

Relatório de estudo de caso para geração de cenários macroeconômicos.

Janeiro, 2024

Sumário

Sumário.....	1
Resumo.....	2
1.Coleta de Dados.....	3
2.Exploração e análise dos dados.....	4
3.Dados Estacionários.....	14
4.Modelagem.....	20
5.Predição.....	26
6.Conclusão.....	31
7.Referências.....	32

Resumo

Este relatório foi elaborado como um estudo voltado à análise e estimativa da taxa Selic, do IPCA e do Dólar, utilizando métodos como o modelo SARIMA, entre outras ferramentas estatísticas. A escolha do modelo SARIMA encontra respaldo em Bonno (2014), que, ao comparar diferentes modelos de séries temporais, apontou o SARIMA como o mais eficaz para prever a inflação brasileira. Durante as análises intermediárias, foram obtidas importantes conclusões, como a identificação de correlações e padrões sazonais. Algumas previsões apresentaram coerência prática no cenário atual, mas destaca-se a necessidade de otimização das variáveis em certos modelos e a importância de considerar os contextos da análise fundamentalista para obter resultados mais robustos.

1.Coleta de Dados

A coleta dos dados se deu pelo site de Recieri (2025). Pois os dados já estavam pré limpos. A fonte possui dados de Selic, IpcA e Dólar, mensalmente no período de jan/2000 até dez/2024.

Mês	SELIC	IPCA	DÓLAR
dez/24	12,25%	4,83%	6,12
nov/24	11,25%	4,87%	5,85
out/24	10,75%	4,76%	5,61

Figura 1.1. Parte dos dados coletados e trabalhados neste projeto.

2.Exploração e análise dos dados

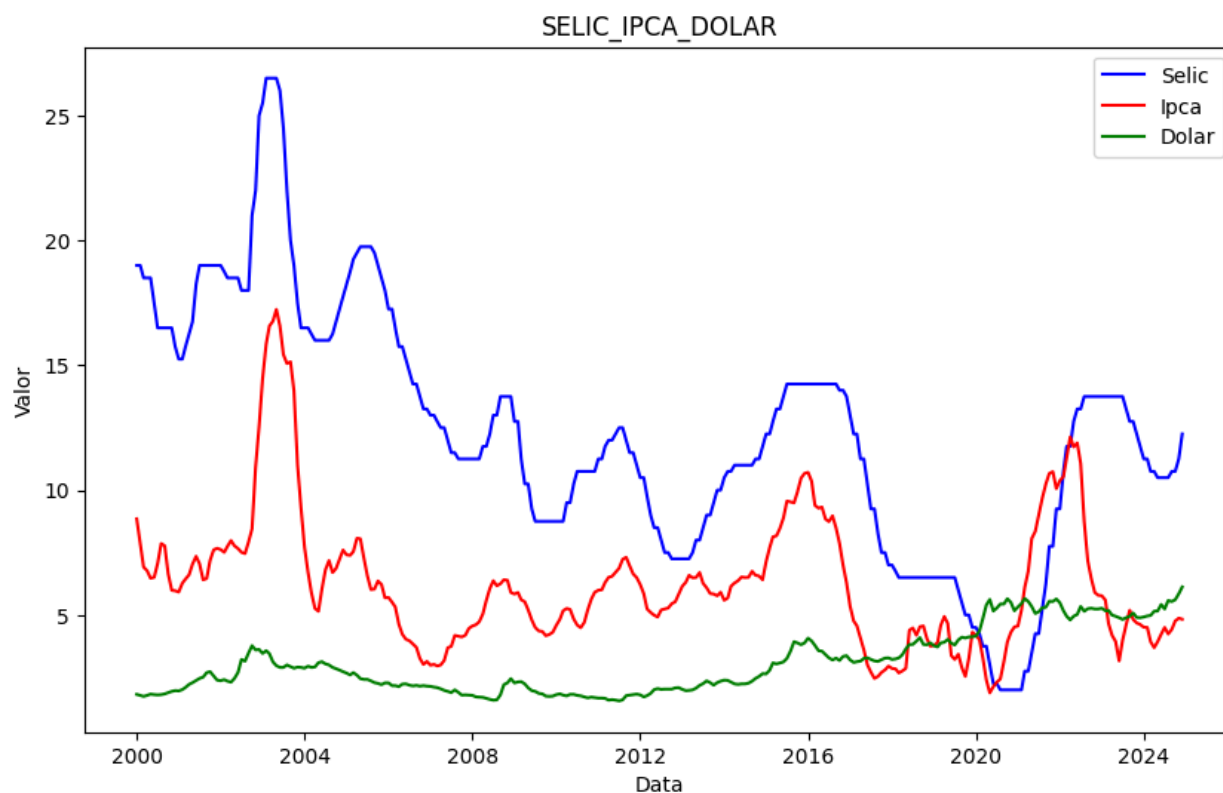


Figura 2.1. Gráfico de Selic, Ipca e Dólar sobre tempo.

Na Figura 2.1 podemos notar um certo grau de sincronismo entre Selic e Ipca, mas não para o dólar.

Um heatmap do coeficiente de Pearson é uma representação visual que mostra as correlações lineares entre diferentes variáveis. O coeficiente de Pearson varia de -1 (correlação negativa perfeita) a +1 (correlação positiva perfeita), com valores próximos de zero indicando pouca ou nenhuma correlação. Esse tipo de gráfico facilita a identificação de padrões e relações entre variáveis em um conjunto de dados.

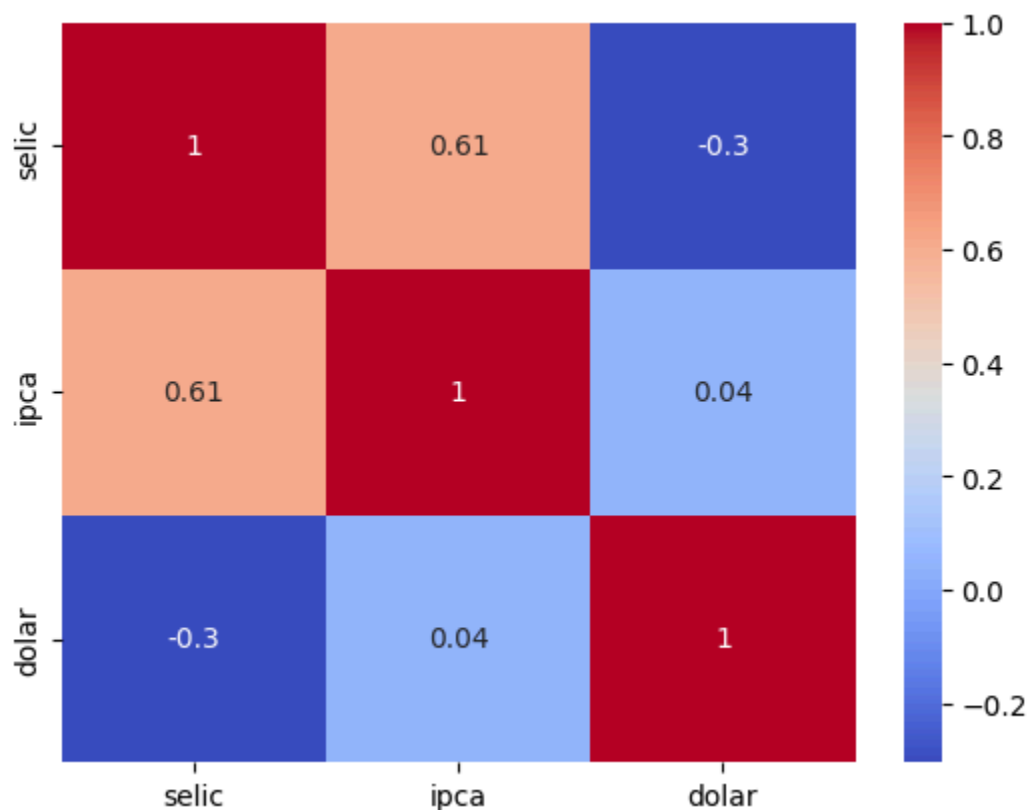


Figura 2.2. Heatmap do coeficiente de pearson.

Segundo a Figura 2.2, temos uma correlação alta entre Selic e Ipca, de 0,61, mas o dólar não se correlaciona com nenhum dos outros dados.

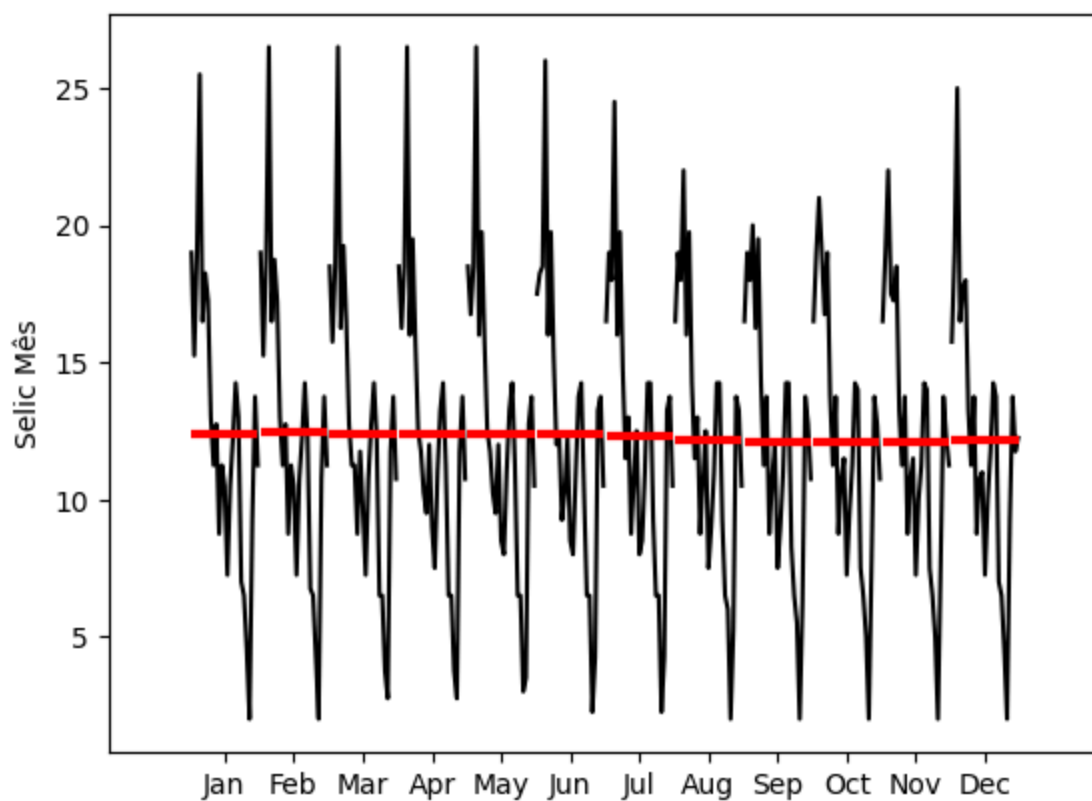


Figura 2.3. Média do Selic agrupada por mês.

