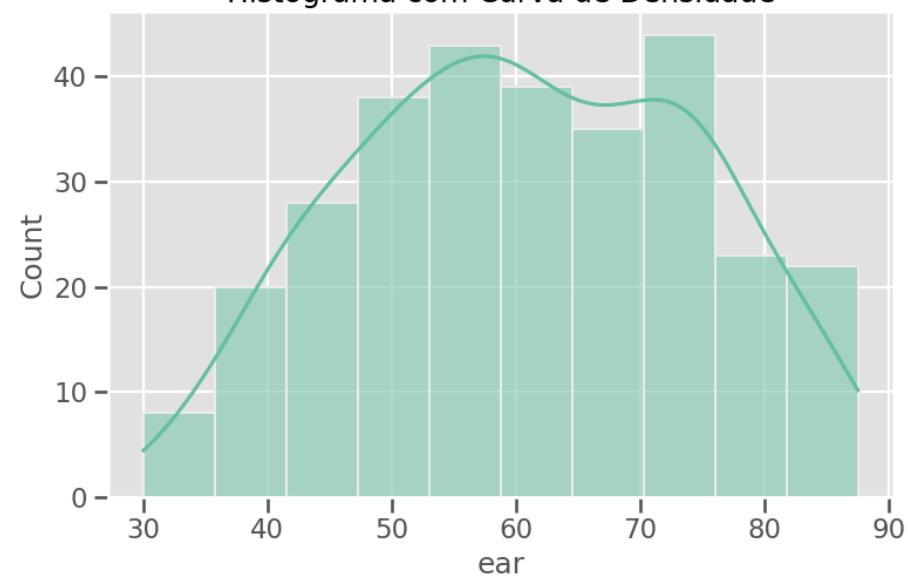
```

```
print(f"{'='*80}")
plot_variable_distribution(df, column)
```

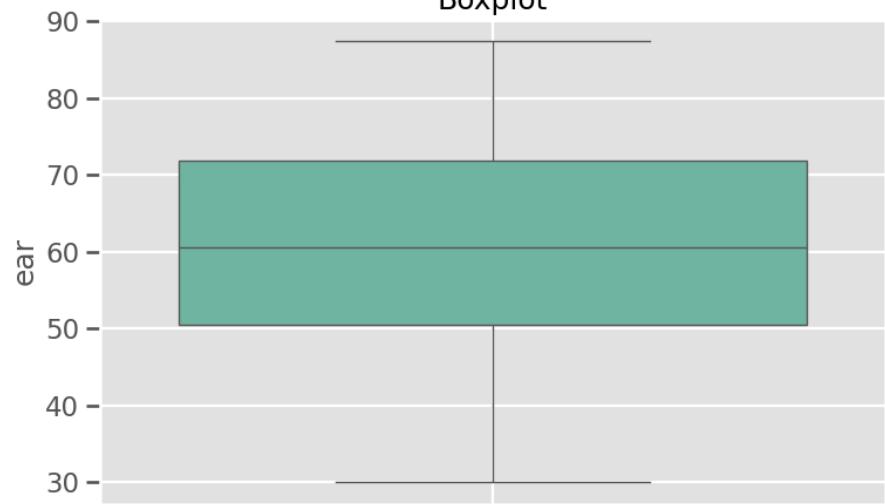
```
=====
Análise de Distribuição para: ear
=====
```

Análise de Distribuição: ear

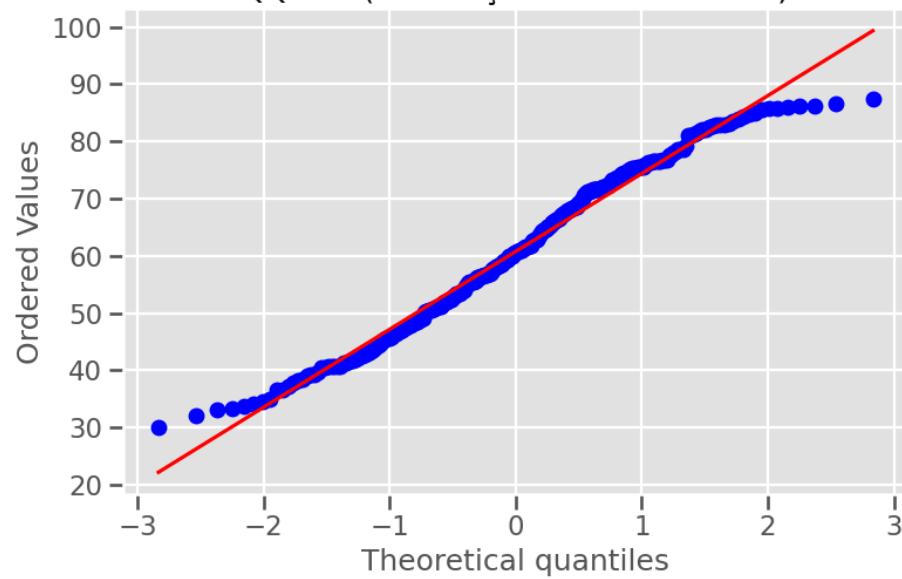
Histograma com Curva de Densidade



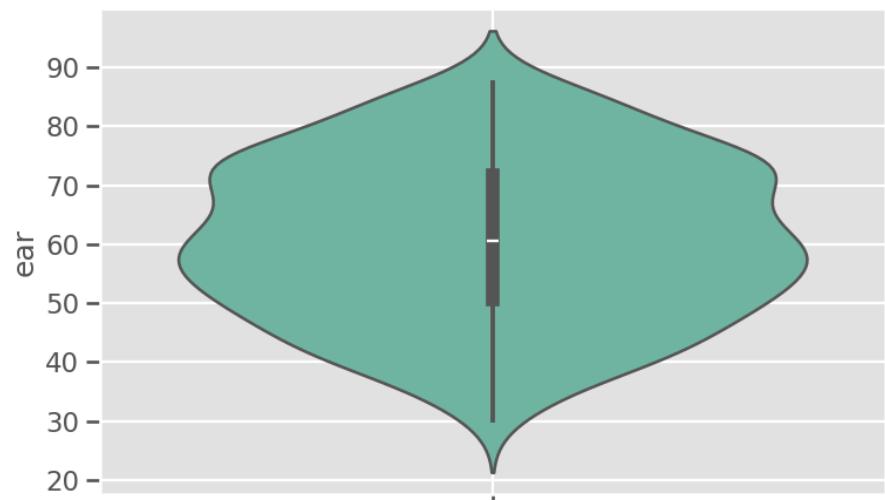
Boxplot



QQ Plot (Verificação de Normalidade)



Violin Plot



Estatísticas descritivas para ear:

	0
<b>Média</b>	60.751847
<b>Mediana</b>	60.566563
<b>Desvio Padrão</b>	13.644838
<b>Variância</b>	186.181591
<b>Mínimo</b>	30.028935
<b>Máximo</b>	87.479710
<b>Assimetria</b>	-0.039185
<b>Curtose</b>	-0.877915

Teste de Normalidade (Shapiro-Wilk) para ear:

Estatística de teste: 0.9806

Valor p: 0.0004

Conclusão: A distribuição não parece ser normal (rejeita hipótese nula).

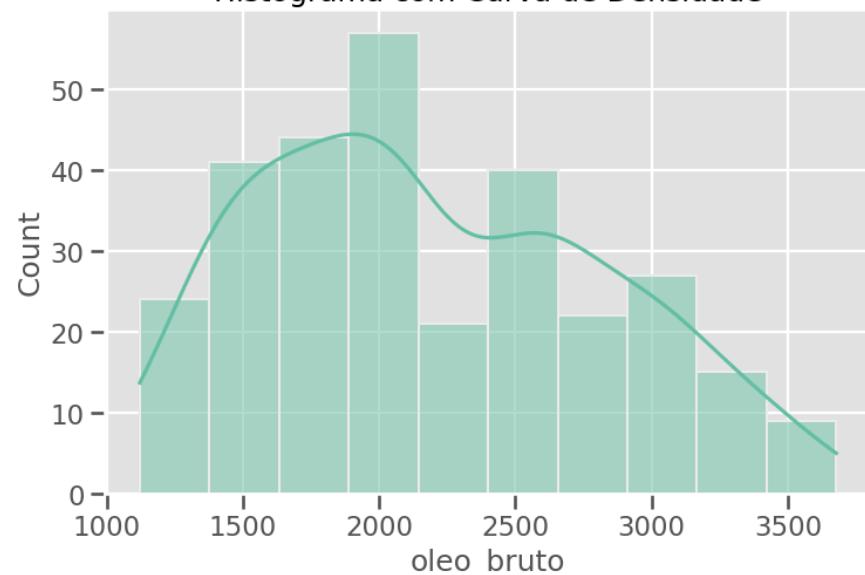
=====

Análise de Distribuição para: oleo\_bruto

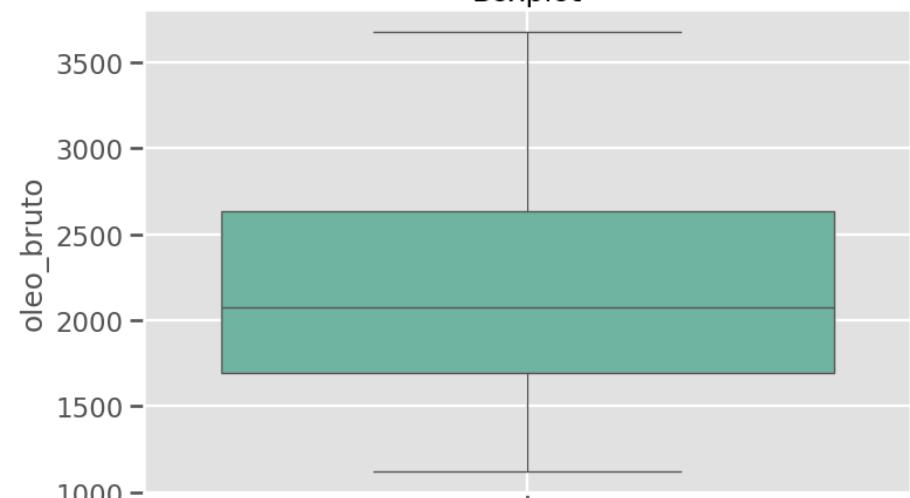
=====

## Análise de Distribuição: oleo\_bruto

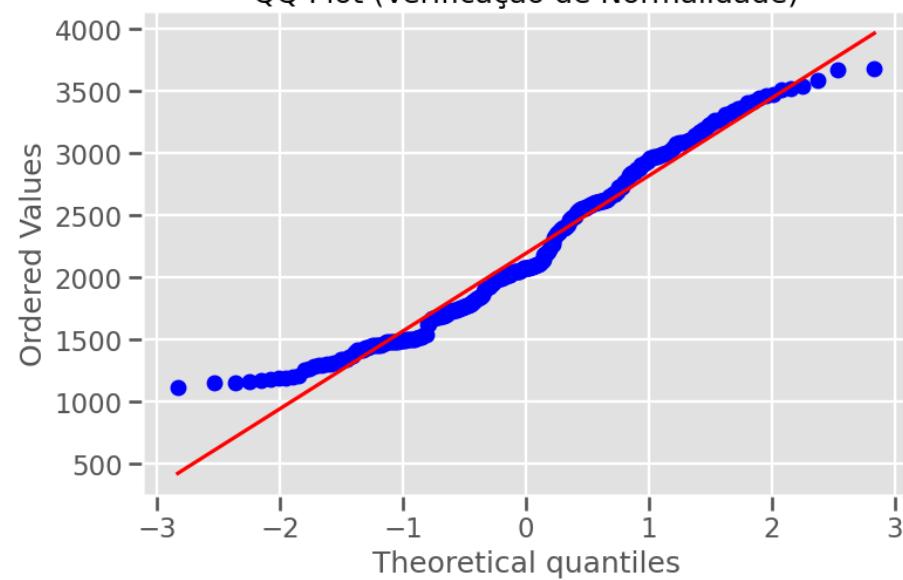
Histograma com Curva de Densidade



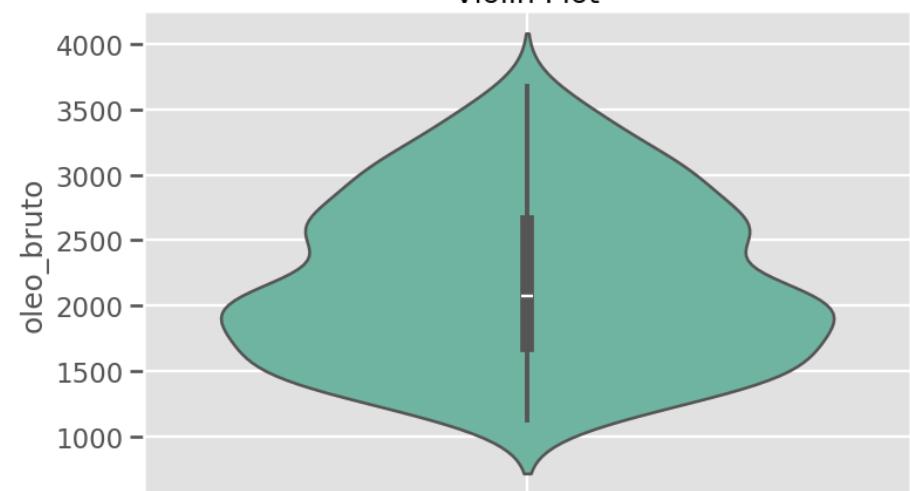
Boxplot



QQ Plot (Verificação de Normalidade)



Violin Plot



Estatísticas descritivas para oleo\_bruto:

0	
<b>Média</b>	2195.160000
<b>Mediana</b>	2077.500000
<b>Desvio Padrão</b>	631.169277
<b>Variância</b>	398374.656589
<b>Mínimo</b>	1121.000000
<b>Máximo</b>	3678.000000
<b>Assimetria</b>	0.343534
<b>Curtose</b>	-0.850090

Teste de Normalidade (Shapiro-Wilk) para oleo\_bruto:

Estatística de teste: 0.9636

Valor p: 0.0000

Conclusão: A distribuição não parece ser normal (rejeita hipótese nula).

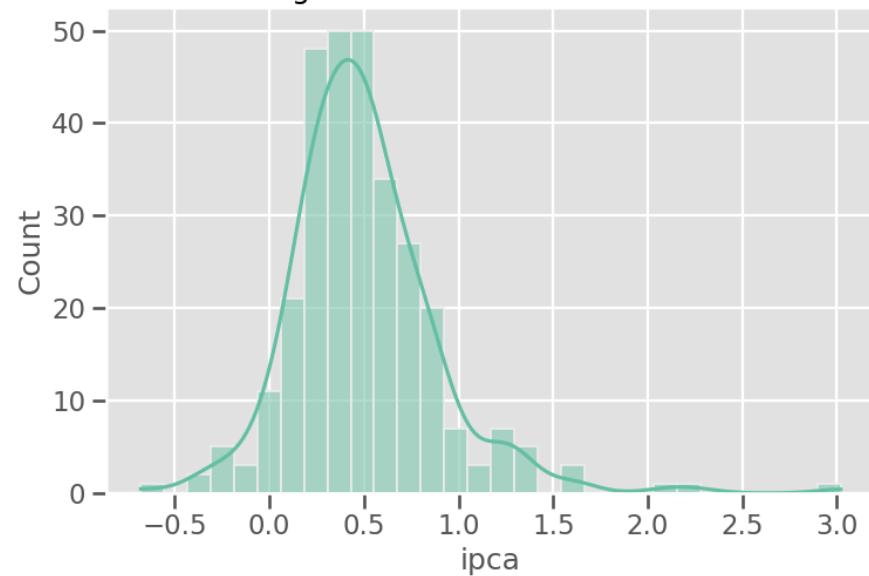
=====

Análise de Distribuição para: ipca

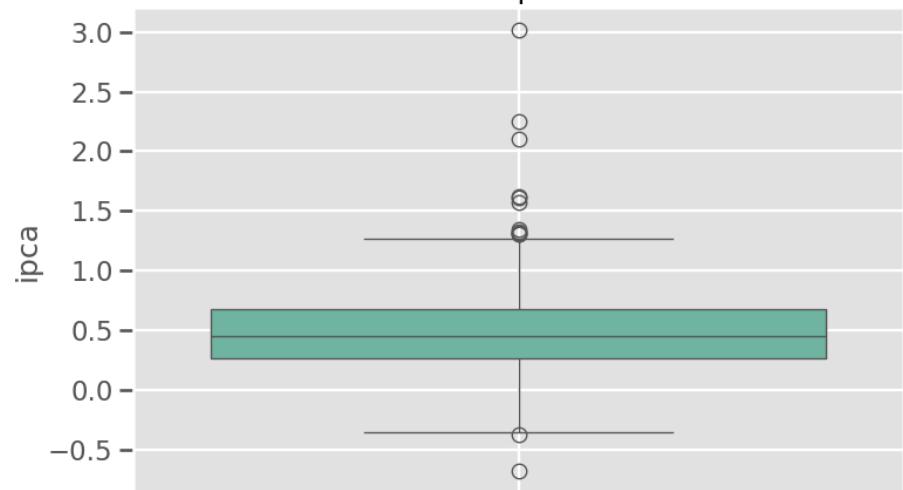
=====

## Análise de Distribuição: ipca

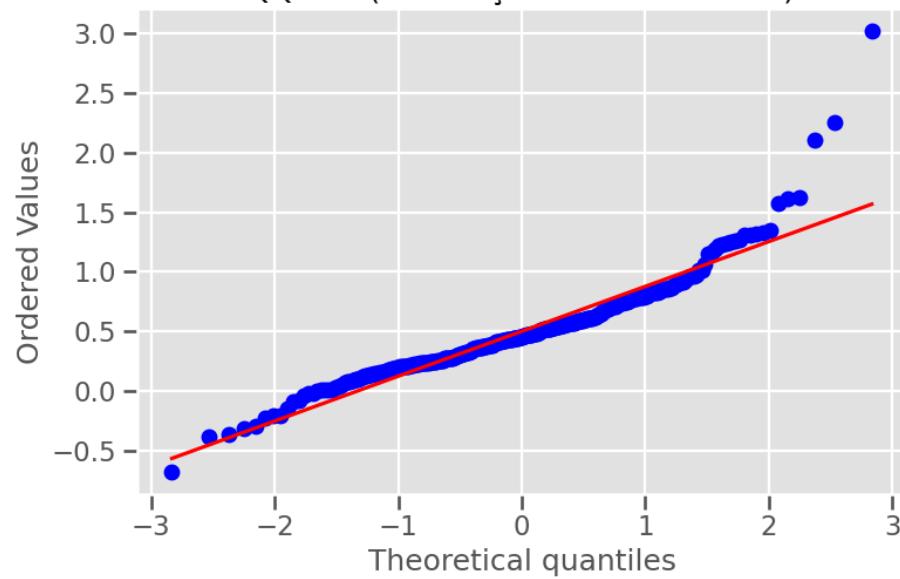
Histograma com Curva de Densidade



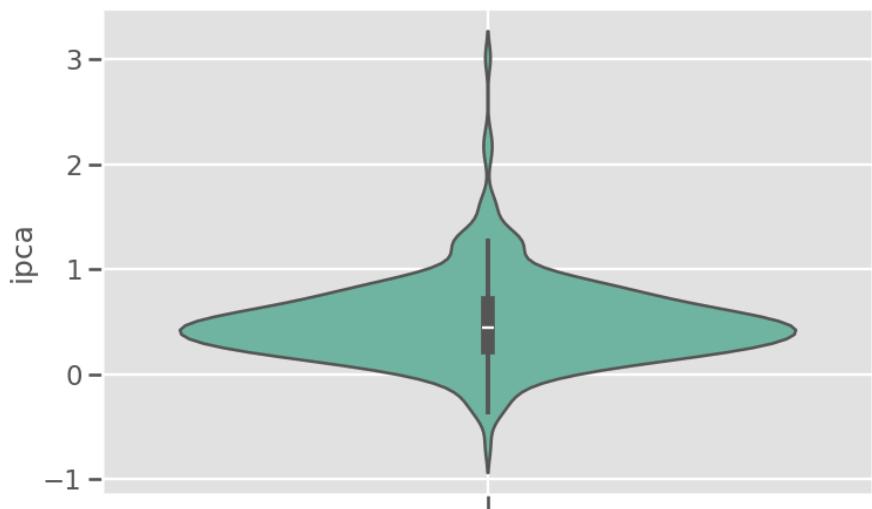
Boxplot



QQ Plot (Verificação de Normalidade)



Violin Plot



Estatísticas descritivas para ipca:

0	
<b>Média</b>	0.501133
<b>Mediana</b>	0.450000
<b>Desvio Padrão</b>	0.394905
<b>Variância</b>	0.155950
<b>Mínimo</b>	-0.680000
<b>Máximo</b>	3.020000
<b>Assimetria</b>	1.565830
<b>Curtose</b>	7.158004

Teste de Normalidade (Shapiro-Wilk) para ipca:

Estatística de teste: 0.9026

Valor p: 0.0000

Conclusão: A distribuição não parece ser normal (rejeita hipótese nula).



## Tabela Resumo: Estatísticas Descritivas e Teste de Normalidade



### Estatísticas Descritivas

Estatística	IPCA	Energia Armazenada (EAR)	Óleo Bruto
Média	0.5011	57.08	2226.83
Mediana	0.4500	56.75	2210.50
Desvio Padrão	0.3949	12.81	865.27
Variância	0.1560	164.06	748694.37
Mínimo	-0.6800	18.66	1121.00
Máximo	3.0200	85.66	3734.00

Estatística	IPCA	Energia Armazenada (EAR)	Óleo Bruto
Assimetria	1.5658	-0.19	0.27
Curtose	7.1580	-0.25	-1.18

## Teste de Normalidade (Shapiro-Wilk)

Estatística	IPCA	Energia Armazenada (EAR)	Óleo Bruto
Estatística do teste	0.9026	0.9870	0.9356
Valor-p	< 0.0001	0.0102	< 0.0001
Conclusão	Não normal	Não normal	Não normal

## Descritivo e Análise dos Resultados

### Média e Mediana

- **Média:** Representa o valor médio das amostras.
- **Mediana:** Indica o valor central quando as amostras estão ordenadas.
- Para o **IPCA** e **Óleo Bruto**, a média e a mediana são próximas, sugerindo uma distribuição mais simétrica. No caso da **Energia Armazenada (EAR)**, a proximidade também indica uma tendência central consistente.

### Desvio Padrão e Variância

- **Desvio Padrão:** Mede a dispersão dos dados em torno da média.
- **Variância:** É o quadrado do desvio padrão e expressa a variabilidade dos dados.
- O **IPCA** apresenta baixa variabilidade (desvio padrão de 0.39), enquanto o **Óleo Bruto** possui alta dispersão (desvio padrão de 865.27), refletindo maiores flutuações nos preços de petróleo.

### Mínimo e Máximo

- Indicam os valores extremos dos dados.
- O **IPCA** varia entre -0.68 e 3.02, mostrando a amplitude de variações inflacionárias.

- A **Energia Armazenada** vai de 18.66 a 85.66, refletindo flutuações nos níveis de armazenamento.
- O **Óleo Bruto** apresenta a maior amplitude (de 1121 a 3734), o que destaca sua volatilidade.

## Assimetria

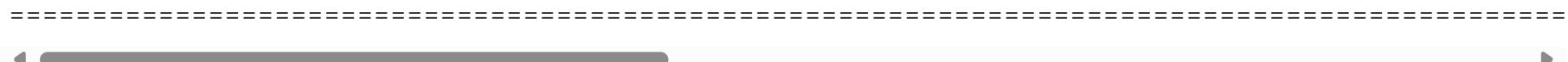
- Mede o grau de distorção da distribuição em relação à simetria.
- O **IPCA** tem assimetria positiva (1.5658), indicando uma cauda longa à direita (mais valores extremos elevados).
- A **EAR** está próxima da simetria (-0.19).
- O **Óleo Bruto** apresenta leve assimetria positiva (0.27), mas não tão marcante.

## Curtose

- Indica o grau de achatamento da distribuição em relação a uma distribuição normal.
- O **IPCA** apresenta alta curtose (7.1580), sugerindo presença de muitos valores extremos (distribuição leptocúrtica).
- A **EAR** e o **Óleo Bruto** têm curtoses negativas, indicando distribuições mais achatadas que a normal (platicúrticas).

## Teste de Normalidade (Shapiro-Wilk)

- Avalia se os dados seguem uma distribuição normal.
- Todos os conjuntos (**IPCA**, **EAR** e **Óleo Bruto**) tiveram **valores-p baixos** (menores que 0.05), levando à rejeição da hipótese de normalidade.
- Isso confirma que as distribuições são **não normais** e podem exigir métodos não paramétricos ou transformações em análises posteriores.



## 4. Teste de Estacionariedade

Para a modelagem econométrica adequada de séries temporais, é fundamental verificar se as variáveis são estacionárias.

Para isso, vamos primeiro entender o conceito de estacionariedade, vejamos:

*"Em linhas gerais um processo estocástico será chamado de estacionário se sua média e variância forem constantes ao longo do tempo eo valor da covariância entre os dois períodos de tempo depender apenas da distância, do intervalo ou da defasagem entre os dois períodos e não o tempo"*

*real ao qual a covariância é computada*" (GUJARATI, Damodar; PORTER, Dawn. Econometria básica. 5. ed. Porto Alegre: AMGH, 2009).

Um processo estocástico, por sua vez é: "*Um processo aleatório ou estocástico é uma coleção de variáveis aleatórias ordenadas no tempo*" (GUJARATI, Damodar; PORTER, Dawn. Econometria básica. 5. ed. Porto Alegre: AMGH, 2009).

A estacionariedade implica que a média, variância e autocorrelação da série são constantes ao longo do tempo. Utilizaremos o **Teste de Dickey-Fuller Aumentado (ADF)** para avaliar a estacionariedade das variáveis do nosso estudo: `ipca`, `ear` e `oleo_bruto`.

O teste ADF tem como hipótese nula a presença de raiz unitária (série não estacionária). Se o p-valor for menor que o nível de significância (geralmente 0,05), rejeitamos a hipótese nula e concluímos que a série é estacionária.

Vamos implementar o teste ADF para cada uma das variáveis e interpretar os resultados:

