

Considerações sobre a Modelagem de Séries Temporais

Anderson Falcão¹
Pedro Tokar¹

Antonio Brych¹
Tomás Lira¹

Gabriel Abad¹
Vitor Moreira¹

¹Escola de Matemática Aplicada – FGV EMAp

1 Introdução

Este trabalho visa modelar e prever a variável ‘volume’ contida no conjunto de dados fornecido, usando as variáveis auxiliares ‘users’ e ‘inv’. A série temporal abrange o período de 2022-10-31 a 2025-10-27 com intervalos semanais, totalizando 150 semanas de observação.

1.1 Exploração inicial do dados

	volume	inv	users
mean	4.820382	0.712956	6.576834
std	5.162304	0.889207	6.596819
min	0.140000	0.000240	0.401000
25%	0.660000	0.076311	1.351000
50%	3.250000	0.464850	3.847000
75%	7.550000	0.859225	9.768000
max	24.52000	5.623875	29.33200

Tabela 1: Descrição Estatística dos Dados

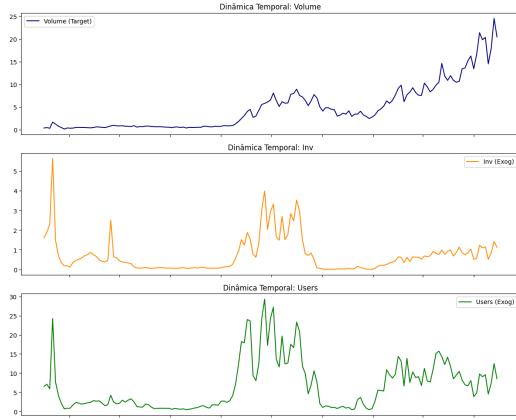


Figura 1: Line plot da dinâmica temporal das variáveis

O lineplot das variáveis em relação ao tempo já começa mostrando padrões interessantes, principalmente para as novas covariáveis. Observamos que o comportamento delas é semelhante durante 2023 e durante os três primeiros trimestres de 2024, apresentando valor baixo e aparentemente constante em 2023 e tendo um aumento em 2024. Apesar dos valores serem diferentes, os momentos de alta das séries parecem sempre ocorrer no mesmo período.

Além disso, o aumento delas também segue o aumento da série volume, o que já pode ser visto como motivação para se incluir as covariáveis na modelagem. Essas observações empíricas podem ser confirmadas quando analisamos a correlação entre as variáveis:

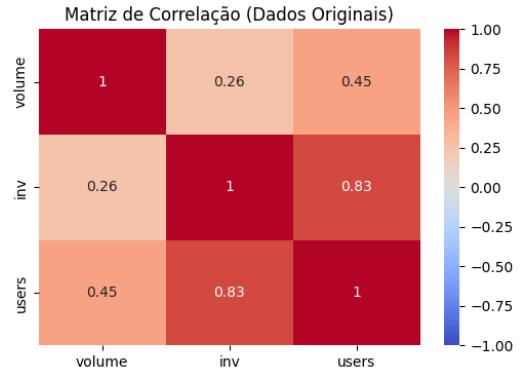


Figura 2: Correlações entre as variáveis

Seguindo a mesma motivação da entrega anterior (heterocedastidade), podemos aplicar uma transformação logarítmica nas variáveis:

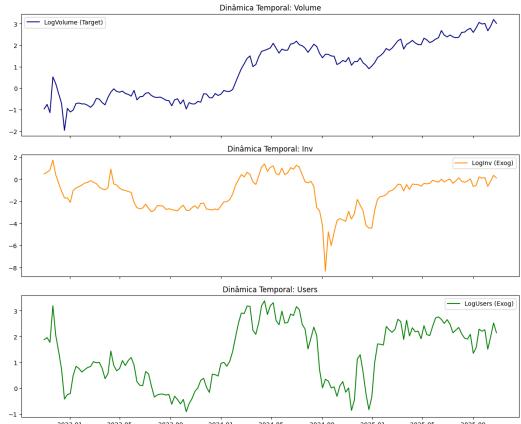


Figura 3: Line plot da dinâmica temporal das variáveis transformadas com log

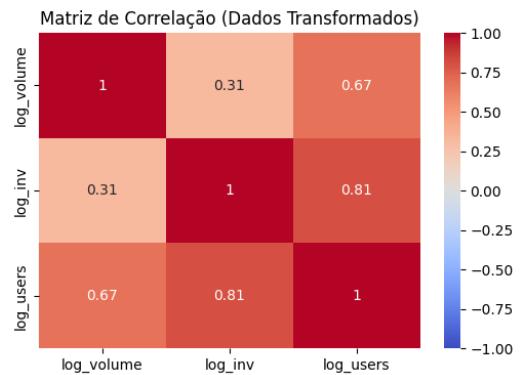


Figura 4: Correlações entre as variáveis transformadas com log

A correlação linear entre as covariáveis e “volume” aumenta se aplicarmos uma transformação logarítmica nelas, isso pode ser interessante para construir os modelos para ter melhores resultados. Vale observar, porém, que a correlação entre as variáveis `inv` e `users` é alta, o que pode ser prejudicial para alguns modelos, já que as informações de cada coluna serão quase colineares.

2 Modelos baseline

Pela própria natureza dos modelos baseline utilizados, nenhuma de suas formulações aceitam covariáveis. À luz disso, esperamos um comportamento similar ao que foi visto anteriormente.

Continuamos adotando o valor unitário para o passo mas desta vez com horizonte igual a 4, isto é, com nosso modelo prevendo 4 semanas consecutivas, sustentado por uma janela de contexto de 52 semanas.