Segundo a Figura 2.3, Selic tem um patamar médio mínimo perto do mês de Setembro. De Janeiro a Junho aparenta uma estabilidade com valores altos.

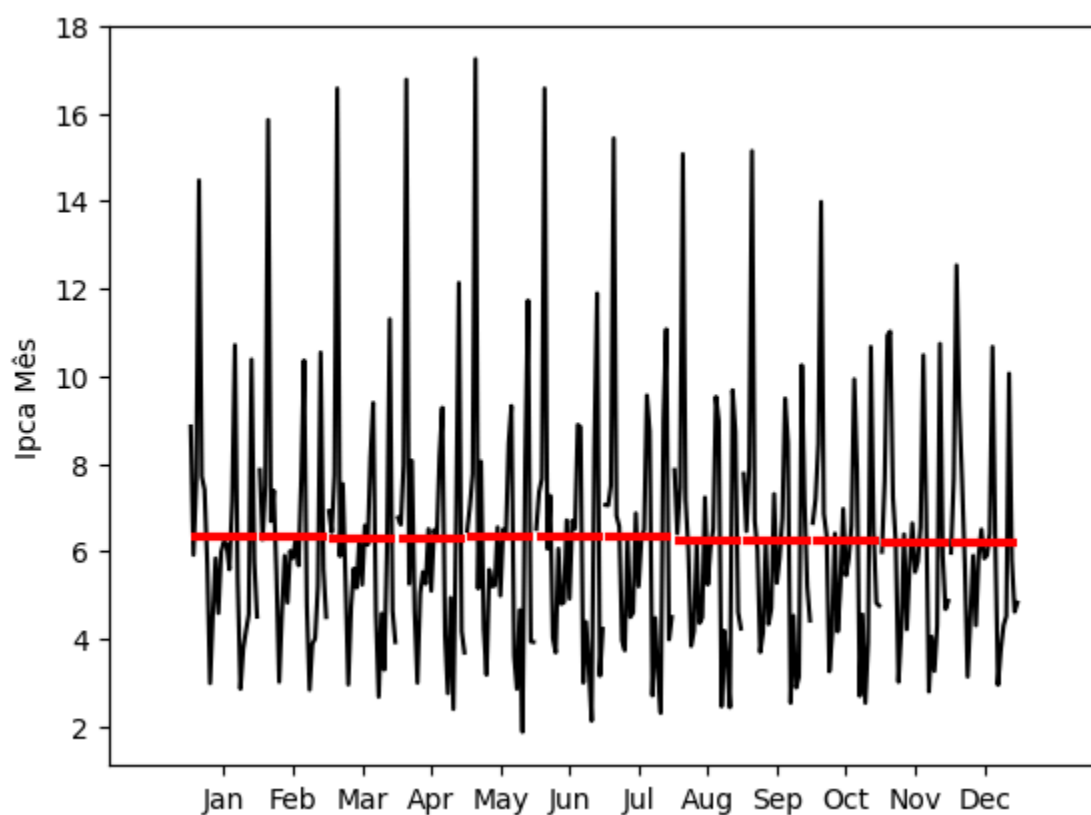


Figura 2.4. Média do Ipca agrupada por mês.

Segundo a Figura 2.4, o Ipca tem um patamar médio mínimo entre os meses de Outubro e Dezembro, mas Novembro tem um leve pico regional. A máxima se dá em Abril.

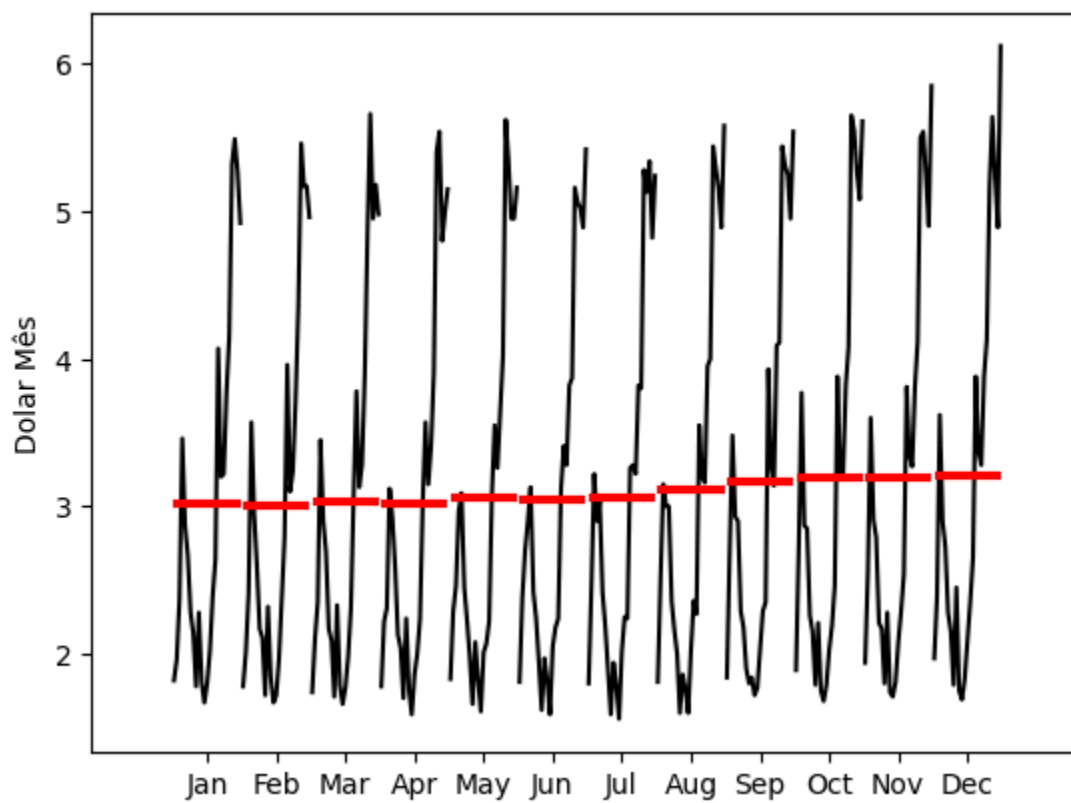


Figura 2.5. Média do Dólar agrupada por mês.

Segundo a Figura 2.5, o dólar tem um patamar médio máximo no final do ano, entre novembro e dezembro. Nos outros meses sofre com algumas oscilações.

A decomposição de séries temporais é uma técnica que separa uma série em componentes fundamentais, como tendência, sazonalidade e resíduos (ou ruído). Essa abordagem ajuda a entender melhor o comportamento da série, permitindo identificar variações estruturais e eliminar a sazonalidade ou tendência para análises mais aprofundadas.

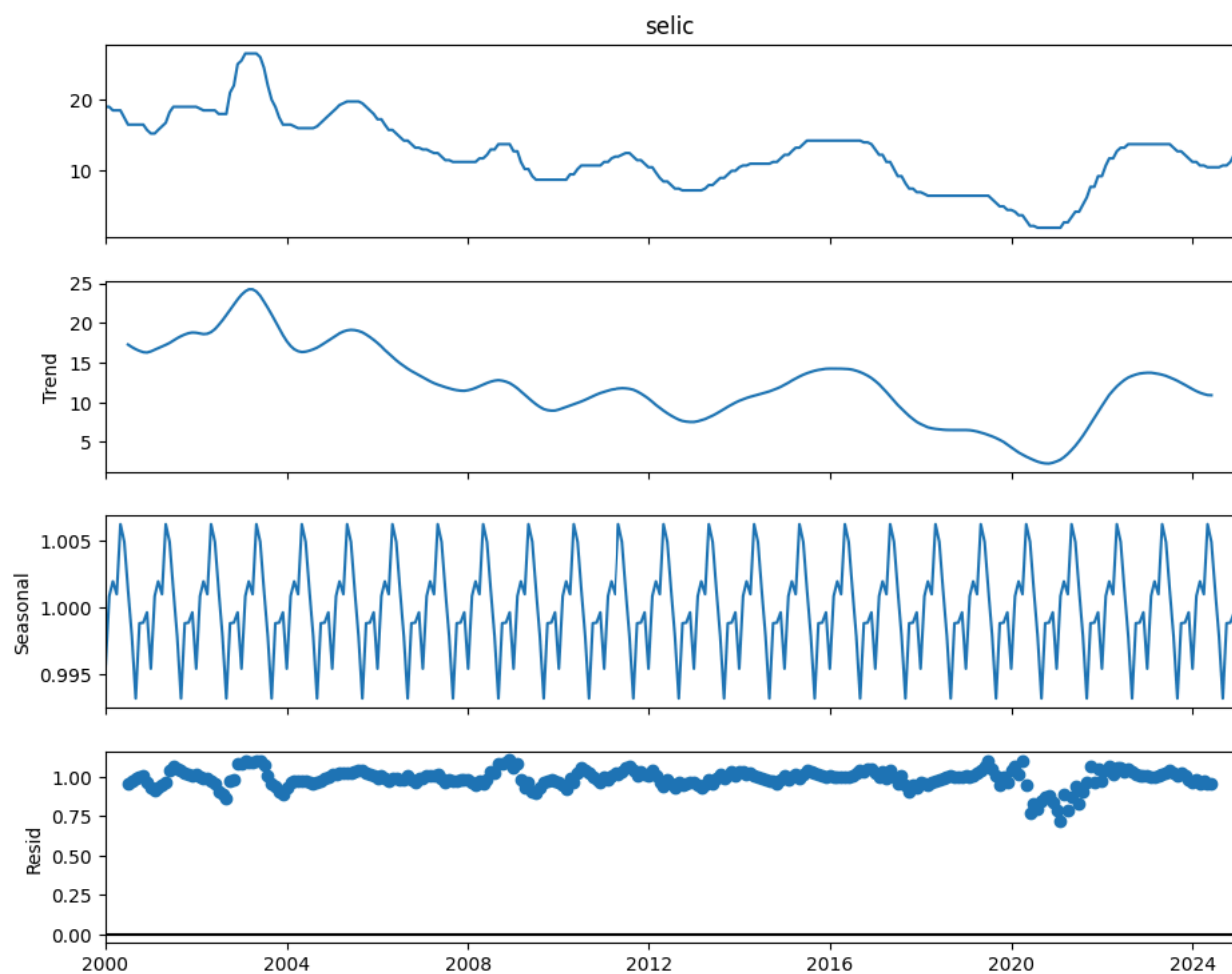


Figura 2.6. Decomposição do Selic.