```
In [10]: # Implementação do Teste de Dickey-Fuller Aumentado (ADF)
from statsmodels.tsa.stattools import adfuller

def test_adf(series, series_name):
    """
    Função para realizar o teste ADF e interpretar os resultados
    """
    print(f"Teste ADF para {series_name}")
    print("-" * 50)

    # Executar o teste ADF
    result = adfuller(series.dropna())

    # Extrair e apresentar resultados
    adf_statistic = result[0]
    p_value = result[1]
    critical_values = result[4]

    # Exibir resultados
    print(f'Estatística ADF: {adf_statistic:.4f}')
    print(f'Valor p: {p_value:.4f}')
    print('Valores críticos:')
    for key, value in critical_values.items():
        print(f'\t{key}: {value:.4f}'')
```

```
# Interpretar resultados
if p_value < 0.05:
    print(f"Conclusão: Rejeitamos a hipótese nula. ",
          f"A série {series_name} é estacionária (não possui raiz unitária).")
else:
    print(f"Conclusão: Não rejeitamos a hipótese nula. ",
          f"A série {series_name} não é estacionária (possui raiz unitária).")
print("\n")
```

```
In [11]: # Aplicar o teste ADF para cada variável
# Certifique-se de que o DataFrame está ordenado por data se for uma série temporal
if 'data' in df.columns:
    df = df.sort_values('data')

# Teste para IPCA
test_adf(df['ipca'], 'IPCA')

# Teste para Energia Armazenada (EAR)
test_adf(df['ear'], 'Energia Armazenada (EAR)')

# Teste para Óleo Bruto
test_adf(df['oleo_bruto'], 'Óleo Bruto')
```

Teste ADF para IPCA

Estatística ADF: -8.3602

Valor p: 0.0000

Valores críticos:

1%: -3.4524

5%: -2.8713

10%: -2.5719

Conclusão: Rejeitamos a hipótese nula. A série IPCA é estacionária (não possui raiz unitária).

Teste ADF para Energia Armazenada (EAR)

Estatística ADF: -2.5510

Valor p: 0.1036

Valores críticos:

1%: -3.4533

5%: -2.8717

10%: -2.5722

Conclusão: Não rejeitamos a hipótese nula. A série Energia Armazenada (EAR) não é estacionária (possui raiz unitária).

Teste ADF para Óleo Bruto

Estatística ADF: -0.4786

Valor p: 0.8961

Valores críticos:

1%: -3.4526

5%: -2.8713

10%: -2.5720

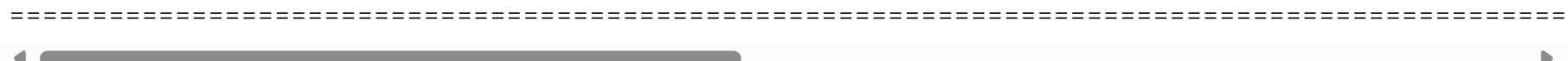
Conclusão: Não rejeitamos a hipótese nula. A série Óleo Bruto não é estacionária (possui raiz unitária).



## Tabela Resumo: Teste de Estacionariedade (ADF)

Variável	Estatística ADF	Valor-p	Valores Críticos (1%, 5%, 10%)	Conclusão	Requer Tratamento?
<b>IPCA</b>	-8.3602	0.0000	-3.4524, -2.8713, -2.5719	✓ Rejeita $H_0 \rightarrow$ Série é <b>estacionária</b>	Não
<b>Energia Armazenada (EAR)</b>	-2.5510	0.1036	-3.4533, -2.8717, -2.5722	✗ Não rejeita $H_0 \rightarrow$ Série <b>não é estacionária</b>	Sim
<b>Óleo Bruto</b>	-0.4786	0.8961	-3.4526, -2.8713, -2.5720	✗ Não rejeita $H_0 \rightarrow$ Série <b>não é estacionária</b>	Sim

Como aqui demonstrado, a variável `ipca` é estacionária, enquanto `ear` e `oleo_bruto` não são. Portanto, deve ocorrer o tratamento das variáveis `ear` e `oleo_bruto` para torná-las estacionárias, pois "Se uma série é não estacionária (...) ela é chamada de série temporal não estacionária (...). Em outras palavras, uma série temporal não estacionária terá uma média que varia com o tempo, ou uma variância que varia com o tempo, ou, ainda, ambas. (...) se uma série temporal é não estacionária, podemos estudar seu comportamento apenas pelo período em consideração. Cada conjunto de dados de série temporal, portanto, será específico a cada episódio. Como consequência, não é possível generalizá-lo para outros períodos. Sendo assim, para o propósito de previsão, tal série temporal (não estacionária) pode ser de pouco valor prático" (GUJARATI, Damodar; PORTER, Dawn. Econometria básica. 5. ed. Porto Alegre: AMGH, 2009).



#### .4.1 Análise Visual da Estacionariedade: Gráficos de Séries Temporais

Para complementar a análise formal do teste ADF, vamos visualizar graficamente as séries temporais. Isso nos permitirá observar visualmente se as séries apresentam características de não-estacionariedade, como tendências, mudanças de variância ao longo do tempo ou padrões sazonais persistentes. O que se faz somente para fins ilustrativos, uma vez que o teste ADF já foi realizado e os resultados foram apresentados anteriormente.

In [12]:

```
# Preparando os dados para visualização de séries temporais
# Verificar se o DataFrame tem uma coluna de data para o índice temporal
if 'data' in df.columns:
    # Converter a coluna de data para índice do DataFrame se ainda não for
    time_series_df = df.set_index('data') if not isinstance(df.index, pd.DatetimeIndex) else df.copy()
```

```

else:
    # Se não tiver coluna de data, verificar se o índice já é do tipo datetime
    time_series_df = df.copy()
    if not isinstance(df.index, pd.DatetimeIndex):
        print("Aviso: Não foi encontrada uma coluna de data para criar série temporal")
        # Criar um índice temporal fictício começando de 01/01/2000 com frequência mensal
        time_series_df.index = pd.date_range(start='2000-01-01', periods=len(df), freq='M')

# Função para plotar séries temporais com análise de estacionariedade
def plot_time_series_stationarity(dataframe, column_name):
    """
    Função para criar gráficos de série temporal com análise de estacionariedade
    - Gráfico da série original
    - Média móvel
    - Desvio padrão móvel
    """
    plt.figure(figsize=(14, 10))

    # Criar uma figura com 3 subplots verticalmente empilhados
    fig, (ax1, ax2, ax3) = plt.subplots(3, 1, figsize=(14, 12), sharex=True)

    # Plot 1: Série temporal original
    series = dataframe[column_name].dropna()
    ax1.plot(series.index, series.values, label='Série Original')
    ax1.set_title(f'Série Temporal: {column_name}', fontsize=14)
    ax1.set_ylabel('Valor')
    ax1.legend(loc='upper left', fontsize=8, framealpha=0.7)
    ax1.grid(True)

    # Plot 2: Média Móvel (janela de 12 meses e 24 meses)
    rolling_mean_12 = series.rolling(window=12).mean()
    rolling_mean_24 = series.rolling(window=24).mean()
    ax2.plot(series.index, series.values, label='Série Original', alpha=0.5, color='gray')
    ax2.plot(rolling_mean_12.index, rolling_mean_12.values, label='Média Móvel (12 meses)', color='red')
    ax2.plot(rolling_mean_24.index, rolling_mean_24.values, label='Média Móvel (24 meses)', color='green')
    ax2.set_title(f'Análise de Tendência (Média Móvel): {column_name}', fontsize=14)
    ax2.set_ylabel('Valor')
    ax2.legend(loc='upper left', fontsize=8, framealpha=0.7, ncol=2)
    ax2.grid(True)

    # Plot 3: Desvio Padrão Móvel (janela de 12 meses)

```

```

rolling_std_12 = series.rolling(window=12).std()
rolling_std_24 = series.rolling(window=24).std()
ax3.plot(rolling_std_12.index, rolling_std_12.values, label='Desvio Padrão Móvel (12 meses)', color='purple')
ax3.plot(rolling_std_24.index, rolling_std_24.values, label='Desvio Padrão Móvel (24 meses)', color='orange')
ax3.set_title(f'Análise de Variância (Desvio Padrão Móvel): {column_name}', fontsize=14)
ax3.set_xlabel('Data')
ax3.set_ylabel('Desvio Padrão')
ax3.legend(loc='upper left', fontsize=8, framealpha=0.7, ncol=2)
ax3.grid(True)

plt.tight_layout()
plt.show()

# Estatísticas básicas da série
media = series.mean()
mediana = series.median()
desvio_padrao = series.std()

print(f"\nEstatísticas básicas para {column_name}:")
print("-" * 50)
print(f"Média: {media:.4f}")
print(f"Mediana: {mediana:.4f}")
print(f"Desvio Padrão: {desvio_padrao:.4f}")
print("-" * 50)

# Avaliação da estacionariedade com base nos gráficos
print(f"\nAnálise visual de estacionariedade para {column_name}:")
print("-" * 70)

# Analisar a presença de tendência (média móvel)
if rolling_mean_12.max() - rolling_mean_12.min() > 0.1 * series.mean():
    print("Tendência: A série parece apresentar tendência (média não constante ao longo do tempo)")
    print(f"Variação na média móvel: {rolling_mean_12.max() - rolling_mean_12.min():.4f}")
else:
    print("Tendência: A série não apresenta tendência significativa (média relativamente constante)")

# Analisar a estabilidade da variância (desvio padrão móvel)
if rolling_std_12.max() / (rolling_std_12.min() + 1e-10) > 1.5: # Evita divisão por zero
    print("Variância: A série parece apresentar heterocedasticidade (variância não constante)")
    print(f"Razão máx/mín do desvio padrão: {rolling_std_12.max() / (rolling_std_12.min() + 1e-10):.4f}")
else:

```

```
print("Variância: A série parece ter variância relativamente constante (homocedasticidade)")

# Conclusão preliminar
if (rolling_mean_12.max() - rolling_mean_12.min() > 0.1 * series.mean()) or (rolling_std_12.max() / (rolling_std_12.min()))
    print("\nConclusão preliminar: A série apresenta características de não-estacionariedade")
else:
    print("\nConclusão preliminar: A série apresenta características de estacionariedade")
print("-" * 70)
```

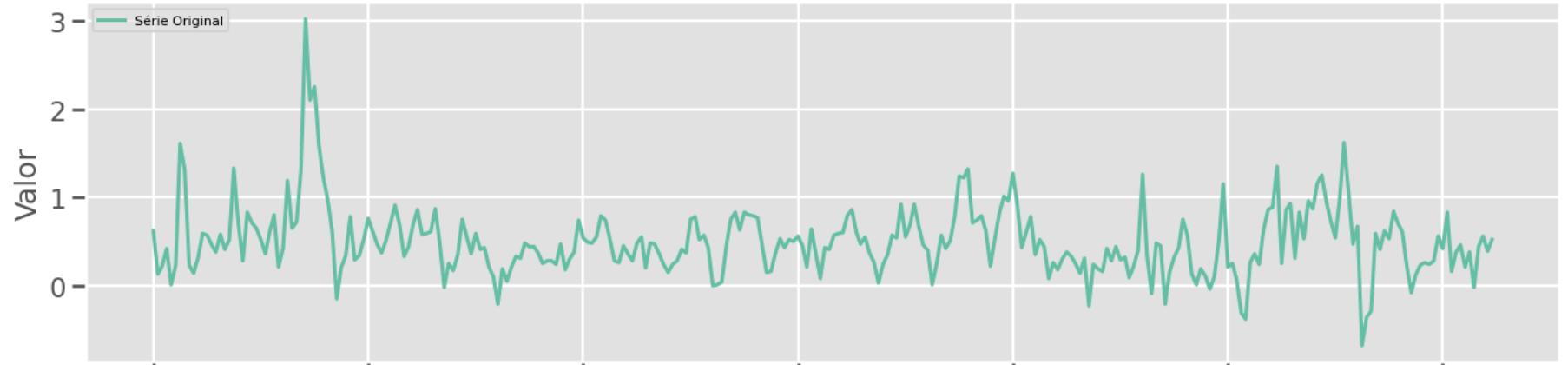
```
In [13]: # Aplicar a análise de série temporal para cada variável
variables_to_analyze = ['ipca', 'ear', 'oleo_bruto']

for var in variables_to_analyze:
    print(f"\n{'=' * 80}")
    print(f"Análise de Série Temporal para: {var}")
    print(f"{'=' * 80}")
    plot_time_series_stationarity(time_series_df, var)
```

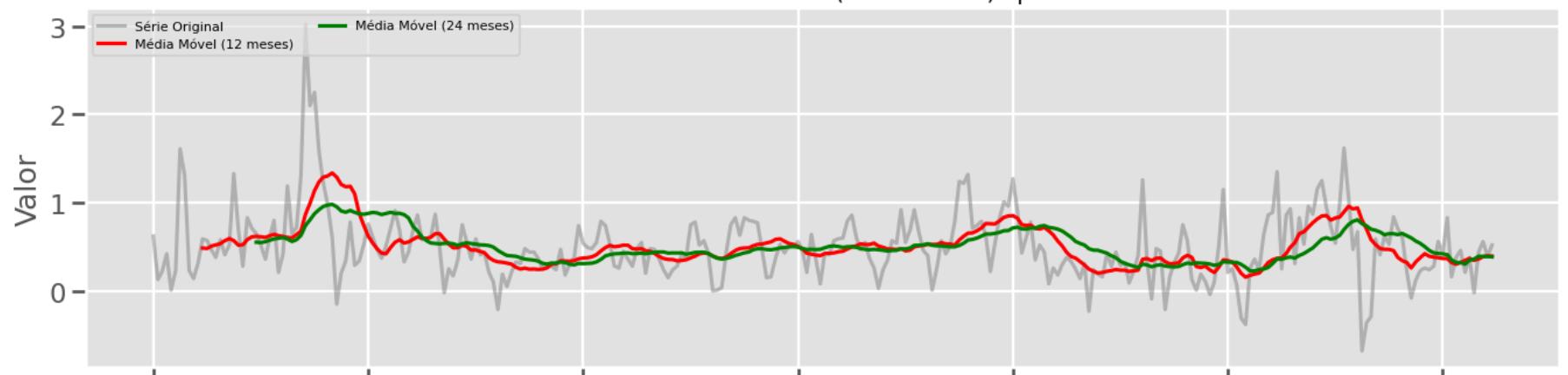
```
=====
Análise de Série Temporal para: ipca
=====
```

```
<Figure size 1400x1000 with 0 Axes>
```

Série Temporal: ipca

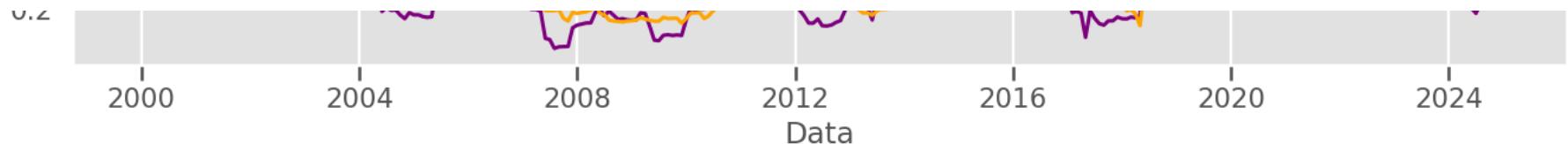


Análise de Tendência (Média Móvel): ipca



Análise de Variância (Desvio Padrão Móvel): ipca





Estatísticas básicas para ipca:

Média: 0.5011  
Mediana: 0.4500  
Desvio Padrão: 0.3949

Análise visual de estacionariedade para ipca:

Tendência: A série parece apresentar tendência (média não constante ao longo do tempo)

Variação na média móvel: 1.1808

Variância: A série parece apresentar heterocedasticidade (variância não constante)

Razão máx/mín do desvio padrão: 10.0771

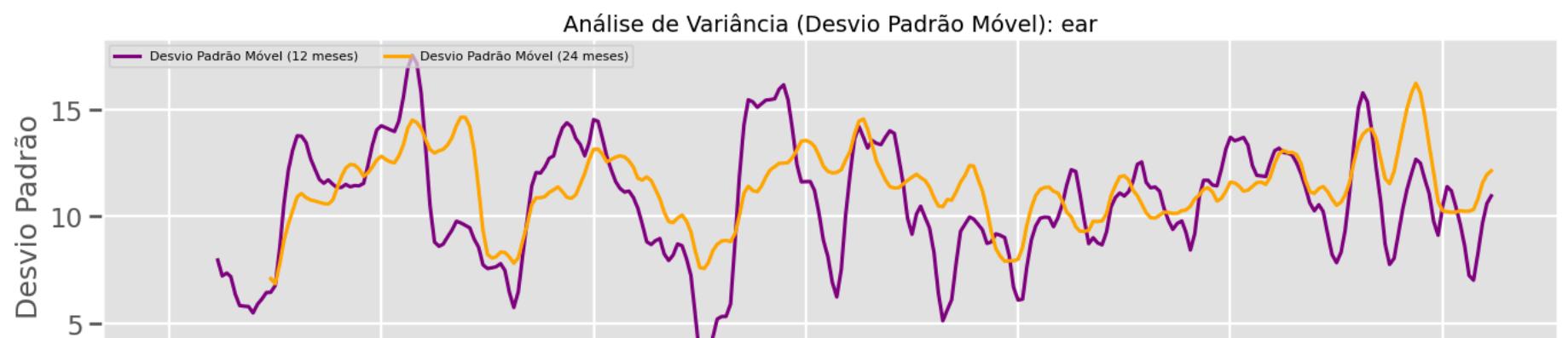
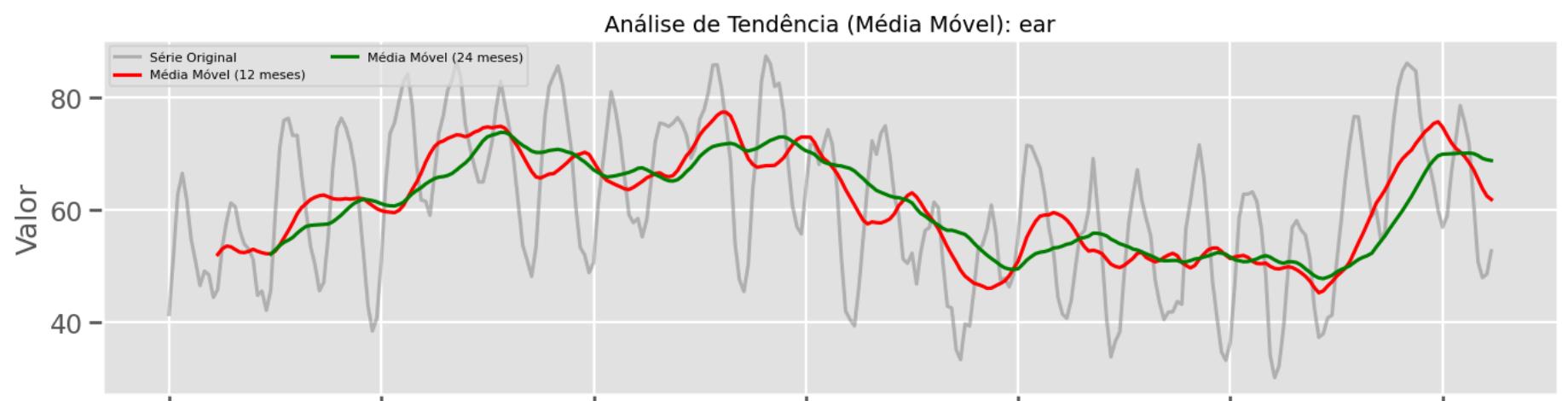
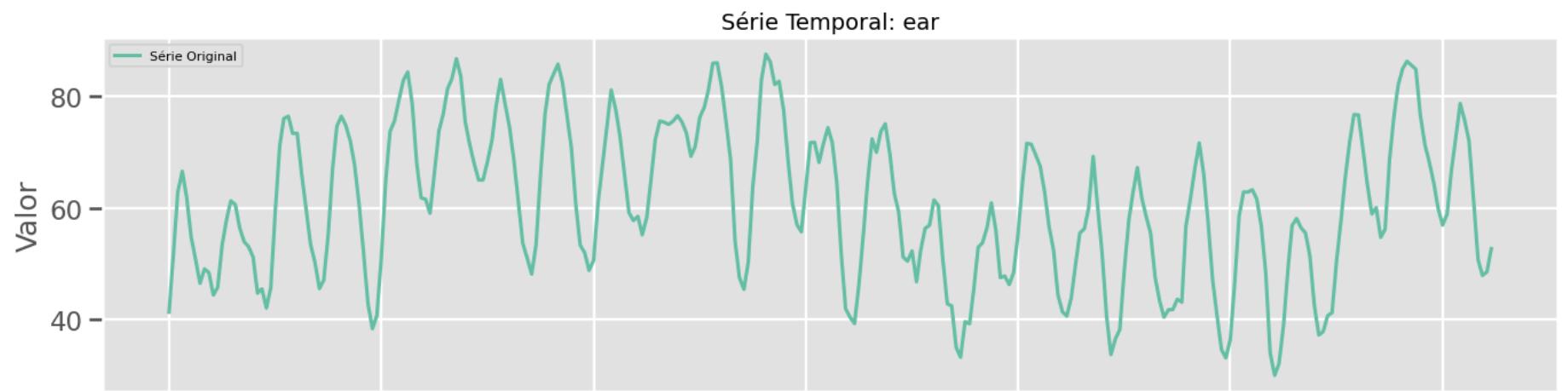
Conclusão preliminar: A série apresenta características de não-estacionariedade

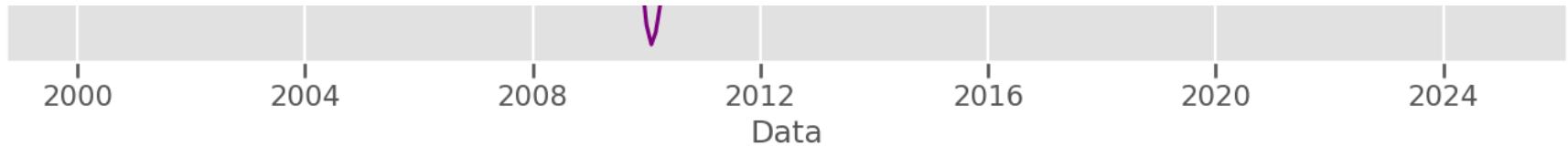
=====

Análise de Série Temporal para: ear

=====

<Figure size 1400x1000 with 0 Axes>





Estatísticas básicas para ear:

Média: 60.7518  
Mediana: 60.5666  
Desvio Padrão: 13.6448

Análise visual de estacionariedade para ear:

Tendência: A série parece apresentar tendência (média não constante ao longo do tempo)  
Variação na média móvel: 32.2463  
Variância: A série parece apresentar heterocedasticidade (variância não constante)  
Razão máx/mín do desvio padrão: 7.0249

Conclusão preliminar: A série apresenta características de não-estacionariedade

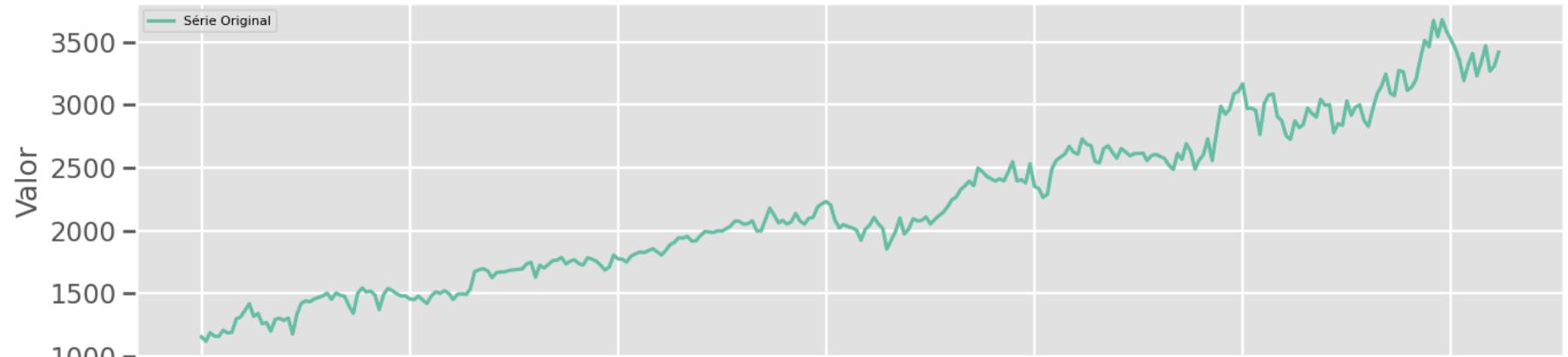
=====

Análise de Série Temporal para: oleo\_bruto

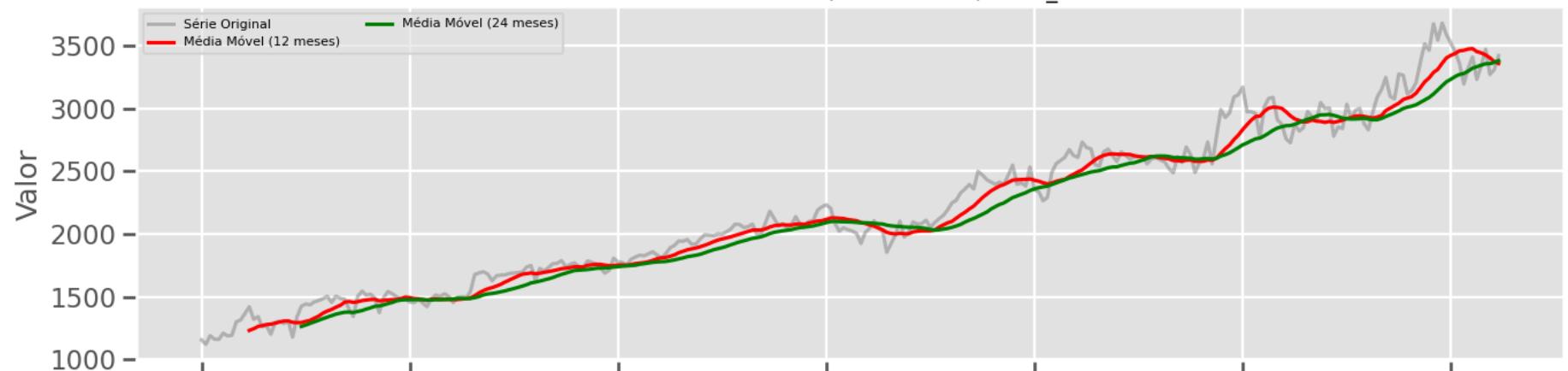
=====

<Figure size 1400x1000 with 0 Axes>

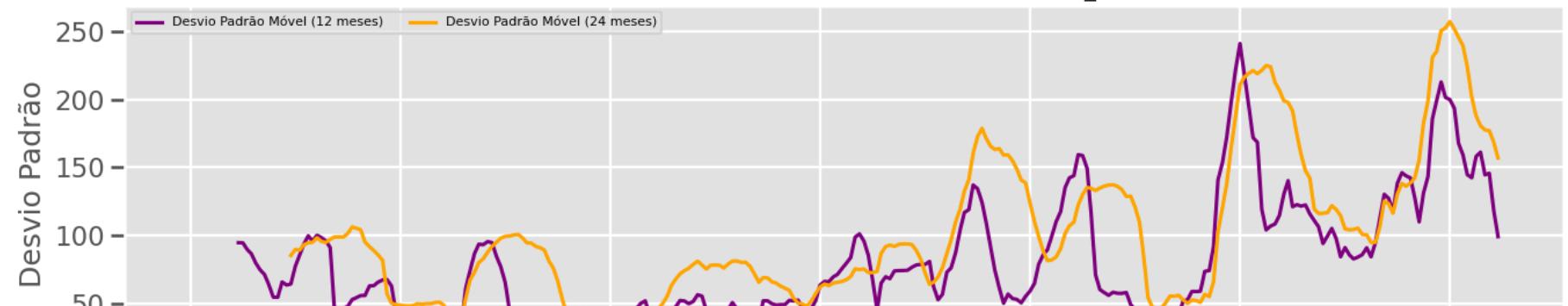
Série Temporal: oleo\_bruto

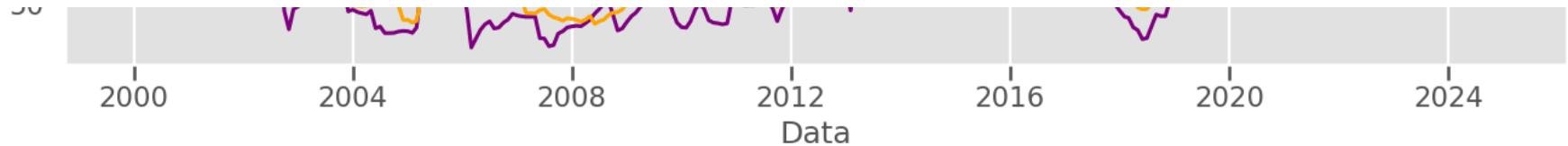


Análise de Tendência (Média Móvel): oleo\_bruto



Análise de Variância (Desvio Padrão Móvel): oleo\_bruto





Estatísticas básicas para oleo\_bruto:

Média: 2195.1600  
 Mediana: 2077.5000  
 Desvio Padrão: 631.1693

Análise visual de estacionariedade para oleo\_bruto:

Tendência: A série parece apresentar tendência (média não constante ao longo do tempo)

Variação na média móvel: 2243.8333

Variância: A série parece apresentar heterocedasticidade (variância não constante)

Razão máx/mín do desvio padrão: 12.5054

Conclusão preliminar: A série apresenta características de não-estacionariedade



#### Tabela Resumo Final – Estacionariedade (Teste ADF + Análise Visual)

Variável	Tendência Aparente	Heterocedasticidade (Razão máx/mín)	Estatística ADF	Valor-p	Conclusão ADF	Conclusão Visual	Requer Tratamento?
<b>IPCA</b>	Sim (variação média: 1.18)	Sim (10.08)	-8.3602	0.0000	<input checked="" type="checkbox"/> Série estacionária	<input type="checkbox"/> Não estacionária visualmente	<b>Não</b>
<b>EAR</b>	Sim (variação média: 32.25)	Sim (7.02)	-2.5510	0.1036	<input type="checkbox"/> Série não estacionária	<input type="checkbox"/> Não estacionária visualmente	<b>Sim</b>
<b>Oleo Bruto</b>	Sim (variação média: 2243.83)	Sim (12.51)	-0.4786	0.8961	<input type="checkbox"/> Série não estacionária	<input type="checkbox"/> Não estacionária visualmente	<b>Sim</b>



#### .4.2 Tratamento de Variáveis

Nesta seção, realizaremos o tratamento das variáveis que apresentaram características de não-estacionariedade ou outros problemas identificados nas análises anteriores. O objetivo é preparar as séries temporais para a modelagem econométrica, garantindo que atendam aos requisitos necessários, como estacionariedade e estabilidade de variância.

As transformações recomendadas serão aplicadas conforme a tabela resumo apresentada anteriormente. Após o tratamento, as séries serão reavaliadas para confirmar a eficácia das transformações realizadas.



## Tabela Resumo: Transformações Recomendadas para Estacionariedade

Variável	Transformação Recomendada	Objetivo Principal	Justificativa Técnica
IPCA	Nenhuma	Já é estacionária	O teste ADF rejeitou a hipótese de raiz unitária; série pode ser usada no formato atual
EAR	1ª diferença ( $x_t - x_{(t-1)}$ )	Remover tendência e estabilizar a média	Série apresenta tendência e leve heterocedasticidade; valores percentuais não exigem log
Óleo Bruto	Log-diferença ( $\log(x_t) - \log(x_{(t-1)})$ )	Reducir escala e estabilizar variância	Série tem escala alta e variância crescente; transformação ajuda a lidar com isso e permite interpretação percentual

In [14]:

```
# Aplicando as transformações recomendadas para alcançar estacionariedade
# Criando uma cópia do DataFrame para não modificar o original
df_transformed = df.copy()

# 1. Aplicando a primeira diferença ( $x_t - x_{(t-1)}$ ) na variável EAR
df_transformed['ear_diff'] = df_transformed['ear'].diff()

# 2. Aplicando a log-diferença ( $\log(x_t) - \log(x_{(t-1)})$ ) na variável Óleo Bruto
df_transformed['oleo_bruto_log_diff'] = np.log(df_transformed['oleo_bruto']).diff()

# Removendo as primeiras linhas que terão valores NaN devido à diferenciação
df_transformed = df_transformed.dropna(subset=['ear_diff', 'oleo_bruto_log_diff'])

# Visualizando as primeiras linhas do DataFrame transformado
print("Primeiras linhas do DataFrame com as variáveis transformadas:")
display(df_transformed[['ipca', 'ear', 'ear_diff', 'oleo_bruto', 'oleo_bruto_log_diff']].head())
```

```
# Visualizando estatísticas descritivas das variáveis transformadas
print("\nEstatísticas descritivas das variáveis transformadas:")
display(df_transformed[['ear_diff', 'oleo_bruto_log_diff']].describe().T)
```

Primeiras linhas do DataFrame com as variáveis transformadas:

	ipca	ear	ear_diff	oleo_bruto	oleo_bruto_log_diff
1	0.13	52.108976	10.764014	1121	-0.029879
2	0.22	62.794431	10.685455	1189	0.058891
3	0.42	66.530108	3.735677	1161	-0.023831
4	0.01	61.894345	-4.635762	1159	-0.001724
5	0.23	54.794167	-7.100179	1209	0.042236

Estatísticas descritivas das variáveis transformadas:

	count	mean	std	min	25%	50%	75%	max
<b>ear_diff</b>	299.0	0.037953	5.903282	-14.593406	-4.305174	-0.20190	4.569417	14.046825
<b>oleo_bruto_log_diff</b>	299.0	0.003630	0.033736	-0.102385	-0.014679	0.00292	0.021242	0.128105

In [15]: # Realizando o teste ADF nas variáveis transformadas

```
print("\n" + "="*80)
print("TESTES DE ESTACIONARIEDADE NAS SÉRIES TRANSFORMADAS")
print("="*80)

# Teste para EAR diferenciada
test_adf(df_transformed['ear_diff'], 'Energia Armazenada (EAR) - Primeira Diferença')

# Teste para Óleo Bruto com log-diferença
test_adf(df_transformed['oleo_bruto_log_diff'], 'Óleo Bruto - Log-Diferença')
```

=====

TESTES DE ESTACIONARIEDADE NAS SÉRIES TRANSFORMADAS

=====

Teste ADF para Energia Armazenada (EAR) - Primeira Diferença

-----

Estatística ADF: -5.7136

Valor p: 0.0000

Valores críticos:

1%: -3.4533

5%: -2.8717

10%: -2.5722

Conclusão: Rejeitamos a hipótese nula. A série Energia Armazenada (EAR) - Primeira Diferença é estacionária (não possui raiz unitária).

Teste ADF para Óleo Bruto - Log-Diferença

-----

Estatística ADF: -11.1966

Valor p: 0.0000

Valores críticos:

1%: -3.4527

5%: -2.8714

10%: -2.5720

Conclusão: Rejeitamos a hipótese nula. A série Óleo Bruto - Log-Diferença é estacionária (não possui raiz unitária).

```
In [16]: # Análise visual de estacionariedade para as variáveis transformadas
# Criando um DataFrame de séries temporais com as variáveis transformadas
time_series_transformed_df = df_transformed.copy()
if 'data' in time_series_transformed_df.columns:
    time_series_transformed_df = time_series_transformed_df.set_index('data')

# Aplicar a análise visual para cada variável transformada
print("\n" + "*80)
print("ANÁLISE VISUAL DE ESTACIONARIEDADE - SÉRIES TRANSFORMADAS")
print("*80)

# Análise para EAR diferenciada
print("\n" + "*80)
print("Análise de Série Temporal para: EAR - Primeira Diferença")
```

```
print("=*80")
plot_time_series_stationarity(time_series_transformed_df, 'ear_diff')