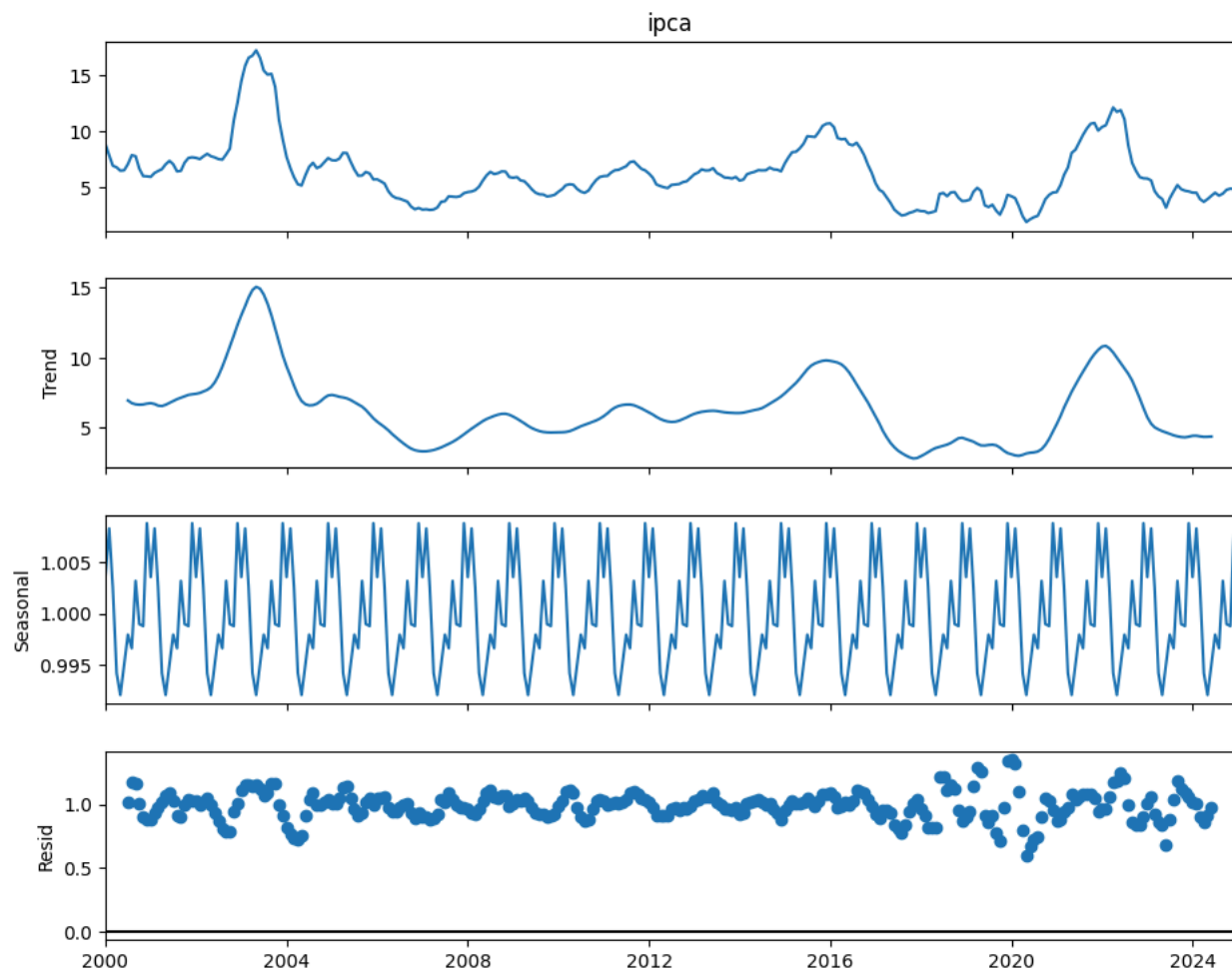


Figura 2.6. Decomposição do Ipca.

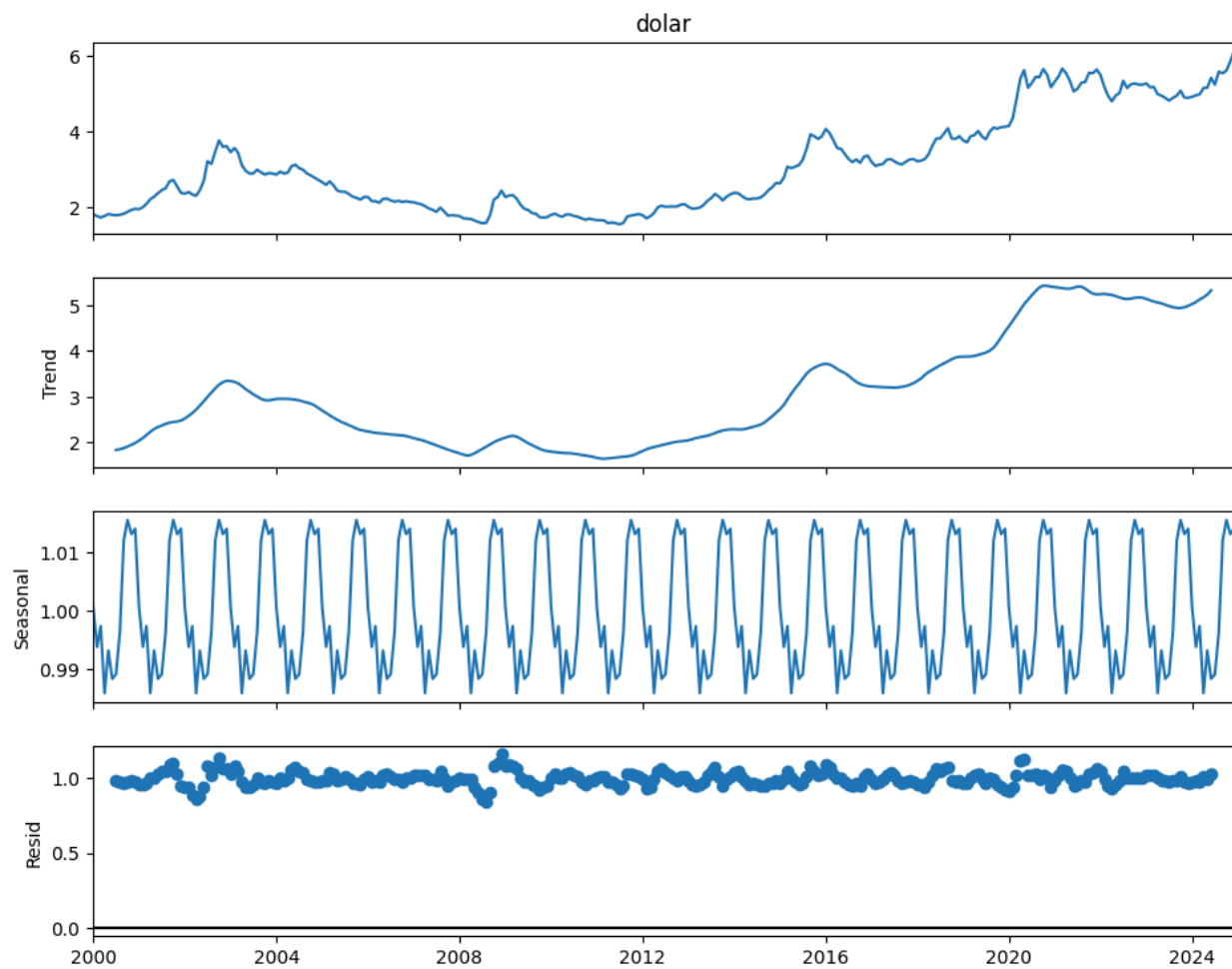


Figura 2.7. Decomposição do Dólar.

ACF (Autocorrelation Function) e PACF (Partial Autocorrelation Function) são ferramentas utilizadas para analisar a dependência temporal em séries temporais. A ACF mede a correlação entre observações em diferentes atrasos (lags), considerando todas as correlações intermediárias, enquanto a PACF isola a correlação direta de uma observação com outra em um determinado lag, eliminando as influências indiretas. Já os lags referem-se ao número de períodos entre os dados observados e são fundamentais para identificar padrões sazonais ou tendências de curto prazo.

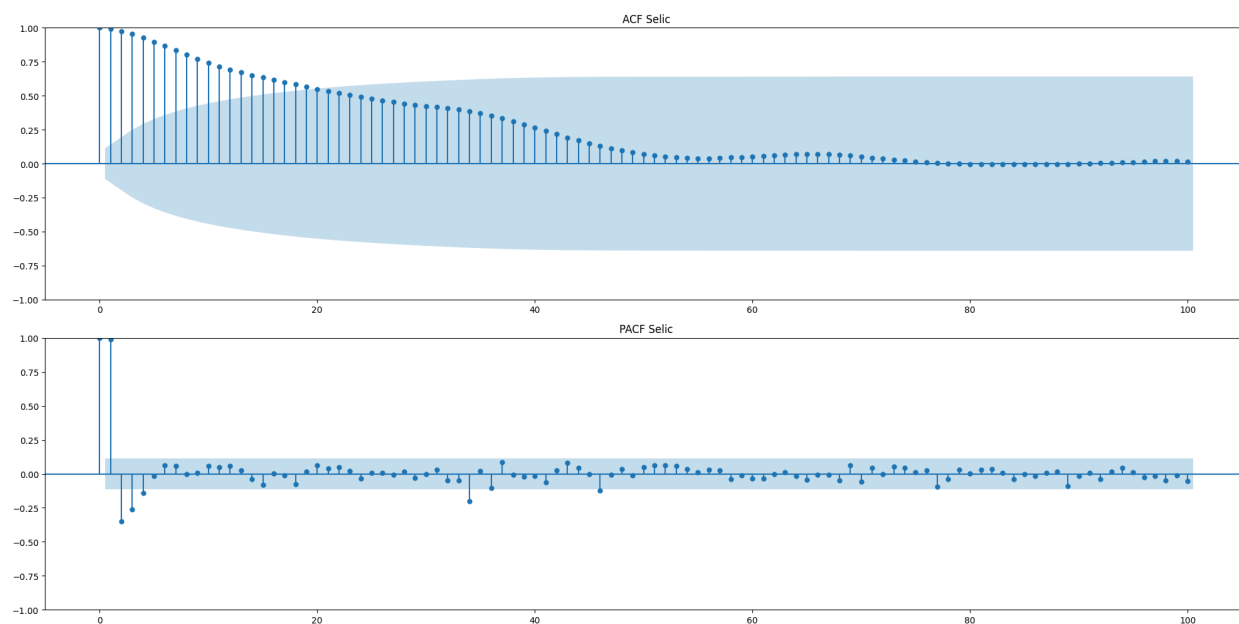


Figura 2.8. Gráficos de ACF e PACF do Selic.

Aparentemente são relevantes os lags 1,2,3,4,5, 34 e 46 no PACF, existe possibilidade de haver correlações.

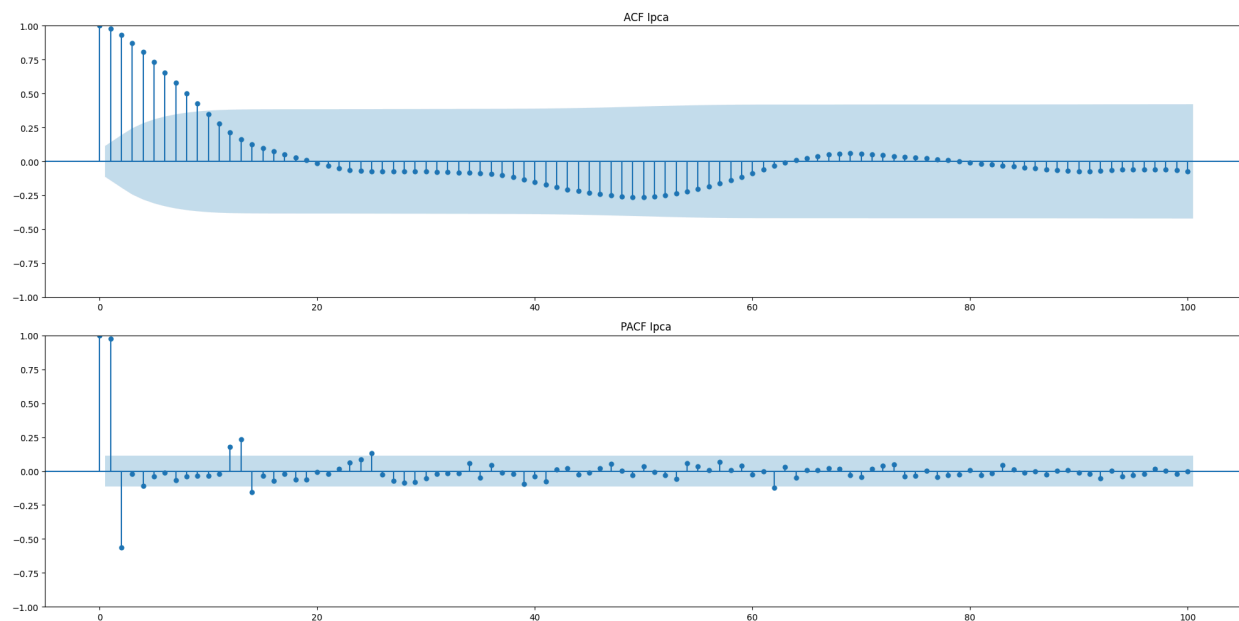


Figura 2.9. Gráficos de ACF e PACF do Ipca.

Aparentemente são relevantes os lags 1, 2, 4, 12, 13, 14, 25, no PACF. Os lags 12 e os seus próximos, indicam uma sazonalidade anual, pois seriam 12 meses.

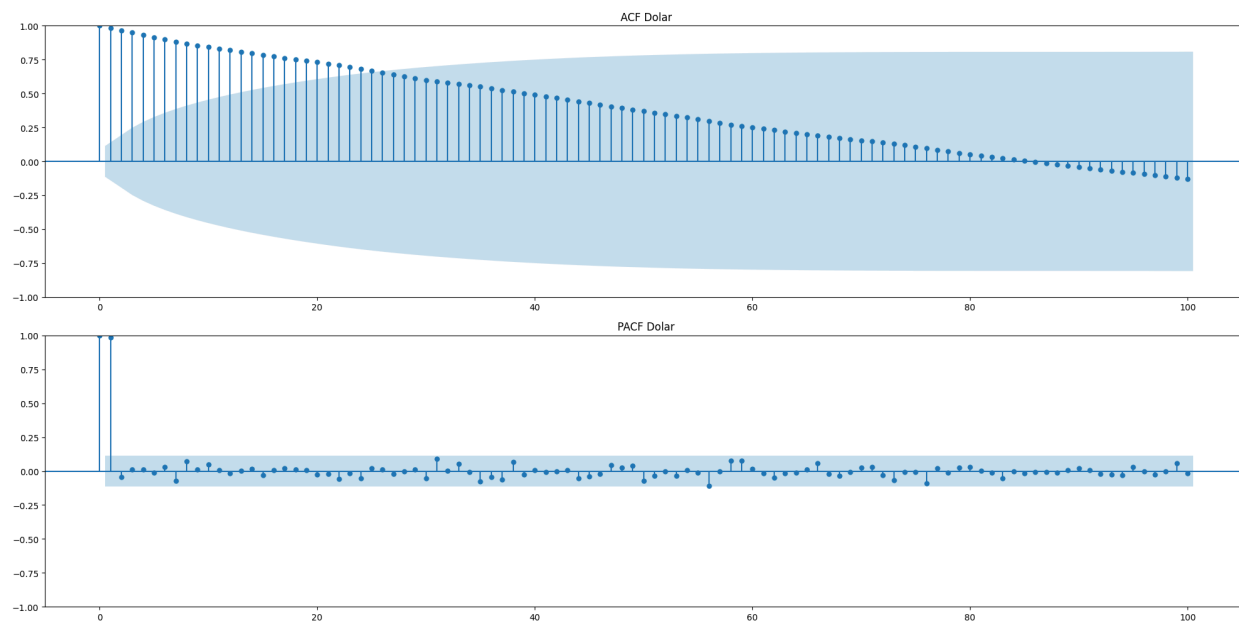


Figura 2.10. Gráficos de ACF e PACF do Dólar.

O dólar aparenta pouca parcial de autocorrelação.

3.Dados Estacionários

O teste Dickey-Fuller (ADF test) é um teste estatístico usado para verificar a estacionariedade de uma série temporal, ou seja, se as propriedades estatísticas da série, como média e variância, permanecem constantes ao longo do tempo. É fundamental para modelos que assumem estacionariedade, como ARIMA, e sua hipótese nula indica que a série não é estacionária.

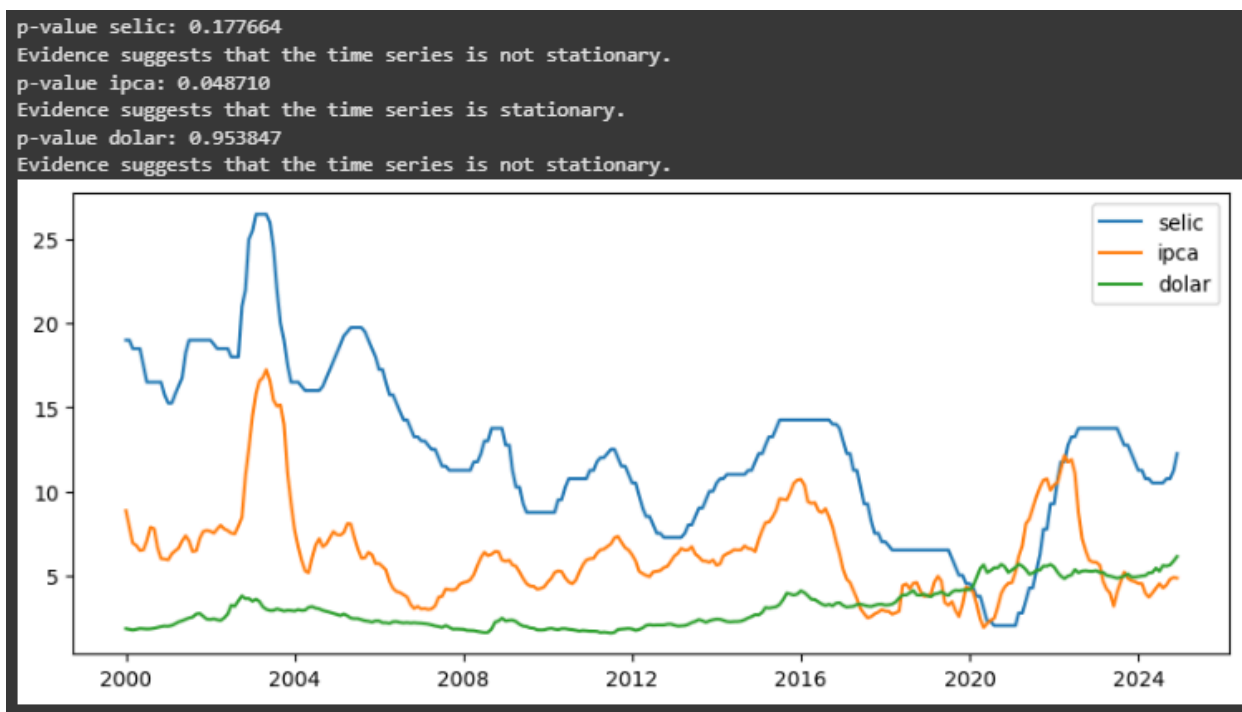


Figura 3.1. Análise se os dados originais são estacionários usando ADF test.

Usando o teste Dickey-Fuller (ADF test) mostra que só o Ipca pode ser considerado estacionário.

A transformação para dados estacionários é um processo que busca estabilizar a média e a variância de uma série temporal, tornando-a apropriada para análise estatística. Técnicas comuns incluem diferenciação (diff), logaritmo e suavização. Essa transformação é essencial para evitar falsos resultados em modelos que pressupõem estacionariedade.

Box-Cox é uma transformação matemática usada para estabilizar a variância e tornar os dados mais próximos da normalidade. Já a diferenciação (diff) é uma técnica que calcula as diferenças entre valores consecutivos de uma série temporal para remover tendências e tornar os dados estacionários.