# Análise para Óleo Bruto com Log-diferença
print("\n" + "*80)
print("Análise de Série Temporal para: Óleo Bruto - Log-Diferença")
print("*80)
plot_time_series_stationarity(time_series_transformed_df, 'oleo_bruto_log_diff')
```

=====

ANÁLISE VISUAL DE ESTACIONARIEDADE - SÉRIES TRANSFORMADAS

=====

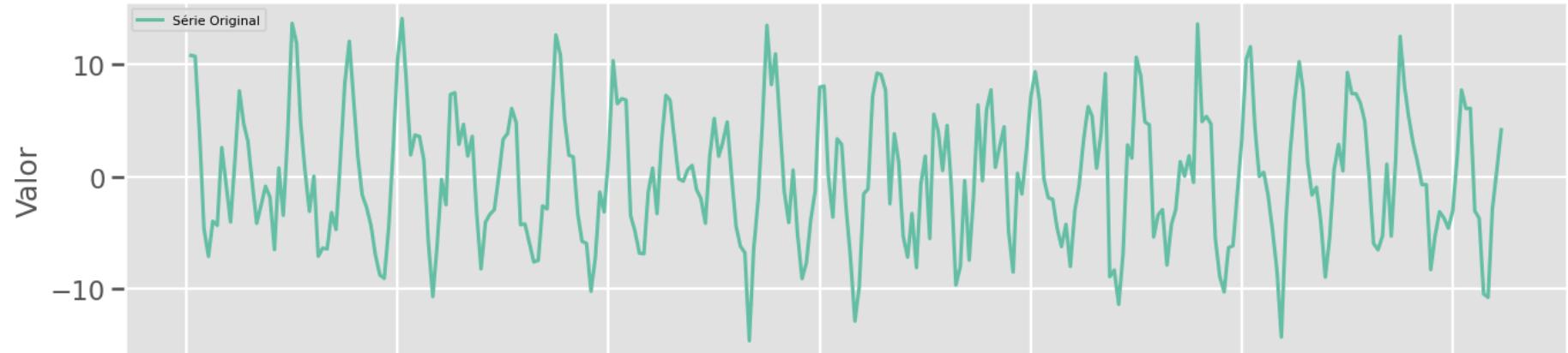
=====

Análise de Série Temporal para: EAR - Primeira Diferença

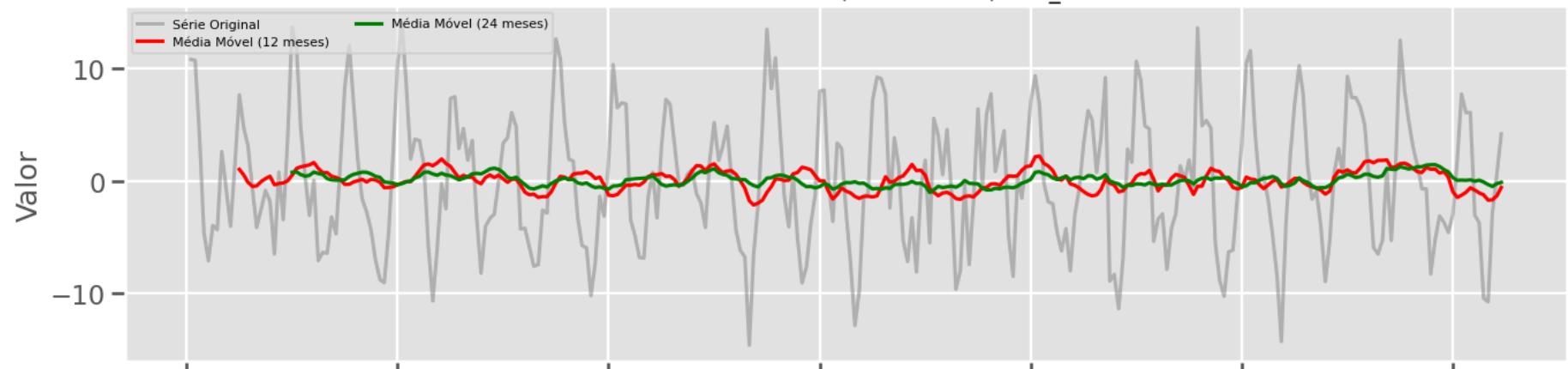
=====

<Figure size 1400x1000 with 0 Axes>

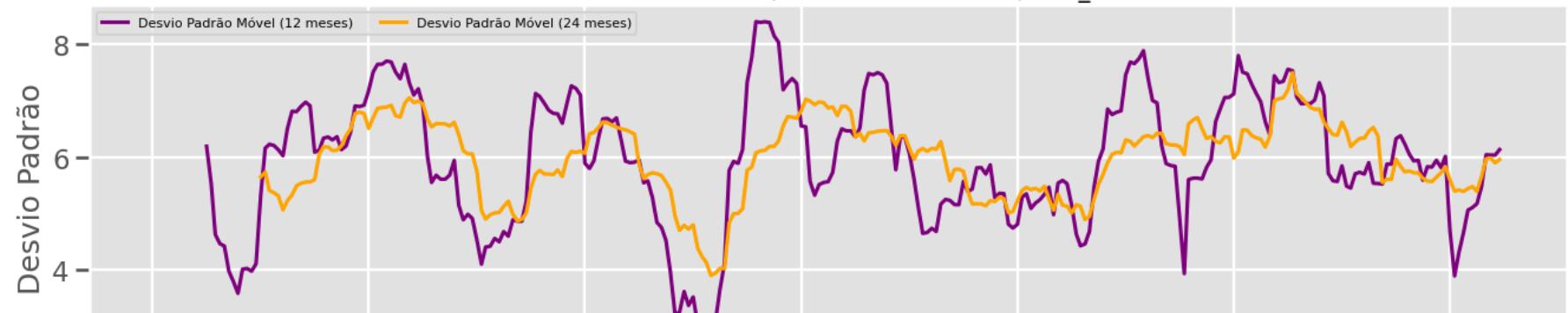
Série Temporal: ear\_diff



Análise de Tendência (Média Móvel): ear\_diff



Análise de Variância (Desvio Padrão Móvel): ear\_diff





Estatísticas básicas para ear\_diff:

Média: 0.0380  
Mediana: -0.2019  
Desvio Padrão: 5.9033

Análise visual de estacionariedade para ear\_diff:

Tendência: A série parece apresentar tendência (média não constante ao longo do tempo)  
Variação na média móvel: 4.3491  
Variância: A série parece apresentar heterocedasticidade (variância não constante)  
Razão máx/mín do desvio padrão: 3.3183

Conclusão preliminar: A série apresenta características de não-estacionariedade

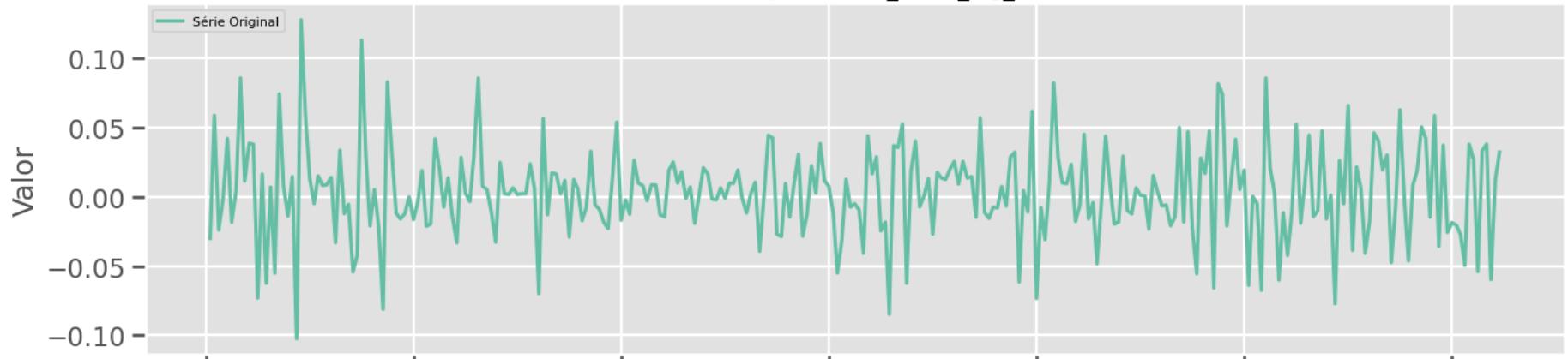
=====

Análise de Série Temporal para: Óleo Bruto - Log-Diferença

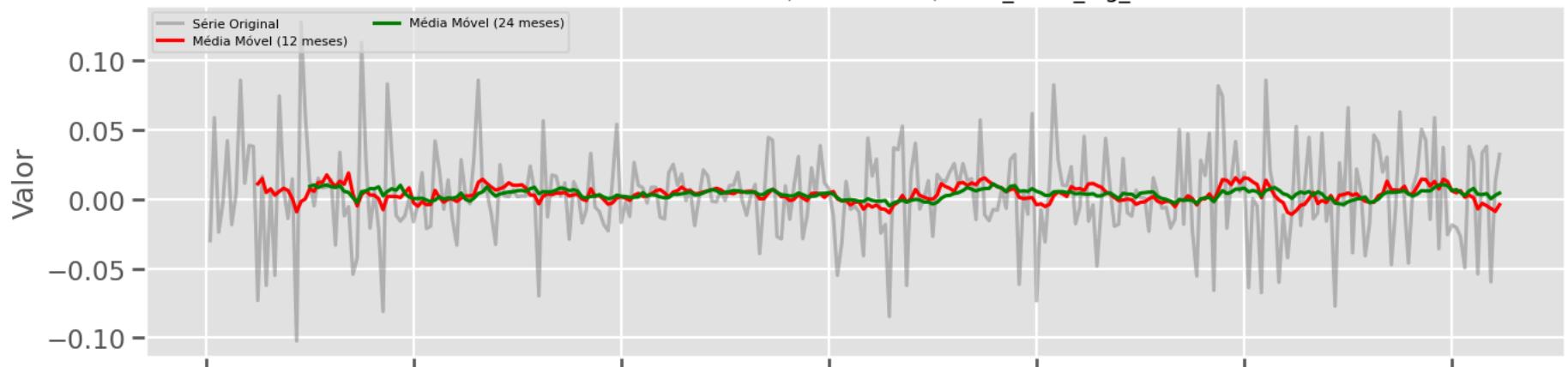
=====

<Figure size 1400x1000 with 0 Axes>

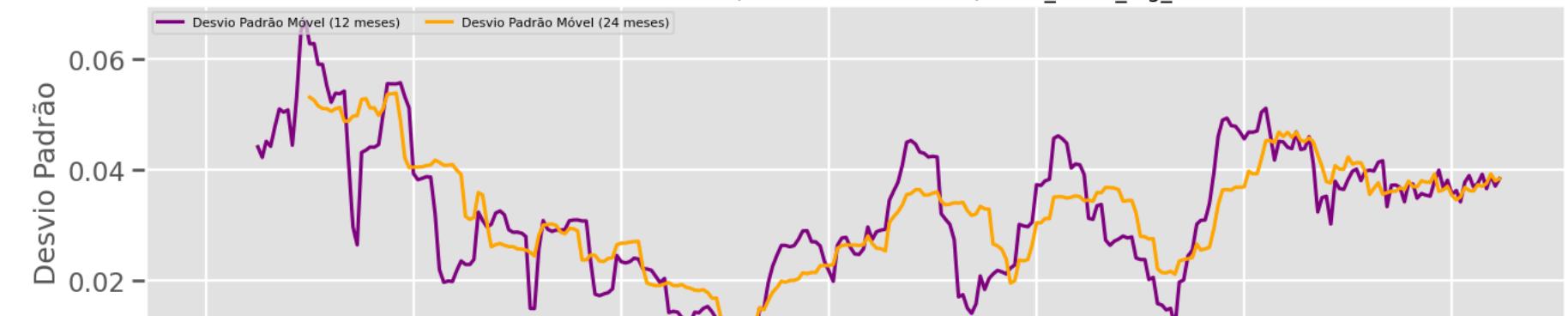
Série Temporal: oleo\_bruto\_log\_diff

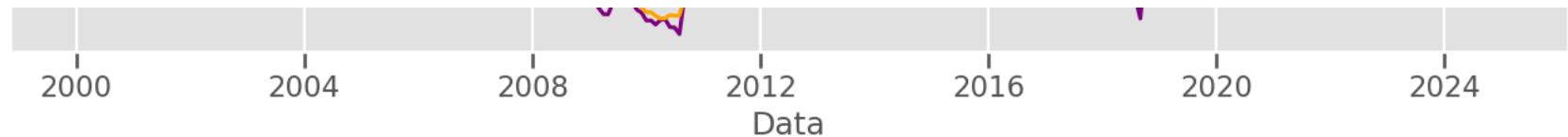


Análise de Tendência (Média Móvel): oleo\_bruto\_log\_diff



Análise de Variância (Desvio Padrão Móvel): oleo\_bruto\_log\_diff





Estatísticas básicas para oleo\_bruto\_log\_diff:

Média: 0.0036  
Mediana: 0.0029  
Desvio Padrão: 0.0337

Análise visual de estacionariedade para oleo\_bruto\_log\_diff:

Tendência: A série parece apresentar tendência (média não constante ao longo do tempo)  
Variação na média móvel: 0.0298  
Variância: A série parece apresentar heterocedasticidade (variância não constante)  
Razão máx/mín do desvio padrão: 7.5026

Conclusão preliminar: A série apresenta características de não-estacionariedade



### Tabela Resumo – Estacionariedade das Séries Transformadas (Visual + ADF)

O teste ADF tem como hipótese nula a presença de raiz unitária (série não estacionária). Se o p-valor for menor que o nível de significância (geralmente 0,05), rejeitamos a hipótese nula e concluímos que a série é estacionária.

Variável Transformada	Transformação Aplicada	Estatística ADF	Valor-p	Conclusão ADF	Variação na Média Móvel	Razão Máx/Mín do Desvio Padrão	Conclusão Visual
EAR (Energia Armazenada)	1ª diferença ( $x_t - x_{(t-1)}$ )	-5.7136	0.0000	<input checked="" type="checkbox"/> Série estacionária	4.3491	3.3183	<span style="color: red;">✗</span> Aparência de não-estacionária
Óleo Bruto	Log-diferença ( $\log(x_t) - \log(x_{(t-1)})$ )	-11.1966	0.0000	<input checked="" type="checkbox"/> Série estacionária	0.0298	7.5026	<span style="color: red;">✗</span> Aparência de não-estacionária

## 5. Análise de Correlação entre IPCA, EAR\_Diff e Óleo Bruto (Log-Diferença)

Nesta seção, realizaremos uma análise de correlação entre as variáveis transformadas `ear_diff` (diferença da energia armazenada), `oleo_bruto_log_diff` (log-diferença da produção de óleo bruto) e o índice de preços ao consumidor amplo (`ipca`).

O objetivo é identificar a força e a direção das relações lineares entre essas variáveis, fornecendo insights iniciais sobre possíveis interdependências que podem ser exploradas em modelagens econométricas futuras. Utilizaremos métricas como o coeficiente de correlação de Pearson e visualizações gráficas para apoiar a análise.

### Análise de Correlação Cruzada (CCF) entre IPCA e Variáveis Transformadas

A correlação cruzada (Cross-Correlation Function - CCF) é uma técnica estatística que mede a relação entre duas séries temporais em diferentes defasagens. Esta análise nos permite identificar se mudanças em uma variável precedem, coincidem ou seguem mudanças em outra variável, e com qual intensidade ocorre essa relação.

Abaixo, visualizaremos os gráficos de CCF entre o IPCA e as variáveis transformadas (`ear_diff` e `oleo_bruto_log_diff`), além de apresentarmos uma tabela com os valores específicos das correlações em diferentes defasagens.

```
In [17]: # Análise de Correlação Cruzada (CCF) usando statsmodels.graphics.tsaplots
```

```
# Configuração para melhorar a visualização
plt.style.use('seaborn-v0_8')
plt.rcParams['figure.figsize'] = (14, 10)

# Definindo o número máximo de lags para análise
lags = 6

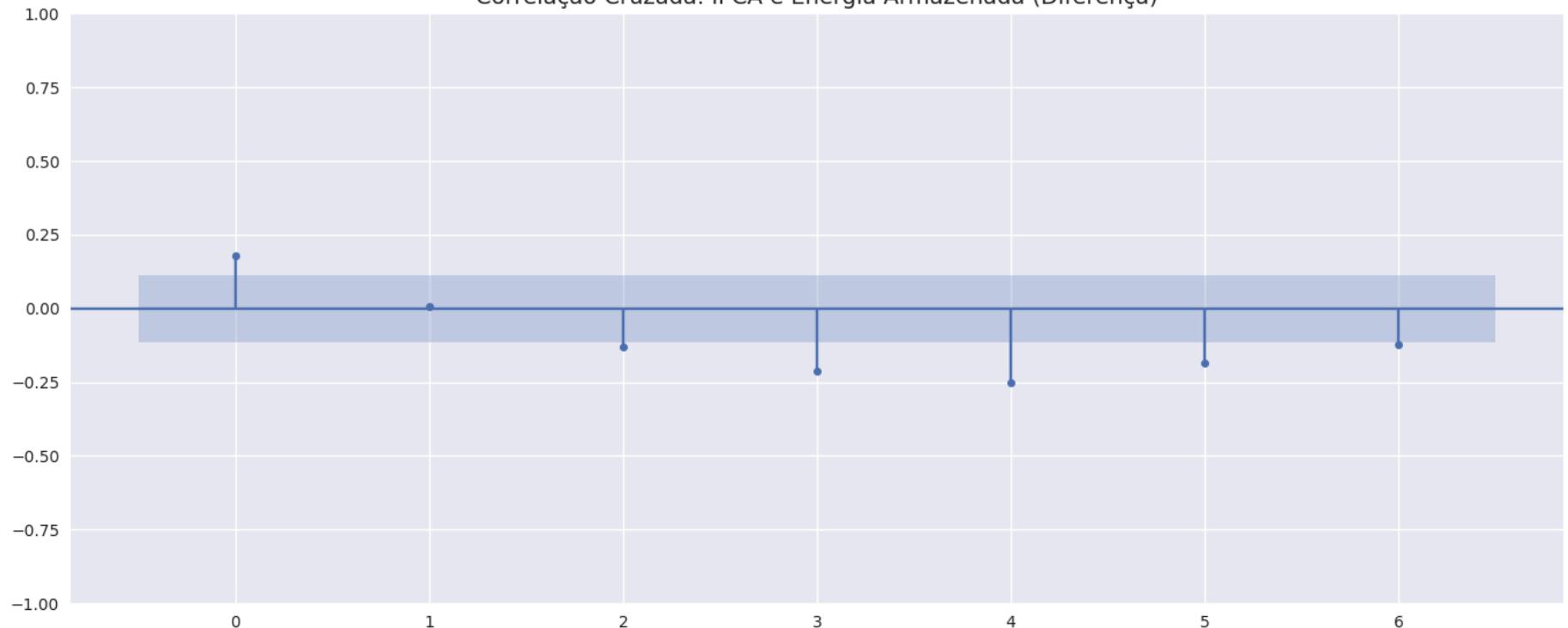
# Criando uma figura com 2 subplots para CCF
fig, axes = plt.subplots(2, 1, figsize=(14, 12))

# Gráfico CCF entre IPCA e EAR (diferença)
plot_ccf(df_transformed['ipca'], df_transformed['ear_diff'], lags=lags, ax=axes[0])
axes[0].set_title('Correlação Cruzada: IPCA e Energia Armazenada (Diferença)', fontsize=14)
axes[0].grid(True)