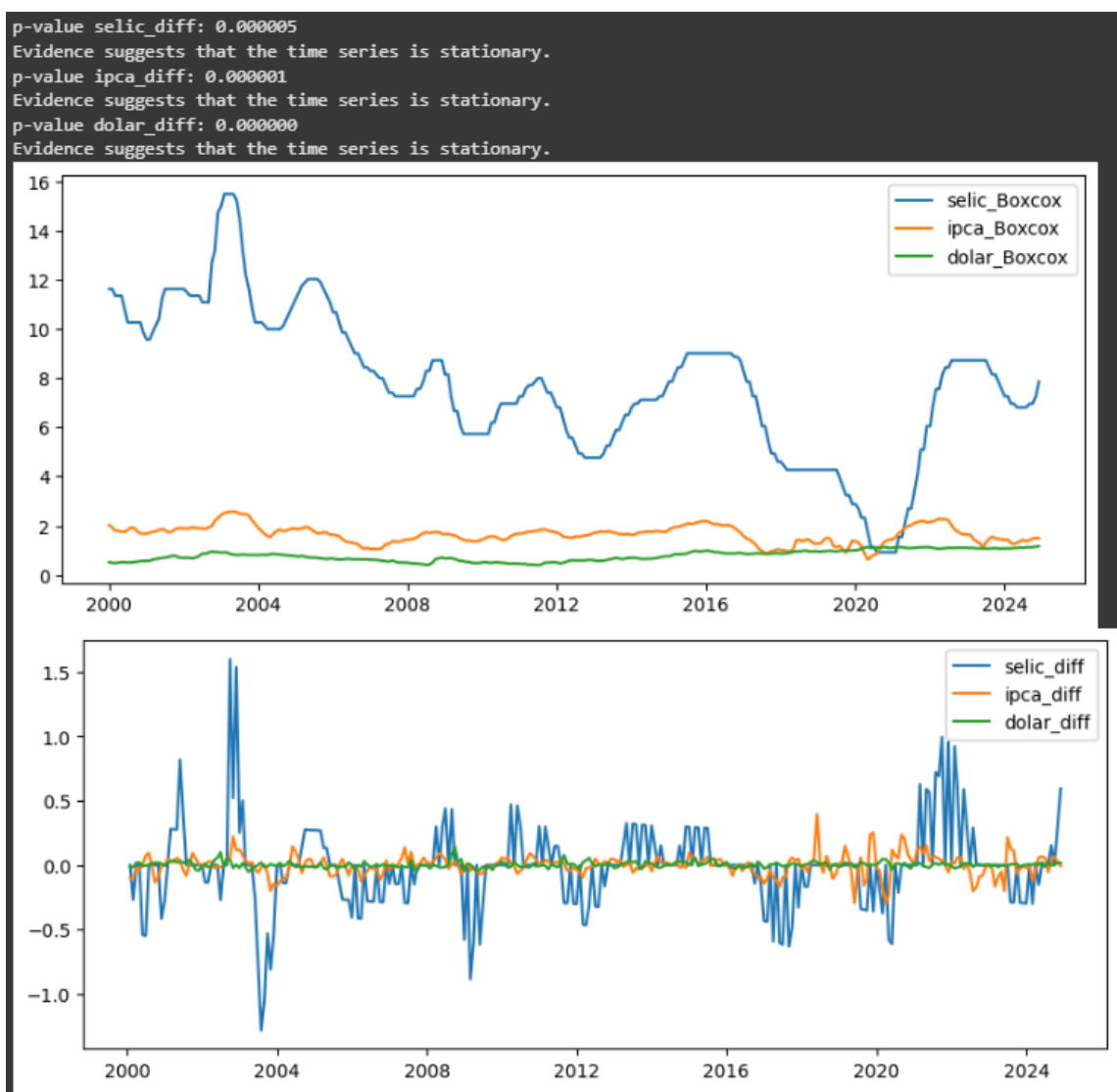


Figura 3.2. Teste ADF em dados aplicados aos métodos de Boxcox e Diff e seus gráficos.

Foi aplicado os métodos de Boxcox e Diff, passando pelo teste ADF, deixando os dados estacionários para as próximas etapas.

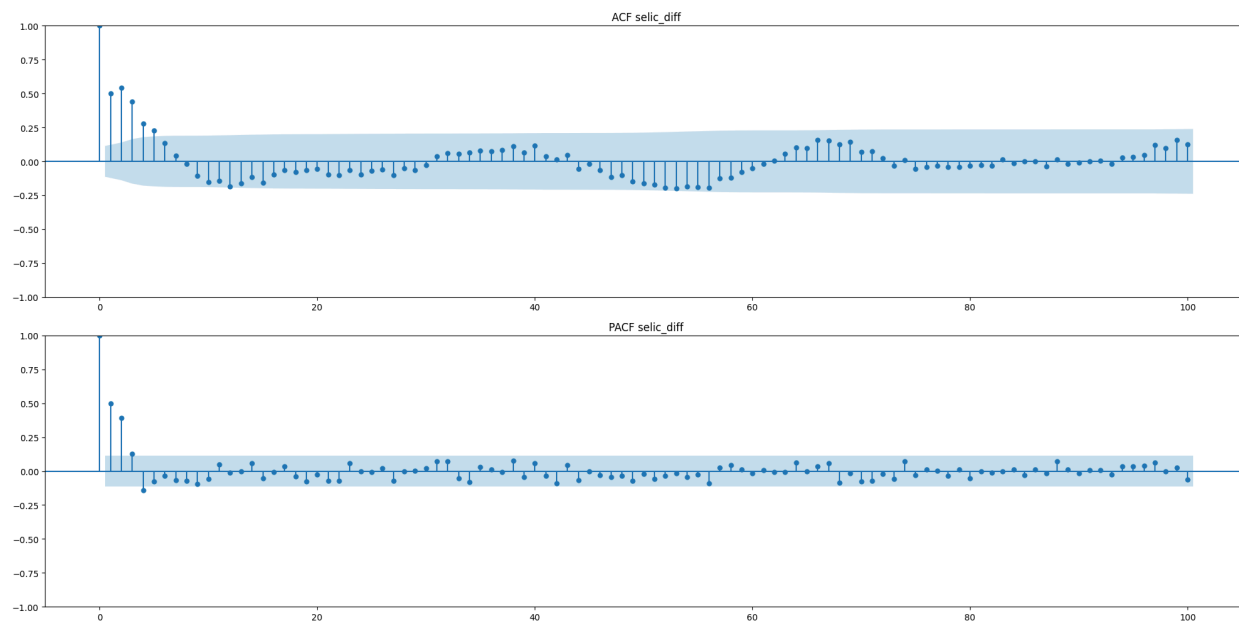


Figura 3.3. ACF e PACF do Selic com dados estacionários.

A Figura 3.3 mostra uma menor auto correlação parcial, com lags maiores para dados estacionários do que para os dados não estacionários. Mostrando que possivelmente os 34 e 46 não são relevantes.

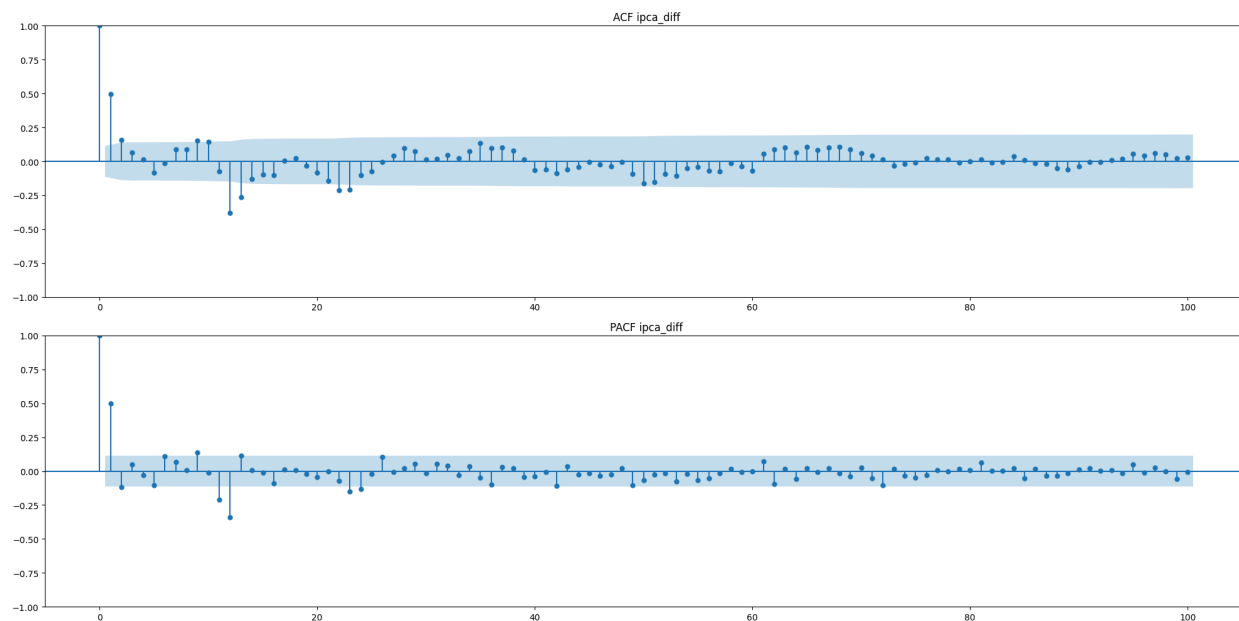


Figura 3.4. ACF e PACF do Ipca com dados estacionários.

A Figura 3.4 corrobora com os dados não estacionários, confirmando uma sazonalidade anual, ou seja de 12 meses.

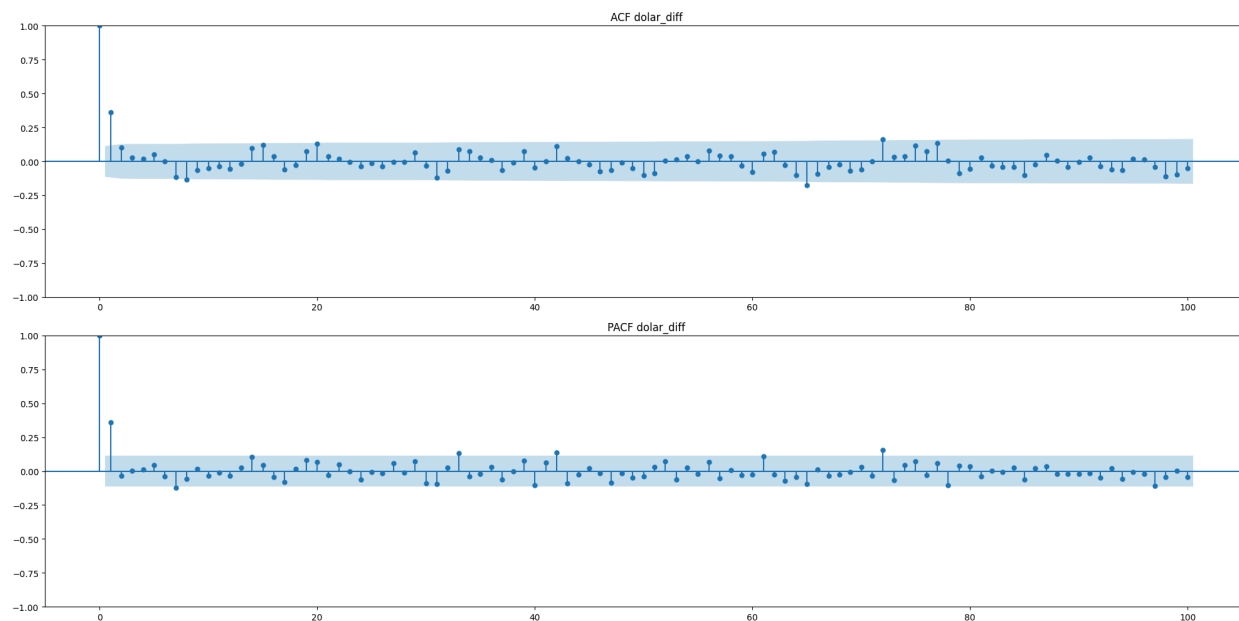


Figura 3.5. ACF e PACF do dólar com dados estacionários.

O dólar, por sua vez, revela apresentar novos lags nos dados estacionários, lags 1, 7, 33, 42 e 72.

4. Modelagem

Primeiramente foi proposto a divisão dos dados, para o treino, na relação de 1 para 4 de teste e treino, mas se mostrou muito fora da realidade, por isso a relação foi ajustada de 1 para 14 de teste e treino. Essa relação se mostra semelhante à utilizada em STATSIM (2025).

SARIMA (Seasonal AutoRegressive Integrated Moving Average) é uma extensão do modelo ARIMA que incorpora componentes sazonais. Ele combina autorregressão (AR), médias móveis (MA) e diferenciação integrada (I) com padrões sazonais para capturar tanto tendências quanto flutuações periódicas em séries temporais.

Os valores de p , d e q no modelo SARIMA representam os parâmetros principais do componente não sazonal. O parâmetro p refere-se à ordem da parte autorregressiva (AR), ou seja, o número de lags anteriores consideradas para prever o valor atual. O parâmetro d indica o número de diferenciações necessárias para tornar a série estacionária. Já q é a ordem do componente de médias móveis (MA), que modela a relação com os erros (resíduos) passados.

Os parâmetros P , D e Q são análogos a p , d e q , mas correspondem aos componentes sazonais do modelo. P é a ordem autorregressiva sazonal, D é o número de diferenciações sazonais necessárias e Q é a ordem sazonal de médias móveis. O parâmetro s representa o número de períodos dentro de um ciclo sazonal (por exemplo, $s=12$ para dados mensais com sazonalidade anual). Esses parâmetros, juntos, permitem ao SARIMA capturar padrões complexos que incluem tanto variações sazonais quanto tendências gerais.

Alguns valores das variáveis do SARIMA, para os modelos, foram ajustados de acordo com os gráficos de ACF e PACF e outros foram ajustados de acordo a palpites baseados em outras tentativas. Foi descartado o uso de parameter tuning por Parameter Grid pois demorava em demasia, o sistema chegava a desconectar dando erro.

Para Selic, a Figura 3.3, sugere valores de $p=4$, $q=5$, $P=0$ e $Q=0$. Para Ipca, a Figura 3.4, sugere valores de $p=2$, $q=2$, $P=1$ e $Q=1$. E finalmente para o dólar, a Figura 3.5, sugere valores de $p=1$, $q=1$, $P=0$ e $Q=0$. Para alguns índices foram criados mais de um modelo com diferentes valores para variáveis com objetivo de diminuir o erro.

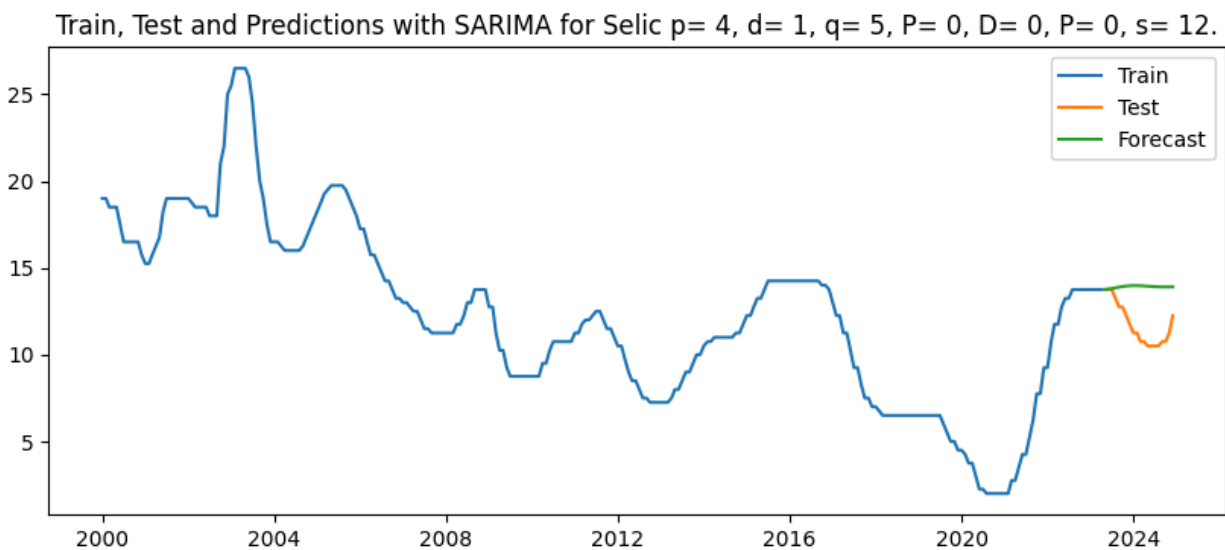


Figura 4.1. Modelagem com SARIMA para Selic.

Os erros obtidos na modelagem da Figura 4.1 para Selic foram de:

- MAE: 2,16
- RMSE: 2,48
- MAPE: 19.54%

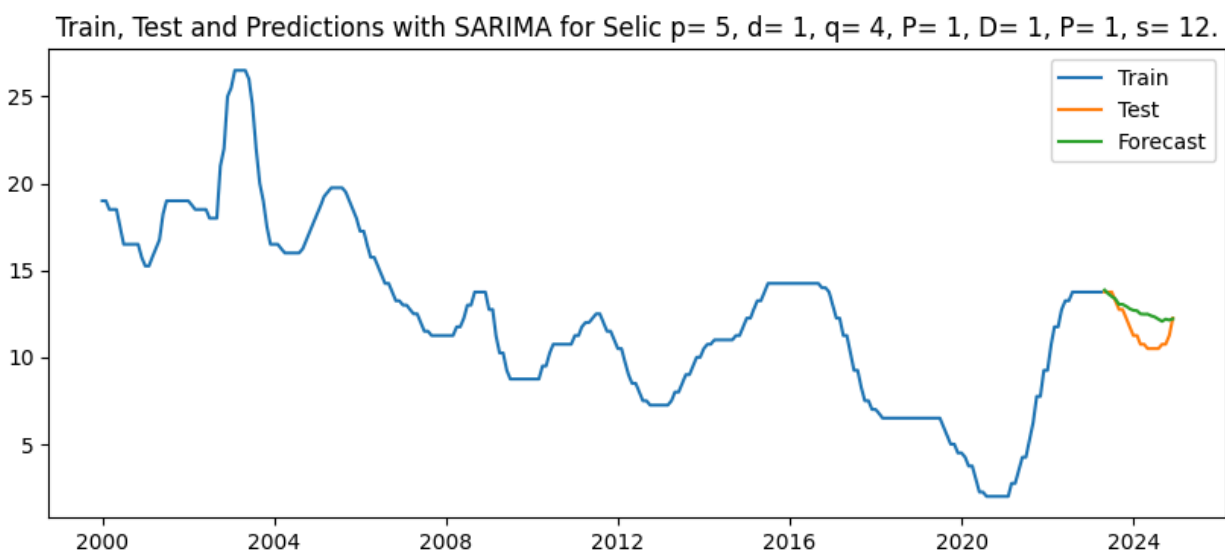


Figura 4.2. Modelagem com SARIMA para Selic com menor erro.

Os erros obtidos na modelagem da Figura 4.2 para Selic foram de:

- MAE: 1,02
- RMSE: 1,23
- MAPE: 9,27%

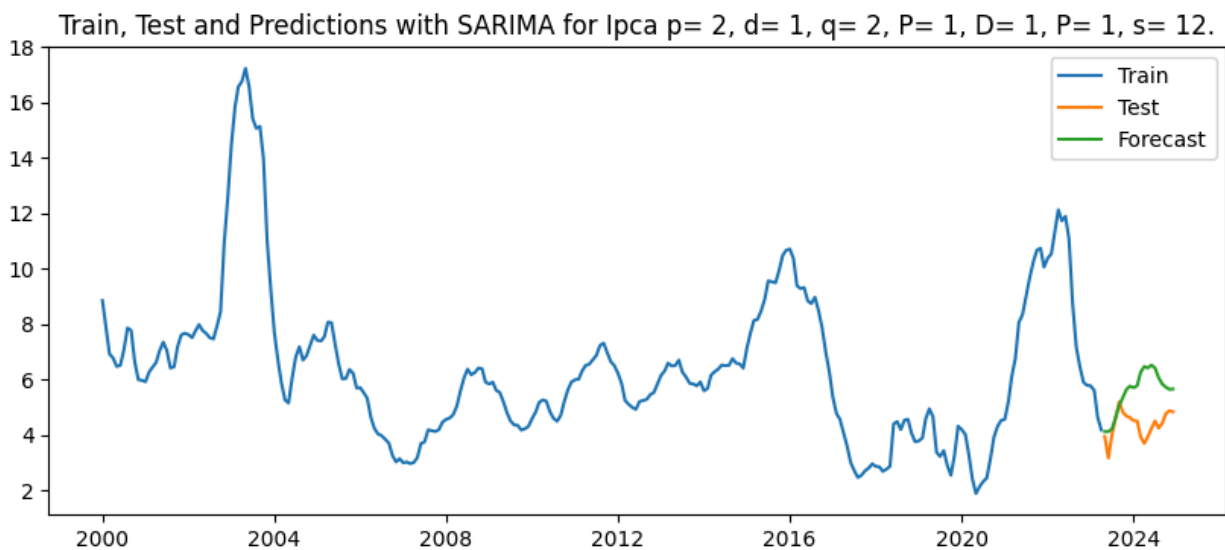


Figura 4.3. Modelagem com SARIMA para Ipca.

Os erros obtidos na modelagem da Figura 4.3 para Ipca foram de:

- MAE: 1,21
- RMSE: 1,45
- MAPE: 28,81%

O erro desse modelo foi suficiente se comparado a outras variantes de modelos para o Ipca.

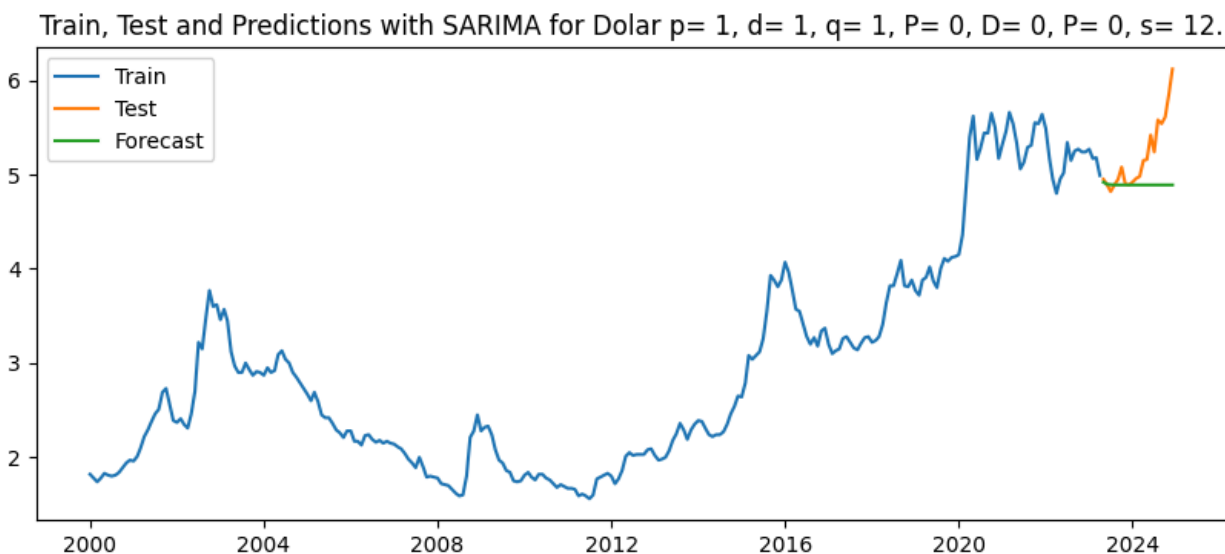


Figura 4.4. Modelagem com SARIMA para Dólar.