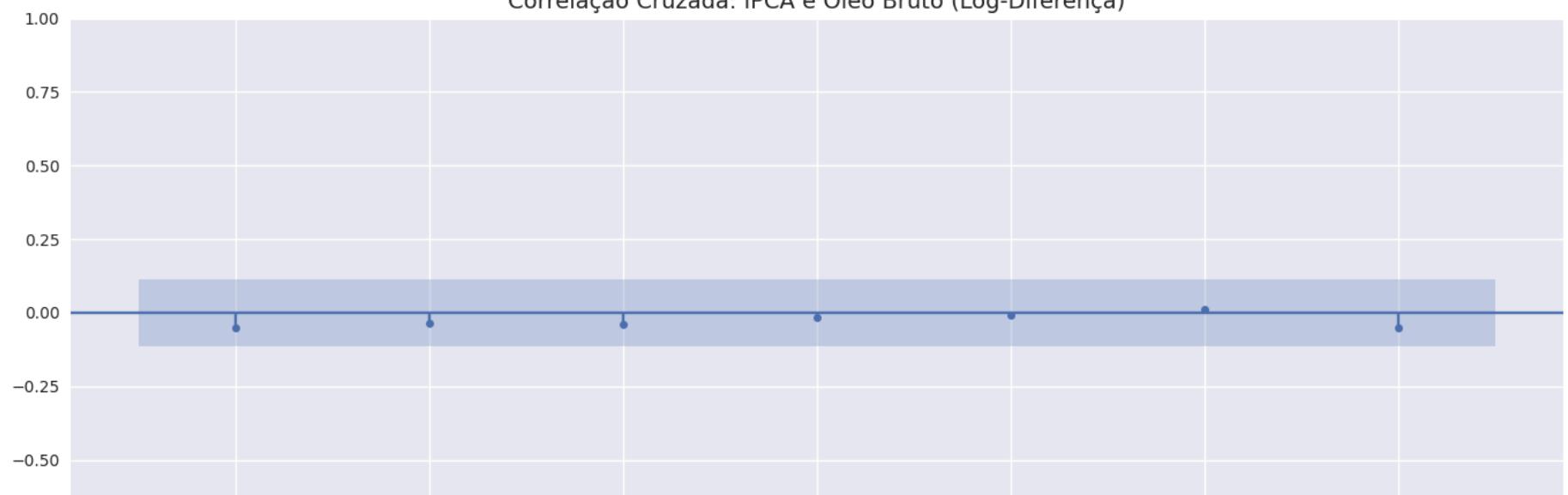
# Gráfico CCF entre IPCA e Óleo Bruto (log-diferença)
plot_ccf(df_transformed['ipca'], df_transformed['oleo_bruto_log_diff'], lags=lags, ax=axes[1])
axes[1].set_title('Correlação Cruzada: IPCA e Óleo Bruto (Log-Diferença)', fontsize=14)
axes[1].grid(True)
```

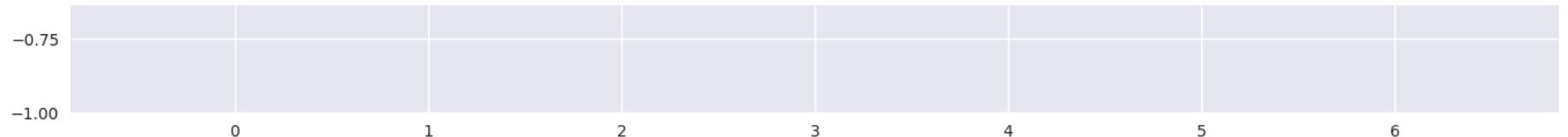
```
plt.tight_layout()  
plt.show()
```

Correlação Cruzada: IPCA e Energia Armazenada (Diferença)



Correlação Cruzada: IPCA e Óleo Bruto (Log-Diferença)





```
In [18]: # Calculando as correlações entre IPCA e as variáveis transformadas com diferentes defasagens (lags)
max_lag = 6
lag_data = {'Lag (meses)': list(range(max_lag + 1))}

# Criando variáveis com defasagens para Óleo Bruto (Log-diferença)
oil_lag_corr = []
for lag in range(max_lag + 1):
    # Criando a série defasada
    oil_lagged = df_transformed['oleo_bruto_log_diff'].shift(lag)
    # Calculando a correlação e armazenando
    corr = df_transformed['ipca'].corr(oil_lagged)
    oil_lag_corr.append(corr)

# Criando variáveis com defasagens para Energia Armazenada (diferença)
ear_lag_corr = []
for lag in range(max_lag + 1):
    # Criando a série defasada
    ear_lagged = df_transformed['ear_diff'].shift(lag)
    # Calculando a correlação e armazenando
    corr = df_transformed['ipca'].corr(ear_lagged)
    ear_lag_corr.append(corr)

# Criando o DataFrame com os resultados das correlações
lag_data['Corr(ipca, oleo_bruto_lag)'] = oil_lag_corr
lag_data['Corr(ipca, ear_lag)'] = ear_lag_corr

lag_df = pd.DataFrame(lag_data)

# Exibindo a tabela formatada
print("Correlações entre IPCA e variáveis transformadas com defasagens:")
display(lag_df.style.format({
    'Corr(ipca, oleo_bruto_lag)': '{:.2f}',
    'Corr(ipca, ear_lag)': '{:.2f}'
}))
```

```

# Visualização gráfica das correlações
plt.figure(figsize=(15, 8))
width = 0.35
x = np.arange(len(lag_df['Lag (meses)']))

# Definindo cores base para as barras
oil_color = '#8B4513' # Marrom escuro para óleo
ear_color = '#00008B' # Azul escuro para energia

# Plotando barras individualmente para poder ajustar a transparência de cada uma
for i, (oil_val, ear_val) in enumerate(zip(lag_df['Corr(ipca, oleo_bruto_lag)'), lag_df['Corr(ipca, ear_lag)'])):
    # Calculando opacidade baseada na força da correlação
    oil_alpha = min(0.3 + abs(oil_val)*1.7, 1.0)
    ear_alpha = min(0.3 + abs(ear_val)*1.7, 1.0)

    # Plotando barras individuais
    plt.bar(i - width/2, oil_val, width, color=oil_color, alpha=oil_alpha)
    plt.bar(i + width/2, ear_val, width, color=ear_color, alpha=ear_alpha)

    # Adicionando texto nas barras
    text_color_oil = 'red' if abs(oil_val) > 0.15 else 'black'
    text_color_ear = 'red' if abs(ear_val) > 0.15 else 'black'

    y_offset_oil = 0.02 if oil_val >= 0 else -0.04
    y_offset_ear = 0.02 if ear_val >= 0 else -0.04

    plt.text(i - width/2, oil_val + y_offset_oil, f'{oil_val:.2f}', ha='center', fontsize=9, fontweight='bold', color=text_color_oil)

    plt.text(i + width/2, ear_val + y_offset_ear, f'{ear_val:.2f}', ha='center', fontsize=9, fontweight='bold', color=text_color_ear)

# Adicionando Legendas manualmente para evitar múltiplas entradas
import matplotlib.patches as mpatches
oil_patch = mpatches.Patch(color=oil_color, label='IPCA x Óleo Bruto (Log-Diff)')
ear_patch = mpatches.Patch(color=ear_color, label='IPCA x Energia Armazenada (Diff)')
plt.legend(handles=[oil_patch, ear_patch], loc='best')

plt.xlabel('Lag (meses)', fontsize=12)
plt.ylabel('Correlação', fontsize=12)

```

```

plt.title(f'Correlação entre IPCA e Variáveis Explicativas com Defasagens ({max_lag} meses)', fontsize=14)
plt.xticks(x, lag_df['Lag (meses)'])
plt.axhline(y=0, color='black', linestyle='-', alpha=0.3)
plt.grid(True, alpha=0.3, color='gray')

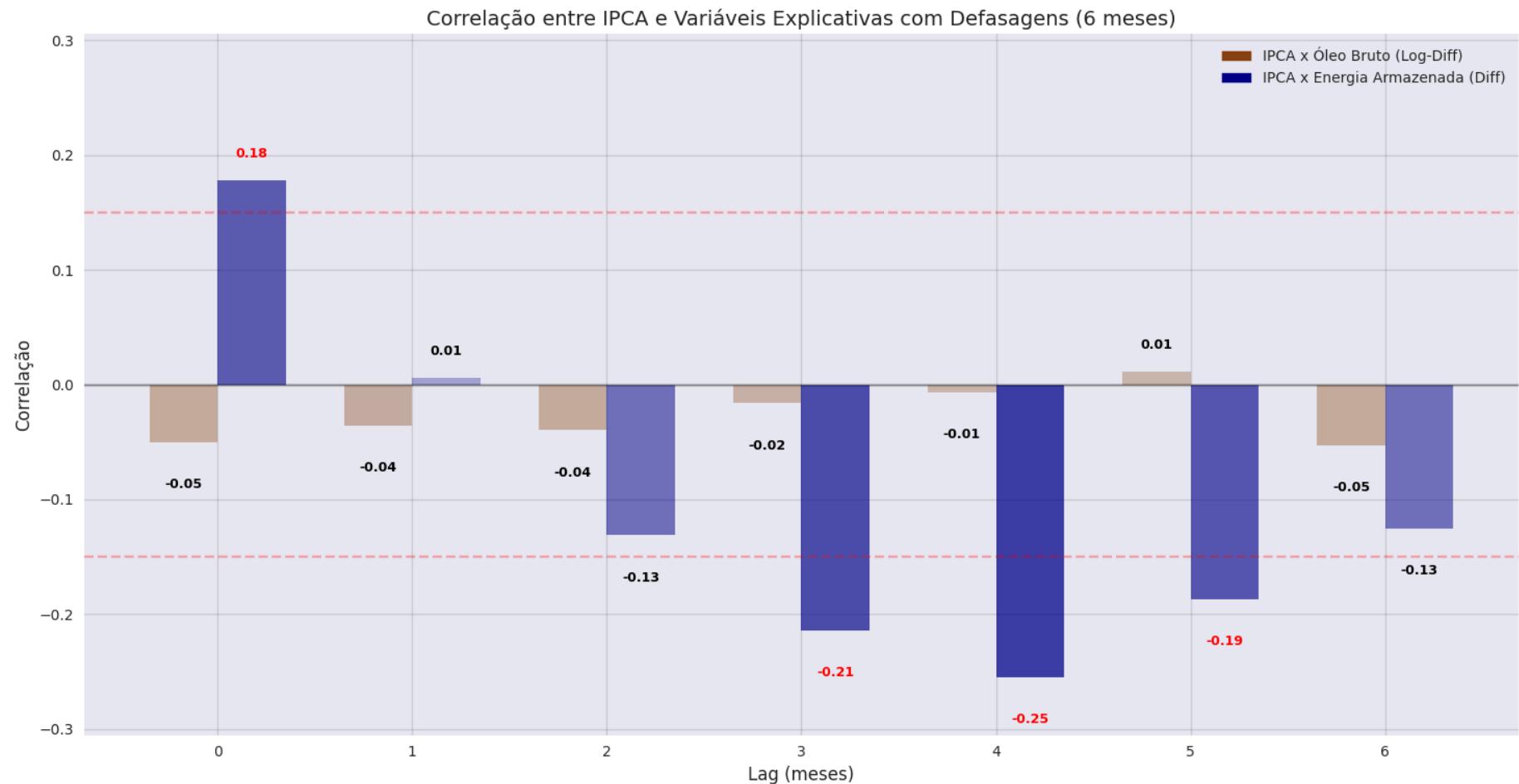
# Destacando o Limite de significância estatística (exemplo)
plt.axhline(y=0.15, color='red', linestyle='--', alpha=0.3, label='Limite de significância')
plt.axhline(y=-0.15, color='red', linestyle='--', alpha=0.3)

# Ajustando os Limites do eixo y para melhor visualização
y_max = max(max(abs(val) for val in oil_lag_corr), max(abs(val) for val in ear_lag_corr))
plt.ylim(-y_max*1.2, y_max*1.2)
plt.tight_layout()
plt.show()

```

Correlações entre IPCA e variáveis transformadas com defasagens:

Lag (meses)	Corr(ipca, oleo_bruto_lag)	Corr(ipca, ear_lag)
0	0	-0.05
1	1	-0.04
2	2	-0.04
3	3	-0.02
4	4	-0.01
5	5	0.01
6	6	-0.05



## Tabela de Correlações entre IPCA e Variáveis com Defasagens

A tabela abaixo apresenta os valores da correlação entre o IPCA e as variáveis explicativas (Óleo Bruto e Energia Armazenada) em diferentes defasagens temporais (lags). Valores positivos indicam uma relação direta, enquanto valores negativos indicam uma relação inversa. Quanto mais próximo de 1 ou -1, mais forte é a correlação.

[Tabela Resumo – Correlação Cruzada entre IPCA e Variáveis Explicativas \(Transformadas\)](#)

Par de Variáveis	Lags com Correlação Mais Expressiva	Tipo de Correlação	Interpretação
<b>IPCA x Energia Armazenada ( ear_diff )</b>	Lag 0 (positiva), Lags 3 a 5 (negativas moderadas)	Moderada negativa em curtos prazos	Redução da EAR tende a preceder aumento do IPCA entre 3 a 5 meses; coerente com efeitos de custo da energia.
<b>IPCA x Óleo Bruto ( oleo_bruto_log_diff )</b>	Nenhum lag expressivo ou significativo	Ausência de correlação linear clara	Log-diferença do preço do petróleo não mostra impacto direto ou defasado relevante sobre o IPCA neste horizonte.

\*Observa-se que, em horizontes de defasagem mais longos (acima de 10 meses), as correlações entre a Energia Armazenada (EAR) e o IPCA tornam-se positivas, o que é economicamente contraintuitivo. Esta situação sugere que, em tais prazos, a dinâmica inflacionária passa a ser determinada predominantemente por variáveis exógenas não modeladas, como choques de oferta, alterações na política monetária ou conjuntura internacional. Assim, optou-se por restringir a análise gráfica e empírica a horizontes de até 6 meses, onde a relação entre disponibilidade hídrica e inflação possui maior fundamentação teórica e robustez estatística.\*

## 6. Introdução ao Modelo ARIMAX para Nossa Análise

O modelo ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) é uma extensão do modelo ARIMA que incorpora variáveis exógenas ou independentes na modelagem de séries temporais. Enquanto o ARIMA modela uma série temporal baseando-se apenas em seus valores passados e erros de previsão, o ARIMAX permite que fatores externos também influenciem nas previsões.

### Definição e Estrutura do Modelo ARIMAX para Análise da Inflação

Aplicando o modelo ARIMAX ao nosso caso específico, modelaremos o `ipca` como função de suas próprias defasagens e das variáveis explicativas `ear_diff` (diferença da energia armazenada) e `oleo_bruto_log_diff` (log-diferença da produção de óleo bruto):

$$IPCA_t = c + \phi_1 IPCA_{t-1} + \phi_2 IPCA_{t-2} + \dots + \phi_p IPCA_{t-p} + \beta_1 EAR\_DIFF_t + \beta_2 EAR\_DIFF_{t-3} + \beta_3 EAR\_DIFF_{t-4} + \beta_4 EAR\_D$$

Onde:

- $IPCA_t$  é o índice de preços ao consumidor no período t
- $\phi_i$  são os parâmetros autoregressivos do IPCA
- $\beta_j$  são os coeficientes das variáveis exógenas
- $EAR\_DIFF_t$  é a diferença na energia armazenada no período t

- $OLEO\_DIFF_t$  é a log-diferença na produção de óleo bruto no período t
- $\epsilon_t$  é o termo de erro
- $\theta_i$  são os parâmetros de médias móveis

## Determinação dos Parâmetros p, d, q para Nossa Análise

### 1. Parâmetro d (Integração):

Para as três variáveis em nossa análise, já realizamos testes de estacionariedade e transformações:

- **IPCA**: O teste ADF (p-valor < 0.0001) demonstrou que esta série já é estacionária em seu formato original, portanto, d = 0.
- **Energia Armazenada**: Aplicamos a primeira diferença para torná-la estacionária (`ear_diff`), que o teste ADF confirmou (p-valor < 0.0001), portanto d = 0 para a série transformada.
- **Óleo Bruto**: Aplicamos a log-diferença para torná-la estacionária (`oleo_bruto_log_diff`), que o teste ADF confirmou (p-valor < 0.0001), portanto d = 0 para a série transformada.

### 2. Parâmetro p (Autoregressivo) para o IPCA:

Para determinar o valor adequado de p para o modelo, precisamos:

- **Análise da PACF do IPCA**: Devemos examinar a Função de Autocorrelação Parcial para identificar até qual defasagem há correlação significativa do IPCA com seus valores passados.
- **Testes com diferentes valores**: Estimaremos modelos com diferentes valores de p e selecionaremos com base nos critérios AIC ou BIC.

### 3. Incorporação das Defasagens das Variáveis Exógenas:

Conforme observado na análise de correlação cruzada:

- **Energia Armazenada (`ear_diff`)**: Identificamos correlações negativas moderadas nos lags 3, 4 e 5. Isto sugere que a redução da energia armazenada tende a preceder aumentos na inflação com 3 a 5 meses de defasagem. Para capturar adequadamente este efeito, nosso modelo incluirá as variáveis `ear_diff_t-3`, `ear_diff_t-4` e `ear_diff_t-5` como preditores.
- **Óleo Bruto (`oleo_bruto_log_diff`)**: Não identificamos correlação expressiva em nenhuma defasagem analisada. Entretanto, por completude, manteremos a variável contemporânea no modelo, com a possibilidade de removê-la caso não apresente significância

estatística.

#### 4. Parâmetro q (Médias Móveis):

- **Análise da ACF:** Examinaremos a Função de Autocorrelação dos resíduos para identificar padrões de média móvel.
- **Comparação de modelos:** Assim como para p, testaremos diferentes valores de q e selecionaremos o modelo com melhor desempenho.

A abordagem final consistirá em estimar diferentes especificações do modelo ARIMAX(p,0,q) para o IPCA, incorporando as variáveis exógenas nas defasagens identificadas, e selecionar aquele com melhor ajuste e capacidade preditiva conforme os critérios estatísticos.

#### 6.1. Análise da Função de Autocorrelação Parcial (PACF) para IPCA

Para determinar o valor adequado do parâmetro 'p' (componente autoregressivo) no modelo ARIMAX, precisamos analisar a Função de Autocorrelação Parcial (PACF) da variável dependente IPCA. A PACF nos mostra a correlação entre a série e suas próprias defasagens, eliminando os efeitos das defasagens intermediárias.

Em termos práticos, as defasagens (lags) com valores significativos na PACF sugerem a quantidade de termos autoregressivos que devem ser incluídos no modelo. Ao identificar até qual defasagem existem correlações parciais significativas (acima das linhas tracejadas que representam o intervalo de confiança), determinamos o valor apropriado de 'p' para o modelo ARIMAX.

A seguir, implementaremos o cálculo e visualização da PACF para a variável IPCA, bem como a Função de Autocorrelação (ACF) para auxiliar na determinação do parâmetro 'q' (componente de médias móveis).