Os erros obtidos na modelagem da Figura 4.4 para Dólar foram de:

- MAE: 0,31
- RMSE: 0,47
- MAPE: 5,58 %

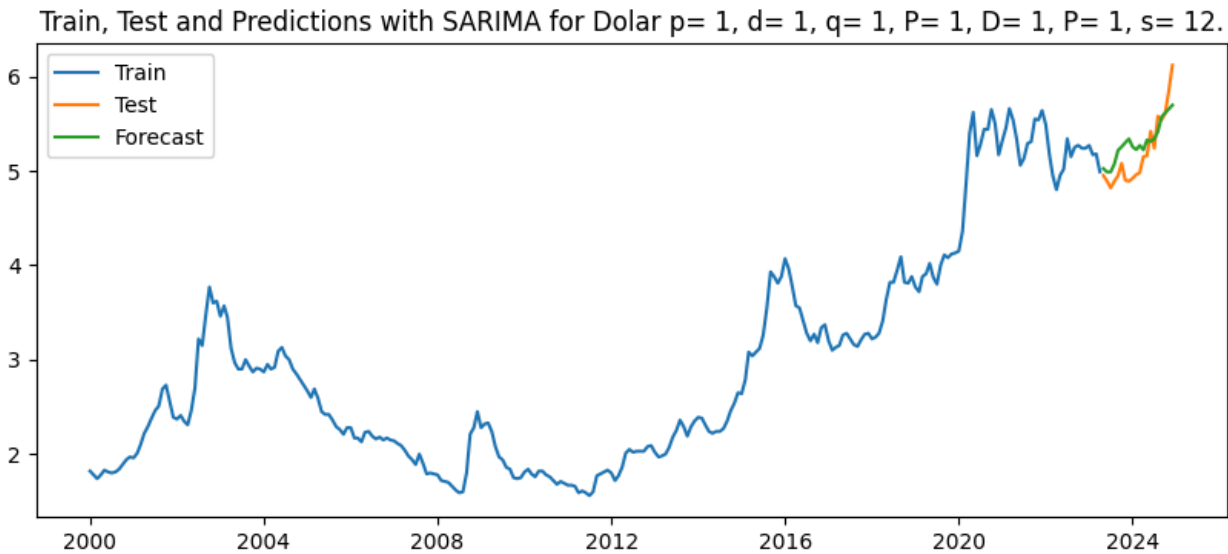
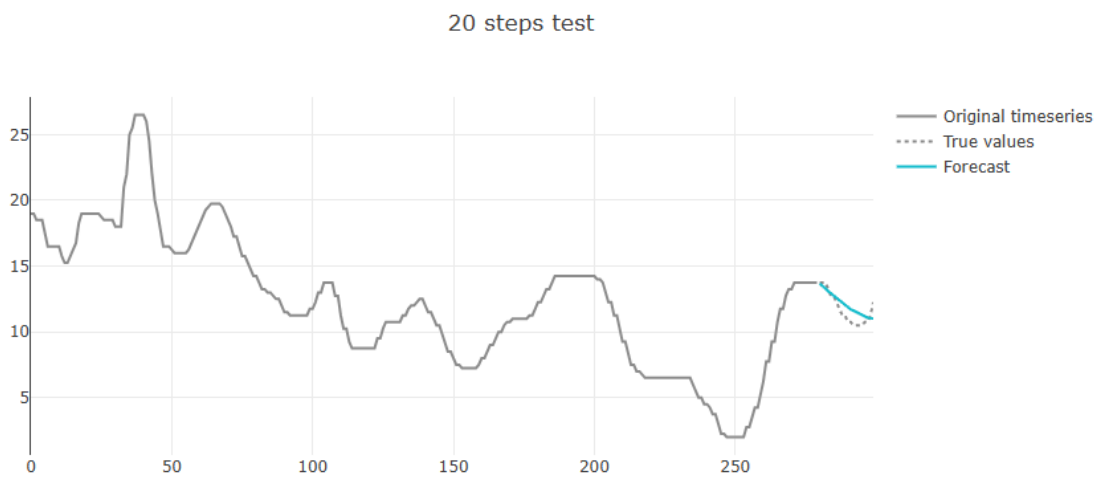


Figura 4.5. Modelagem com SARIMA para Dólar com menor erro.

Os erros obtidos na modelagem da Figura 4.5 para Dólar foram de:

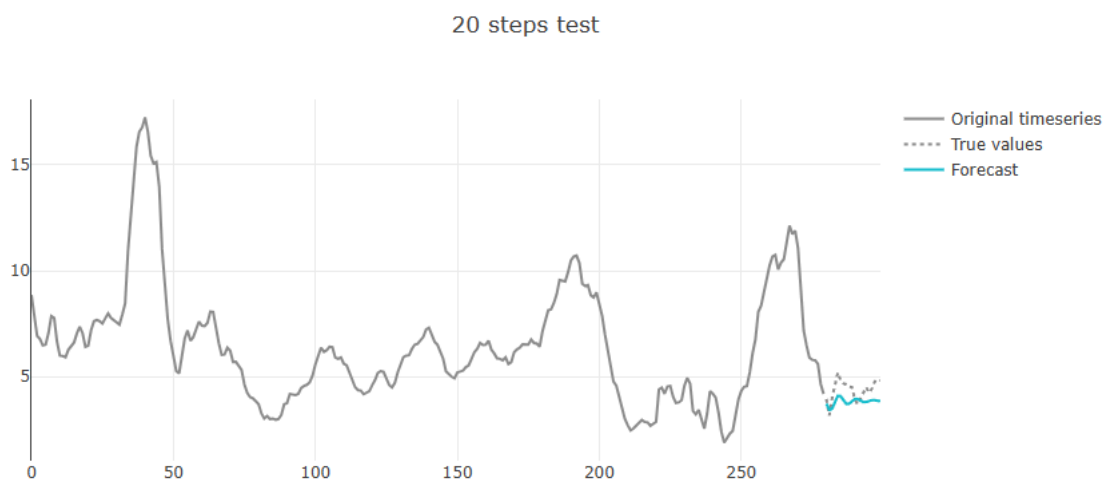
- MAE: 0,20
- RMSE: 0,24
- MAPE: 3,86%

Também foi utilizado a ferramenta em STATSIM (2025) para os 3 índices.



p:8, d: 1, q: 3

Figura 4.6. Modelagem com STATSIM para Selic.



p:4, d: 1, q: 4

Figura 4.7. Modelagem com STATSIM para Ipca.



p:9, d: 2, q: 2

Figura 4.8. Modelagem com STATSIM para Dólar.