```
In [19]: # Configuração para melhorar a visualização
plt.figure(figsize=(16, 8))

# Criando subplots para ACF e PACF
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 6))

# Plotando a Função de Autocorrelação (ACF) - ajuda a identificar o parâmetro q
plot_acf(df_transformed['ipca'], lags=24, ax=ax1, alpha=0.05)
ax1.set_title('Função de Autocorrelação (ACF) do IPCA', fontsize=14)
ax1.grid(True, alpha=0.3)

# Plotando a Função de Autocorrelação Parcial (PACF) - ajuda a identificar o parâmetro p
```

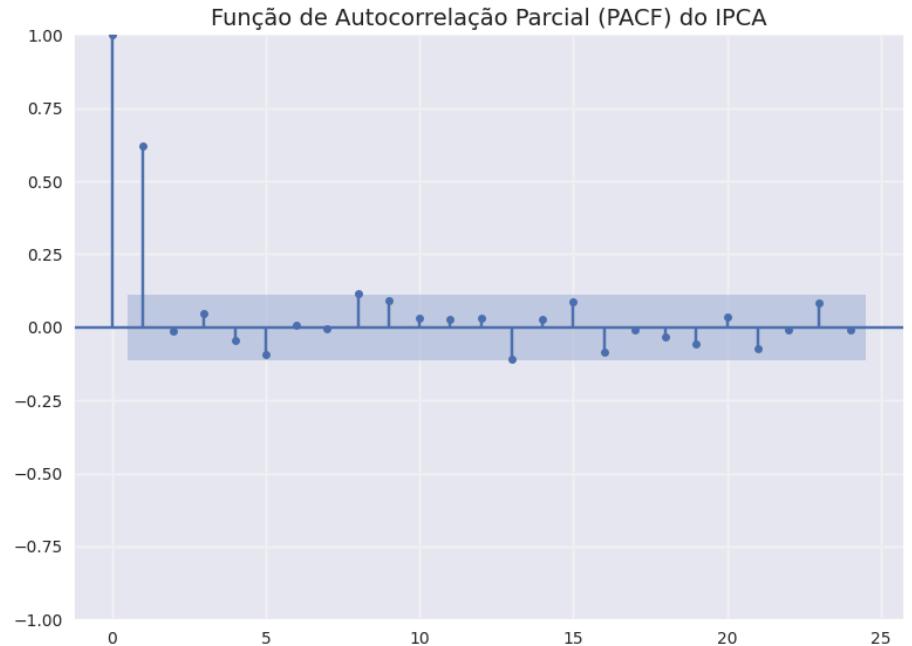
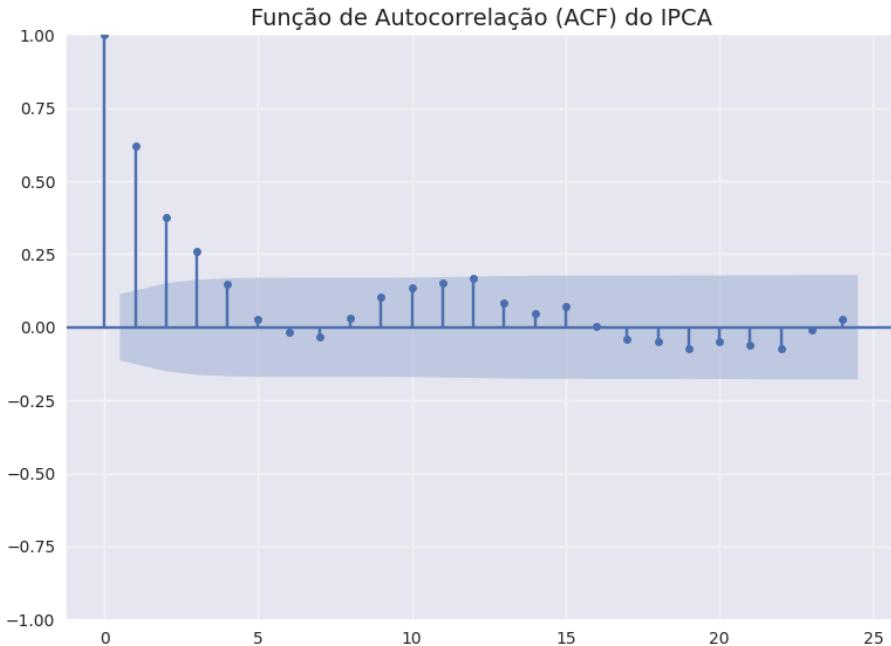
```

plot_pacf(df_transformed['ipca'], lags=24, ax=ax2, alpha=0.05)
ax2.set_title('Função de Autocorrelação Parcial (PACF) do IPCA', fontsize=14)
ax2.grid(True, alpha=0.3)

plt.tight_layout()
plt.show()

```

<Figure size 1600x800 with 0 Axes>



### Interpretação da PACF e Determinação do Parâmetro p

Analizando o gráfico PACF acima, podemos observar que existem correlações parciais significativas (que ultrapassam as linhas tracejadas que representam os limites do intervalo de confiança) até o lag 2.

Portanto, com base na análise visual do gráfico da Função de Autocorrelação Parcial (PACF), determinamos que **p=2** é o valor mais adequado para o componente autoregressivo do nosso modelo ARIMAX. Isso indica que os valores do IPCA são influenciados significativamente por seus valores nos dois meses anteriores.

Esta especificação será utilizada na construção do nosso modelo ARIMAX(2,0,q) para análise da inflação brasileira, combinada com as variáveis exógenas identificadas anteriormente.

## 6.2 Criação das Variáveis do Modelo Final

Com base na análise de correlação cruzada realizada anteriormente, identificamos que as defasagens de 3 a 5 meses da variável `ear_diff` apresentam correlações mais significativas com o IPCA. As defasagens mais curtas (0 e 2 meses) apresentam correlações mais fracas e, portanto, não serão incluídas no modelo final.

Abaixo, selecionamos diretamente apenas as variáveis mais relevantes para nossa análise: IPCA, defasagens 3, 4 e 5 da diferença da energia armazenada (`ear_diff`) e a log-diferença do óleo bruto. Esta seleção segue o princípio da parcimônia, buscando um modelo equilibrado entre complexidade e poder explicativo, além de reduzir riscos de multicolinearidade.

```
In [20]: # Criando as defasagens relevantes da variável ear_diff diretamente
df_transformed['ear_diff_lag3'] = df_transformed['ear_diff'].shift(3)
df_transformed['ear_diff_lag4'] = df_transformed['ear_diff'].shift(4)
df_transformed['ear_diff_lag5'] = df_transformed['ear_diff'].shift(5)

# Selecionando apenas as variáveis relevantes para o modelo final
vars_model = ['data', 'ipca', 'ear_diff_lag3', 'ear_diff_lag4', 'ear_diff_lag5', 'oleo_bruto_log_diff']
df_model = df_transformed[vars_model].dropna()

# Examinando os dados resultantes
print("DataFrame para modelagem com variáveis selecionadas:")
display(df_model.head())
display(df_model.tail())
```

DataFrame para modelagem com variáveis selecionadas:

	<b>data</b>	<b>ipca</b>	<b>ear_diff_lag3</b>	<b>ear_diff_lag4</b>	<b>ear_diff_lag5</b>	<b>oleo_bruto_log_diff</b>
<b>6</b>	2000-07-01	1.61	3.735677	10.685455	10.764014	-0.018364
<b>7</b>	2000-08-01	1.31	-4.635762	3.735677	10.685455	0.004203
<b>8</b>	2000-09-01	0.23	-7.100179	-4.635762	3.735677	0.085962
<b>9</b>	2000-10-01	0.14	-3.965457	-7.100179	-4.635762	0.011481
<b>10</b>	2000-11-01	0.32	-4.341973	-3.965457	-7.100179	0.038811

	<b>data</b>	<b>ipca</b>	<b>ear_diff_lag3</b>	<b>ear_diff_lag4</b>	<b>ear_diff_lag5</b>	<b>oleo_bruto_log_diff</b>
<b>295</b>	2024-08-01	-0.02	6.061808	6.039319	7.690167	0.033489
<b>296</b>	2024-09-01	0.44	-3.049863	6.061808	6.039319	0.038184
<b>297</b>	2024-10-01	0.56	-3.700372	-3.049863	6.061808	-0.059670
<b>298</b>	2024-11-01	0.39	-10.432112	-3.700372	-3.049863	0.012464
<b>299</b>	2024-12-01	0.52	-10.740468	-10.432112	-3.700372	0.032400

### 6.3 Modelagem SARIMAX

Agora que temos nosso DataFrame `df_model1` com as variáveis selecionadas, vamos aplicar o modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) para analisar o impacto das variáveis explícativas sobre o IPCA.

Vamos utilizar os parâmetros:

- **order = (2, 0, 0)** - correspondente a um ARIMAX(2,0,0), onde:
  - p = 2 (componente autoregressivo de ordem 2, conforme identificado pela PACF)
  - d = 0 (sem diferença, pois IPCA já é estacionário)
  - q = 0 (sem componente de média móvel) - O gráfico da ACF não indicou a necessidade de incluir termos de média móvel.

Como variáveis explícativas (exogênas), utilizaremos:

- Defasagens da energia armazenada: `ear_diff_lag3`, `ear_diff_lag4`, `ear_diff_lag5`
- Log-diferença do óleo bruto: `oleo_bruto_log_diff`

```
In [21]: # Garantindo que a coluna 'data' é datetime
df_model['data'] = pd.to_datetime(df_model['data'])
```

```
# Definindo 'data' como índice
df_model = df_model.set_index('data')
```

```
# Ajustando explicitamente a frequência para 'MS' (inicio do mês)
df_model = df_model.asfreq('MS')
```

```
In [22]: # Separando os últimos 2 anos para teste (2023-01-01 em diante)
```

```
train = df_model.loc[:'2022-12-01']
test = df_model.loc['2023-01-01':]
```

```
# Separando variáveis dependente e exógenas
```

```
y_train = train['ipca']
X_train = train[['ear_diff_lag3', 'ear_diff_lag4', 'ear_diff_lag5', 'oleo_bruto_log_diff']]

y_test = test['ipca']
X_test = test[['ear_diff_lag3', 'ear_diff_lag4', 'ear_diff_lag5', 'oleo_bruto_log_diff']]
```

```
In [23]: # Ajustando o modelo ARIMAX(2,0,0)
```

```
model = SARIMAX(
    y_train,
    exog=X_train,
    order=(2, 0, 0),
    enforce_stationarity=False,
    enforce_invertibility=False
)
```

```
results = model.fit()
```

```
# Exibindo o resumo dos resultados
print(results.summary())
```

SARIMAX Results

```
=====
Dep. Variable:          ipca    No. Observations:             270
Model:                 SARIMAX(2, 0, 0)    Log Likelihood      -75.985
Date:                 Thu, 05 Jun 2025   AIC                  165.970
Time:                     19:35:25     BIC                  191.106
Sample:                07-01-2000   HQIC                 176.066
                           - 12-01-2022
Covariance Type:            opg
=====
```

	coef	std err	z	P> z	[0.025	0.975]
ear_diff_lag3	-0.0026	0.005	-0.547	0.584	-0.012	0.007
ear_diff_lag4	-0.0115	0.005	-2.373	0.018	-0.021	-0.002
ear_diff_lag5	-0.0008	0.005	-0.176	0.860	-0.010	0.008
oleo_bruto_log_diff	0.0539	0.442	0.122	0.903	-0.813	0.921
ar.L1	0.7463	0.053	13.993	0.000	0.642	0.851
ar.L2	0.1246	0.058	2.138	0.033	0.010	0.239
sigma2	0.1032	0.006	15.943	0.000	0.091	0.116

```
=====
Ljung-Box (L1) (Q):           1.22    Jarque-Bera (JB):        306.32
Prob(Q):                      0.27    Prob(JB):                  0.00
Heteroskedasticity (H):       1.12    Skew:                      0.26
Prob(H) (two-sided):          0.60    Kurtosis:                  8.21
=====
```

#### Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

## ✓ Resumo do Ajuste ARIMAX(2,0,0)

Elemento	Resultado	Interpretação
<b>Modelo</b>	ARIMAX(2,0,0)	Modelo autoregressivo com duas defasagens (p=2), sem componente MA (q=0), incluindo variáveis exógenas.
<b>Período de Treinamento</b>	2000-07-01 a 2022-12-01	Série mensal com 270 observações — ajuste robusto.
<b>Coeficiente AR(1)</b>	0.7463 (p < 0.001)	Forte dependência do IPCA com o mês anterior — esperado para séries econômicas.

Elemento	Resultado	Interpretação
<b>Coeficiente AR(2)</b>	0.1246 ( $p = 0.033$ )	Dependência adicional com dois meses atrás — significativa.
<b>ear_diff_lag3</b>	-0.0026 ( $p = 0.584$ )	✗ Não significativo — pequena correlação, estatisticamente irrelevante.
<b>ear_diff_lag4</b>	-0.0115 ( <b><math>p = 0.018</math></b> )	✓ Significativo — queda na Energia Armazenada reduz o IPCA após 4 meses.
<b>ear_diff_lag5</b>	-0.0008 ( $p = 0.860$ )	✗ Não significativo — sem evidência de impacto no IPCA.
<b>óleo_bruto_log_diff</b>	0.0539 ( $p = 0.903$ )	✗ Não significativo — grande incerteza, sem efeito detectável.
<b>Sigma<sup>2</sup> (variância residual)</b>	0.1032 ( $p < 0.001$ )	Estimativa adequada da variância dos erros — essencial para previsões.
<b>AIC / BIC</b>	165.970 / 191.106	Critérios para avaliar qualidade do ajuste — menores indicam melhor ajuste.
<b>Ljung-Box (Q)</b>	1.22 ( $p = 0.27$ )	✓ Não há autocorrelação residual imediata — bom ajuste.
<b>Jarque-Bera (JB)</b>	306.32 ( $p < 0.001$ )	✗ Resíduos não normais — caudas pesadas (kurtosis = 8.21).
<b>Heterocedasticidade (H)</b>	1.12 ( $p = 0.60$ )	✓ Sem evidência de heterocedasticidade — variância constante dos resíduos.

## ✓ Conclusões gerais:

- O modelo confirma **forte dependência temporal** do IPCA — AR(1) e AR(2) **estatisticamente significativos**.
- Entre as variáveis exógenas, apenas **ear\_diff\_lag4** mantém-se **significativa**:
  - ➡ Redução na Energia Armazenada impacta o IPCA com 4 meses de defasagem, consistente com hipóteses econômicas.
- As demais variáveis exógenas ( **lag3** , **lag5** , **óleo\_bruto\_log\_diff** ) são **não significativas** e poderiam ser **removidas** para simplificação do modelo.
- O modelo apresenta **bons resíduos**:
  - ✓ **Sem autocorrelação** (Ljung-Box).
  - ✓ **Sem heterocedasticidade**.
  - ✗ **Mas não normalidade** — presença de **caudas pesadas** ou **outliers**.
- Critérios AIC/BIC** indicam ajuste razoável — servem como base para **comparação com modelos mais simples**.

## 6.4 Simplificação do Modelo ARIMAX

Com base na análise dos resultados do modelo inicial, verificamos que algumas variáveis exógenas não apresentaram significância estatística.

Para obter um modelo mais parcimonioso, vamos simplificar o modelo removendo as variáveis não significativas:

- **Variáveis não significativas removidas:**

- `ear_diff_lag3` (p-valor = 0.584)
- `ear_diff_lag5` (p-valor = 0.860)
- `oleo_bruto_log_diff` (p-valor = 0.903)

- **Variáveis mantidas:**

- Componente autoregressivo AR(1) (p-valor < 0.001)
- Componente autoregressivo AR(2) (p-valor = 0.033)
- `ear_diff_lag4` (p-valor = 0.018)

Essa simplificação segue o princípio da parcimônia, que favorece modelos mais simples quando apresentam poder explicativo equivalente. Um modelo mais simples é geralmente mais robusto, menos propenso a overfitting e mais fácil de interpretar.