5. Predição

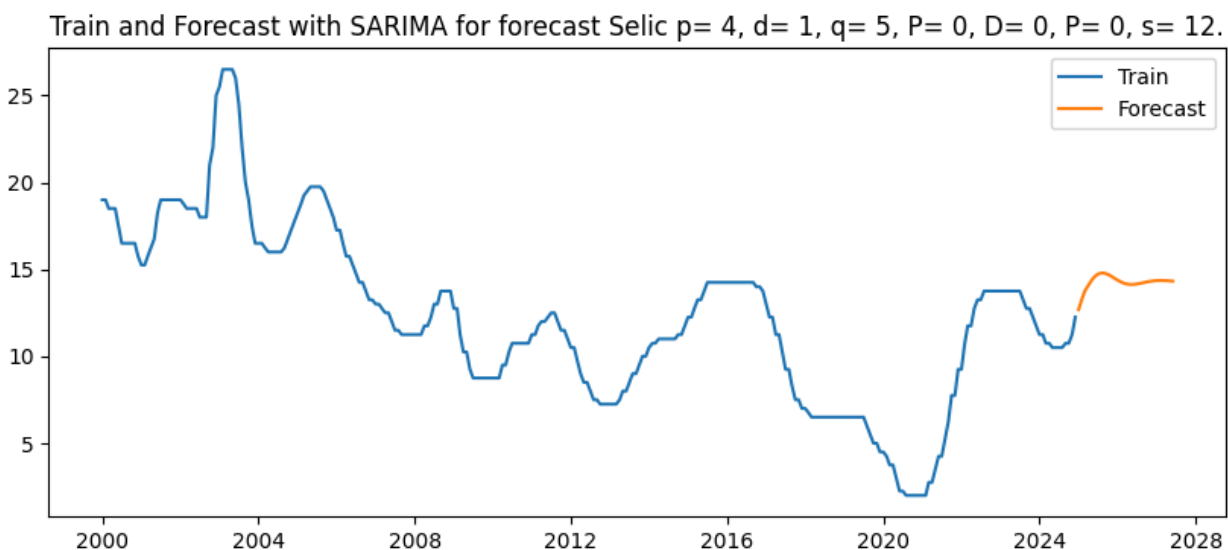


Figura 5.1. Predição para o primeiro modelo de Selic

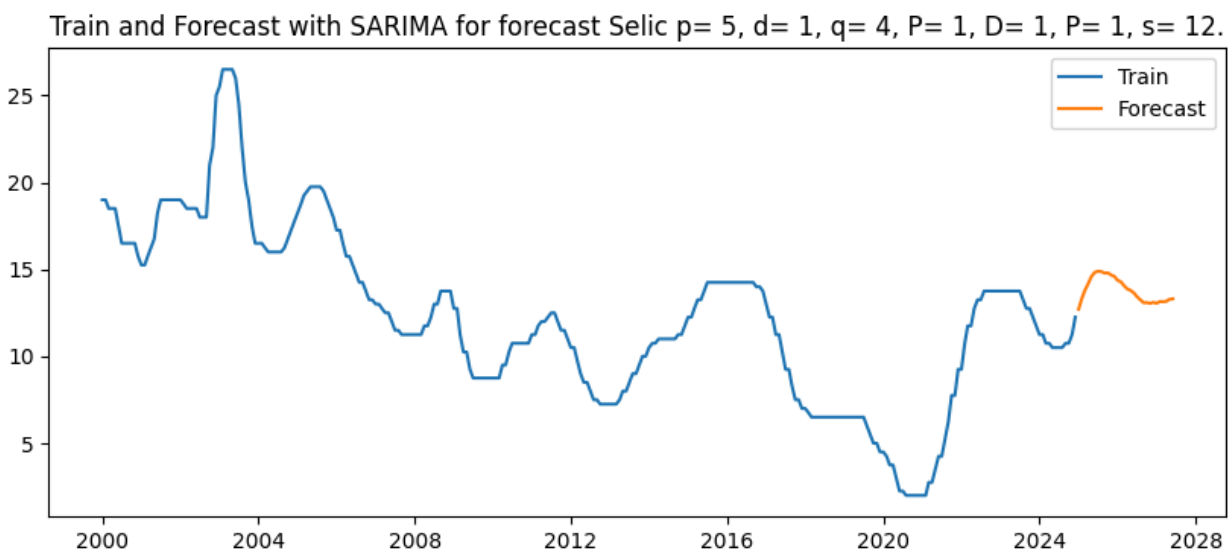


Figura 5.2. Predição para o segundo modelo de Selic

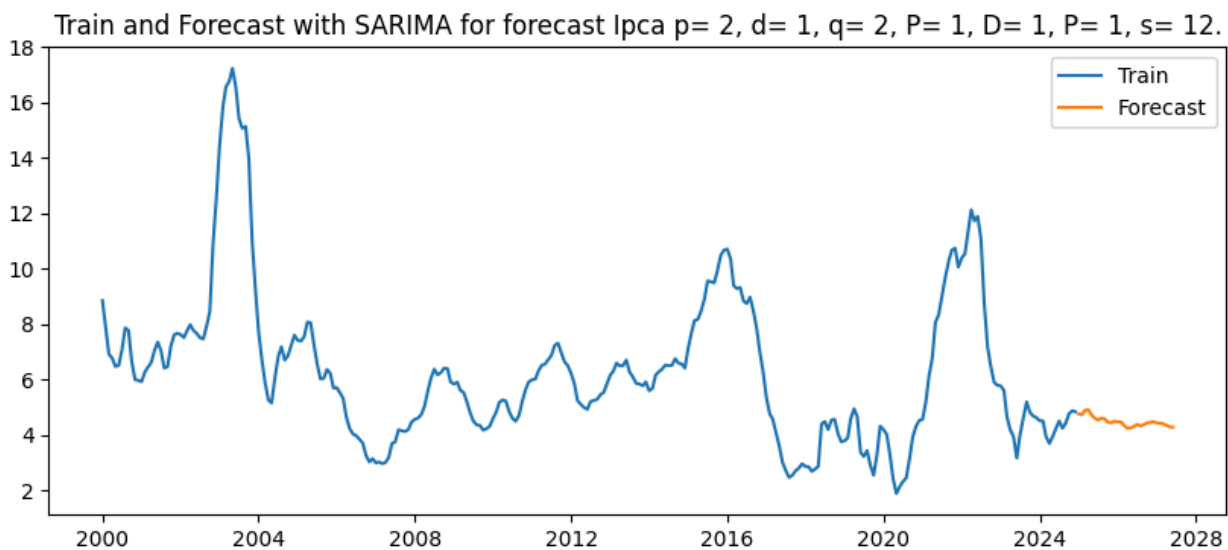


Figura 5.3. Predição para o modelo de Ipca

Train and Forecast with SARIMA for forecast Dolar $p=1, d=1, q=1, P=0, D=0, P=0, s=12$.



Figura 5.4. Predição para o primeiro modelo de Dólar

Train and Forecast with SARIMA for forecast Dolar $p=1, d=1, q=1, P=1, D=1, P=1, s=12$.



Figura 5.5. Predição para o segundo modelo de Dólar

Também foi utilizado a ferramenta em STATSIM (2025) para predição dos 3 índices.

20 steps forecast

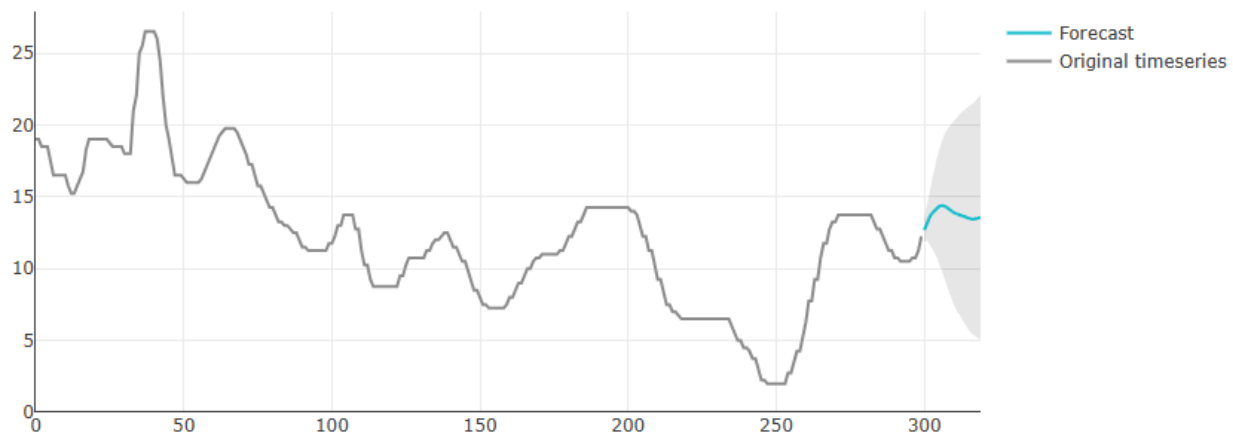


Figura 5.6. Predição do Selic usando a ferramenta em STATSIM (2025)

20 steps forecast

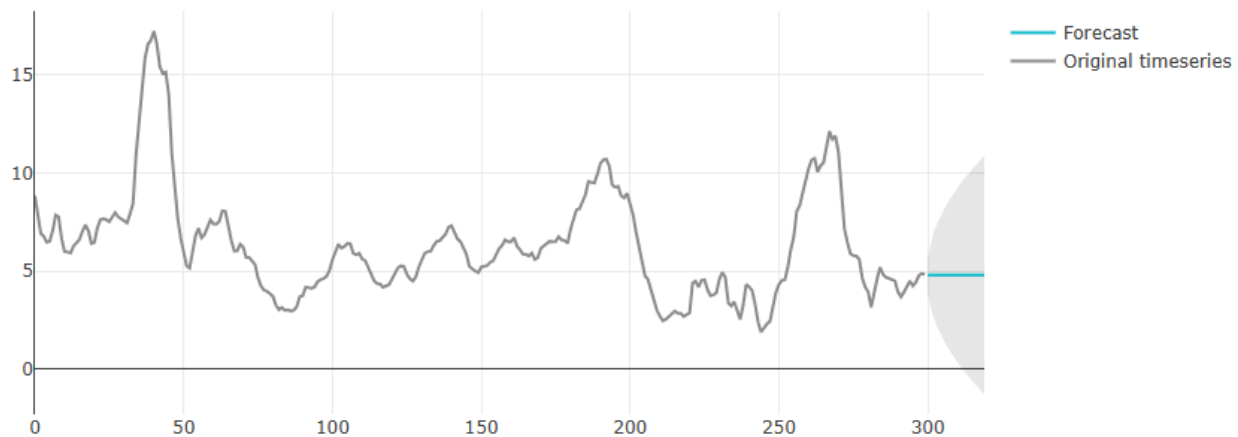


Figura 5.7. Predição do Ipca usando a ferramenta em STATSIM (2025)

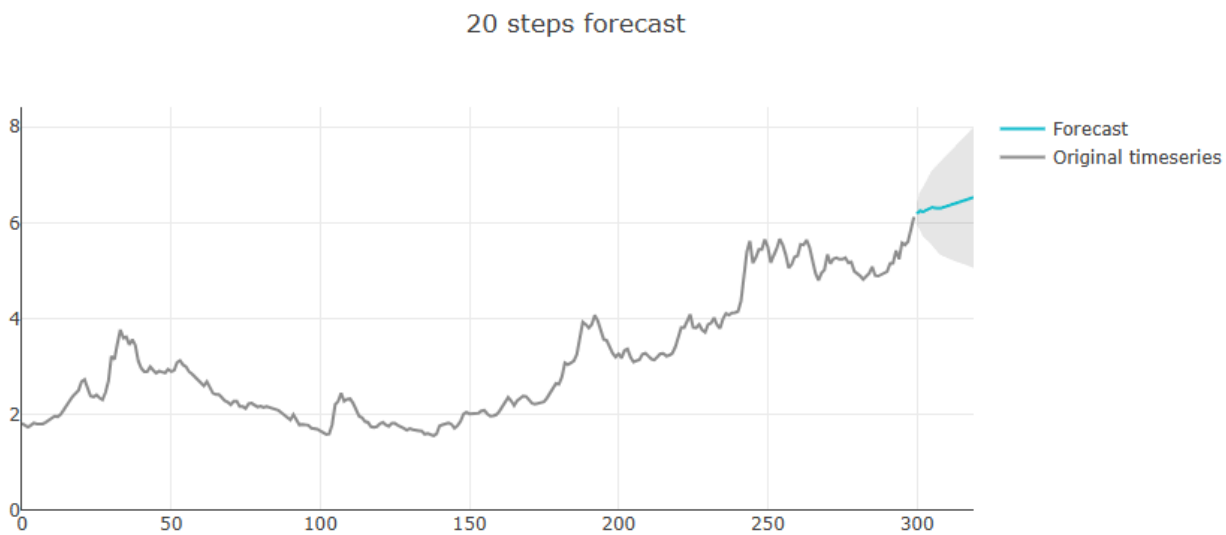


Figura 5.8. Predição do Dólar usando a ferramenta em STATSIM (2025)

6. Conclusão

Utilizando o teste de correlação de Pearson, evidenciou-se uma correlação positiva de 0,61 entre a Selic e o IPCA. No entanto, o dólar não apresentou correlação significativa. Também foram identificados padrões sazonais: a Selic tende a cair próximo a agosto, o IPCA diminui no final do ano, e o dólar se mantém relativamente estável, com aumentos no fim do ano.

Os gráficos de ACF e PACF foram essenciais para determinar os parâmetros do modelo SARIMA, sendo necessário aplicar métodos para tornar a série temporal estacionária. As modelagens foram realizadas utilizando SARIMA e um programa externo, resultando em modelos tanto satisfatórios quanto insatisfatórios.

É importante destacar as limitações do sistema, que utiliza modelos matemáticos para previsão sem considerar aspectos de uma análise fundamentalista. Trabalhos futuros podem incluir o refinamento dos parâmetros dos modelos ou a exploração de abordagens alternativas, como redes neurais recorrentes (RNN) e LSTM, desde que haja dados suficientes para treinamento.

7.Referências

BONNO, Simone Jager Patrocinio. *Previsão de inflação utilizando modelos de séries temporais*. 2014. Dissertação (Mestrado em Finanças e Economia Empresarial) – Fundação Getúlio Vargas, Escola de Pós-Graduação em Economia, Rio de Janeiro, 2014.

RECIERI. *SELIC x IPCA x Dólar: série histórica e conexões*. Disponível em: <https://recieri.com/selic-x-ipca-x-dolar-serie-historica-e-conexoes/>. Acesso em: 26 jan. 2025.

STATSIM. Statistical Simulation and Inference in the Browser. Disponível em: <https://statsim.com/forecast/>. Acesso em: 26 jan. 2025.

MEDIUM. Dicas para criar um modelo de previsão de séries Temporais <https://medium.com/techbloghotmart/dicas-para-criar-um-modelo-de-previs%C3%A3o-de-s%C3%A9ries-temporais-d4bb2e32e148>. Acesso em: 26 jan. 2025.

HOWELL, Egor. *Time-Series Crash Course*. GitHub, 2025. Disponível em: <https://github.com/egorhowell/Youtube/tree/main/Time-Series-Crash-Course>. Acesso em: 26 jan. 2025.

RESENDE, Diogo Alves de. *Time Series with Deep Learning (LSTM, TFT, N-BEATS), GenAI (Amazon Chronos), Prophet, Silverkite, ARIMA. Demand Forecast*. Udemy, 2025. Disponível em: <https://www.udemy.com/course/forecasting-python/learn/lecture/45336399#overview>. Acesso em: 26 jan. 2025.