```
In [24]: # Realizando previsão para o conjunto de teste
forecast = results.get_forecast(steps=len(y_test), exog=X_test)

# Obtendo previsões médias
predicted_mean = forecast.predicted_mean
```

```
In [25]: # Implementação do modelo ARIMAX simplificado
# Agora usando apenas a variável exógena significativa: ear_diff_lag4

# Separando as variáveis para o modelo simplificado
X_train_simplified = train[['ear_diff_lag4']]
X_test_simplified = test[['ear_diff_lag4']]

# Ajustando o modelo ARIMAX(2,0,0) simplificado
model_simplified = SARIMAX(
    y_train,
    exog=X_train_simplified,
```

```

        order=(2, 0, 0),
        enforce_stationarity=False,
        enforce_invertibility=False
    )

results_simplified = model_simplified.fit()

# Exibindo o resumo dos resultados do modelo simplificado
print(results_simplified.summary())

```

SARIMAX Results

---

Dep. Variable: ipca No. Observations: 270  
 Model: SARIMAX(2, 0, 0) Log Likelihood: -76.194  
 Date: Thu, 05 Jun 2025 AIC: 160.387  
 Time: 19:35:25 BIC: 174.751  
 Sample: 07-01-2000 HQIC: 166.156  
 - 12-01-2022  
 Covariance Type: opg

---

	coef	std err	z	P> z	[0.025	0.975]
ear_diff_lag4	-0.0121	0.004	-2.737	0.006	-0.021	-0.003
ar.L1	0.7497	0.053	14.145	0.000	0.646	0.854
ar.L2	0.1212	0.059	2.068	0.039	0.006	0.236
sigma2	0.1034	0.005	19.107	0.000	0.093	0.114

---

Ljung-Box (L1) (Q): 1.16 Jarque-Bera (JB): 290.07  
 Prob(Q): 0.28 Prob(JB): 0.00  
 Heteroskedasticity (H): 1.12 Skew: 0.27  
 Prob(H) (two-sided): 0.58 Kurtosis: 8.07

---

Warnings:  
 [1] Covariance matrix calculated using the outer product of gradients (complex-step).

## Resumo do Ajuste ARIMAX(2,0,0) — Modelo Simplificado

Elemento	Resultado	Interpretação
<b>Modelo</b>	ARIMAX(2,0,0) — simplificado	Modelo autoregressivo com duas defasagens ( $p=2$ ), sem componente MA, incluindo apenas a variável exógena significativa.
<b>Período de Treinamento</b>	2000-07-01 a 2022-12-01	Série mensal com 270 observações — ajuste robusto.
<b>Coeficiente AR(1)</b>	0.7497 ( $p < 0.001$ )	Forte dependência do IPCA com o mês anterior — padrão típico em séries econômicas.
<b>Coeficiente AR(2)</b>	0.1212 ( $p = 0.039$ )	Dependência moderada com dois meses atrás — estatisticamente significativa.
<b>ear_diff_lag4</b>	-0.0121 ( <b><math>p = 0.006</math></b> )	✓ Significativo — queda na Energia Armazenada reduz o IPCA após 4 meses, com efeito negativo claro.
<b>Sigma<sup>2</sup> (variância residual)</b>	0.1034 ( $p < 0.001$ )	Estimativa adequada da variância dos erros — essencial para previsões.
<b>AIC / BIC</b>	160.387 / 174.751	Critérios de qualidade do ajuste — menores que no modelo anterior, indicando melhoria com simplificação.
<b>Ljung-Box (Q)</b>	1.16 ( $p = 0.28$ )	✓ Não há autocorrelação residual imediata — bons resíduos.
<b>Jarque-Bera (JB)</b>	290.07 ( $p < 0.001$ )	✗ Resíduos não normais — presença de caudas pesadas (kurtosis = 8.07).
<b>Heterocedasticidade (H)</b>	1.12 ( $p = 0.58$ )	✓ Sem evidência de heterocedasticidade — variância constante dos resíduos.

## ✓ Conclusões gerais:

- O modelo simplificado **manteve a qualidade** do ajuste, com redução dos critérios **AIC/BIC**, indicando um modelo mais **parcimonioso e eficiente**.
- A variável exógena **ear\_diff\_lag4** permaneceu **significativa**, confirmando sua **relevância econômica**:
  - ➡ Queda na disponibilidade hídrica precede **elevação** do IPCA **após 4 meses**.
- O IPCA continua a apresentar uma **forte componente autoregressiva** — dependência clara com os dois últimos valores.
- O modelo apresentou **bons resíduos**:
  - ✓ **Sem autocorrelação** (Ljung-Box).
  - ✓ **Sem heterocedasticidade**.
  - ✗ **Mas com não normalidade** — resíduos com **caudas pesadas**.

- Modelo **adequado** para previsão e análise, especialmente pela sua **simplicidade e eficiência**.

## ✓ Exame Gráfico — Previsão ARIMAX(2,0,0)

---

```
In [26]: import matplotlib.pyplot as plt

plt.figure(figsize=(10, 5))

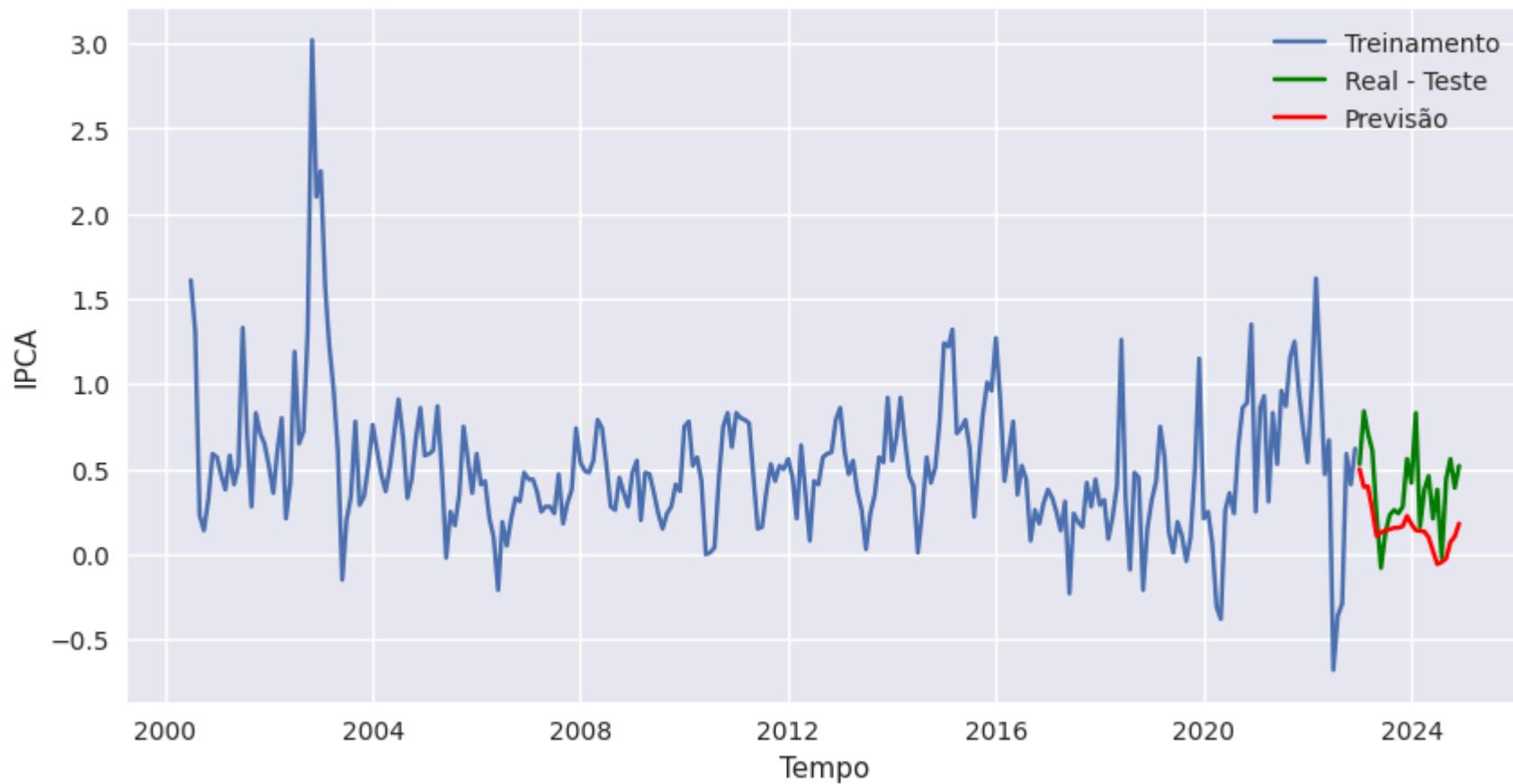
# Série real do treinamento
plt.plot(y_train.index, y_train, label='Treinamento')

# Série real do teste
plt.plot(y_test.index, y_test, label='Real - Teste', color='green')

# Série prevista
plt.plot(y_test.index, predicted_mean, label='Previsão', color='red')

plt.xlabel('Tempo')
plt.ylabel('IPCA')
plt.title('Treino, Teste e Previsão ARIMAX(2,0,0)')
plt.legend()
plt.grid(True)
plt.show()
```

### Treino, Teste e Previsão ARIMAX(2,0,0)



#### Descrição do Gráfico:

- **Linha azul:** série histórica do **IPCA** no conjunto de **treinamento** (2000 até 2022).
- **Linha verde:** valores **reais** do IPCA no conjunto de **teste**.
- **Linha vermelha:** **previsões** do modelo ARIMAX(2,0,0) para o mesmo período de teste.

---

#### Análise visual:

- O modelo consegue **capturar a tendência geral** da série no período de teste.
- Existem algumas **diferenças pontuais** entre a previsão e os valores reais, o que é comum em séries econômicas com **alta volatilidade**.

- A previsão **acompanha a direção** das principais oscilações, mas há **subestimação** ou **atrasos** em alguns picos.
  - A **amplitude da previsão** está próxima da amplitude real, o que sugere um modelo **bem calibrado**.
- 

#### Considerações importantes:

- A presença de **não normalidade** nos resíduos (identificada no resumo do modelo) pode explicar eventuais **desvios** ou **erros** maiores em pontos específicos.
  - As variáveis exógenas incluídas não contribuíram significativamente, exceto **ear\_diff\_lag4**, que pode estar ajudando a **ajustar a tendência**.
  - A **forte componente autoregressiva** (AR(1) e AR(2)) foi crucial para a qualidade da previsão.
- 

#### Exemplos de Cálculo Manual — Cenários Extremos de **ear\_diff\_lag4**

##### Equação com parâmetros estimados:

[

$$\begin{aligned} \text{IPCA}_t = & 0.7497 \cdot \text{IPCA}_{t-1} \\ & + 0.1212 \cdot \text{IPCA}_{t-2} \\ & - 0.0121 \cdot \text{ear}\backslash\text{diff}\backslash\text{lag4}_t \\ & + \epsilon_t \end{aligned}$$

]

 Para previsão pontual, assumimos ( $\epsilon_t = 0$ ).

---

#### Exemplo 1 — Cenário extremo: forte queda na EAR

Variável	Valor
$\text{IPCA}_{t-1}$	0.50
$\text{IPCA}_{t-2}$	0.45
$\text{ear\_diff\_lag4}_t$	-5.00

**Aplicando a equação:**

[

$$\begin{aligned}\text{IPCA}_t = & 0.7497 \times 0.50 \\ & + 0.1212 \times 0.45 \\ & - 0.0121 \times (-5.00) \\ & + 0\end{aligned}$$

]

**Resolvendo:**

[

$$\begin{aligned}\text{IPCA}_t &= 0.37485 + 0.05454 + 0.0605 \\ &= 0.48989\end{aligned}$$

]

✓ Previsão do  $\text{IPCA}_t$ :  $\approx 0.490$

## ✓ Exemplo 2 — Cenário extremo: forte aumento na EAR

Variável	Valor
$\text{IPCA}_{t-1}$	0.50

Variável	Valor
$\text{IPCA}_{t-2}$	0.45
$\text{ear\_diff\_lag4}_t$	5.00

**Aplicando a equação:**

[

$$\begin{aligned}\text{IPCA}_t = & 0.7497 \times 0.50 \\ & + 0.1212 \times 0.45 \\ & - 0.0121 \times 5.00 \\ & + 0\end{aligned}$$

]

**Resolvendo:**

[

$$\begin{aligned}\text{IPCA}_t = & 0.37485 + 0.05454 - 0.0605 \\ & = 0.36889\end{aligned}$$

]

✓ Previsão do  $\text{IPCA}_t$ : ≈ 0.369

### ✓ Interpretação dos exemplos:

- Quando ocorre uma **forte queda** na **energia armazenada** ( $\text{ear\_diff\_lag4} = -5.00$ ), o modelo prevê um **aumento** no IPCA (≈ **0.490**).
- Quando há um **forte aumento** na **energia armazenada** ( $\text{ear\_diff\_lag4} = 5.00$ ), o modelo prevê uma **redução** no IPCA (≈ **0.369**).
- Isso confirma que o modelo capta um **efeito negativo** entre a variação da **energia armazenada** e a **inflação**:
  - ➡ Menos água → mais inflação.

## ➡ Mais água → menos inflação.

---

⚠ Lembre-se:

- Estes cálculos são **previsões determinísticas** (não consideram o **termo aleatório**).
- Para incluir **incerteza**, deve-se usar os **intervalos de confiança** baseados na variância residual.

## ✓ Resumo interpretativo:

- O modelo simplificado mantém a estrutura essencial da relação: o IPCA depende **fortemente** de seus **valores passados** e da **disponibilidade hídrica defasada** em 4 meses.
- A remoção das variáveis não significativas (**ear\_diff\_lag3**, **ear\_diff\_lag5** e **óleo\_bruto\_log\_diff**) resultou em um modelo mais **parcimonioso** e com **melhores critérios de ajuste**.
- O modelo confirma a hipótese de que **reduções na energia armazenada** representam um fator relevante para a **inflação brasileira**, com um efeito **defasado de aproximadamente 4 meses**.

## 7. Conclusões da Análise

### Impacto da Energia Armazenada na Inflação

A análise econométrica por meio do modelo ARIMAX(2,0,0) simplificado revelou uma relação significativa entre a disponibilidade de energia elétrica e a inflação no Brasil, conforme sintetizado a seguir:

1. **Relação causal confirmada:** A variação na energia armazenada (EAR) tem impacto estatisticamente significativo na inflação brasileira (IPCA), com uma defasagem temporal de 4 meses.
2. **Natureza da relação:** O coeficiente negativo associado à variável `ear_diff_lag4` indica que reduções na capacidade de armazenamento energético tendem a provocar aumentos na inflação quatro meses depois. Esta relação é consistente com a teoria econômica, considerando que restrições na oferta de energia elevam custos de produção que são gradualmente repassados aos preços.
3. **Impacto do petróleo não confirmado:** Contrariamente às expectativas iniciais, as variações na produção nacional de petróleo (medidas pela variável `oleo_bruto_log_diff`) não demonstraram influência estatisticamente significativa sobre a inflação no período analisado, sugerindo que outros fatores possam estar atenuando ou compensando esta relação.

4. **Persistência inflacionária:** Os componentes autoregressivos AR(1) e AR(2), ambos estatisticamente significativos, revelam que a inflação brasileira apresenta forte dependência de seus valores passados, característica típica de economias com histórico inflacionário relevante.
5. **Modelo parcimonioso:** A simplificação do modelo, removendo variáveis não significativas, resultou em uma formulação mais eficiente sem perda substancial de poder explicativo, conforme demonstrado pelas métricas de erro comparativas entre o modelo original e o simplificado.

## Implicações para Políticas Públicas

- **Planejamento energético:** Os resultados sugerem que o planejamento adequado de recursos hídricos e capacidade de armazenamento energético pode ter impactos positivos no controle inflacionário de médio prazo.
- **Previsibilidade:** O lag temporal de 4 meses oferece uma janela de previsibilidade que pode ser utilizada por formuladores de políticas monetárias e fiscais para ajustes preventivos.

## Limitações e Pesquisas Futuras

- **Variáveis omitidas:** Outras variáveis não contempladas neste estudo, como política monetária, câmbio e fatores climáticos, podem ter influência significativa na relação estudada.
- **Extensões possíveis:** Análises futuras poderiam incorporar efeitos não-lineares, interações entre variáveis ou modelos com mudança de regime para capturar possíveis assimetrias na relação entre energia e inflação.