O horizonte de previsão foi ampliado em relação ao trabalho anterior para evitar a superestimação sistemática do desempenho do modelo Naive, uma distorção que ocorria ao utilizar apenas um passo à frente ($h = 1$).

Realizamos um grid search para determinar o melhor hiperparâmetro para o modelo **Seasonal Naive**:

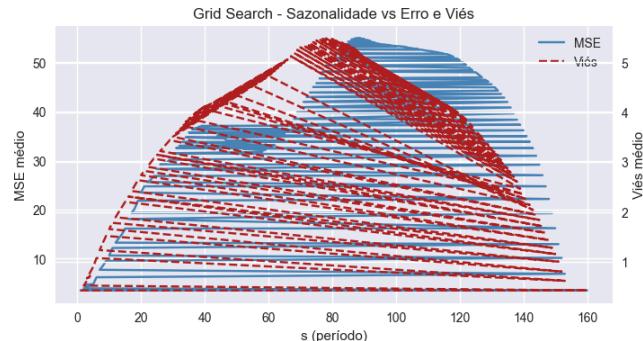


Figura 5: Grid-Search para modelo Seasonal Naive

O gráfico acima sugere que o melhor modelo para esses dados é o Naive simples (olhar apenas para o valor imediatamente anterior, $s = 1$), pois tentar buscar padrões sazonais mais longos apenas aumentou o erro.

Mo- delo	R ²	MAE	MSE	RMSE	Média Resí- duos (\bar{y})
Naive	0.849886	1.330956	3.526443	1.877883	0.431838
Sea- sonal Naive	0.849886	1.330956	3.526443	1.877883	0.431838
Mean	-0.603601	4.651042	37.671449	6.137707	4.648174
Drift	0.853469	1.327347	3.442281	1.855339	0.299659

Tabela 2: Métricas de desempenho dos modelos baseline com Seasonal Naive $s = 1$.

Considerando as métricas de resumo acima, observamos que os modelos Drift e Seasonal Naive apresentaram os resultados mais aceitáveis, sendo o primeiro superior por exibir menor RMSE e maior R^2 .

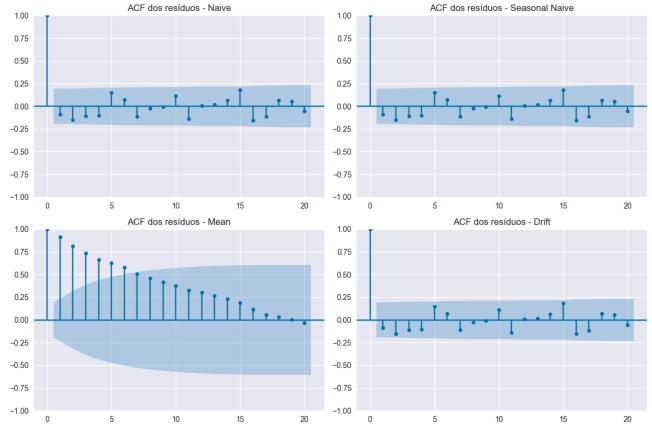


Figura 6: Avaliação usando o ACF para cada modelo

Tanto o Naive quanto o Seasonal Naive apresentaram bons resultados, com resíduos sem padrões fortes e contidos na margem de segurança, sendo que o Seasonal acabou se comportando de forma idêntica ao Naive pelo fato de seu ciclo ótimo ter sido igual a 1.

Por outro lado, o modelo Mean mostrou-se um baseline ruim ao exibir forte autocorrelação e barras grandes que evidenciam sua falha em explicar a dinâmica da série, enquanto o Drift se confirmou como a melhor opção, gerando resíduos comportados.

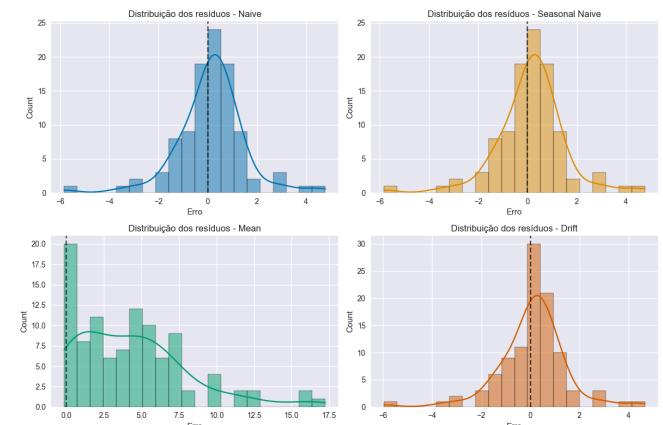


Figura 7: Comparação dos histogramas de resíduos para cada modelo

A análise dos histogramas de resíduos mostram que o Naive e o Seasonal Naive apresentaram comportamentos muito semelhantes e positivos, com distribuições razoavelmente simétricas e centradas, indicando estabilidade e baixo viés.

O modelo Mean exibiu forte assimetria com resíduos concentrados à direita do zero, demonstrando que ele subestima os valores reais. Por fim, o Drift destacou-se novamente como a melhor opção, apresentando uma distribuição mais concentrada em zero e com menor dispersão que os demais.

2.1 Conclusão

Diferentemente do restante do trabalho, onde adotamos amplamente um horizonte curto ($h=1$) e o modelo Naive predominava, a ampliação do horizonte nesta etapa revelou um comportamento distinto. Neste cenário de longo prazo, o Drift apresentou desempenho superior, pois é o único baseline capaz de projetar a tendência da série, ao contrário do Naive, que apenas replica o último valor observado e perde precisão conforme o horizonte cresce.

3 Diferenciação

A análise de diferenciação é uma etapa crítica na modelagem de séries temporais, especialmente ao utilizar modelos da família

ARIMA. A principal utilidade deste processo é transformar uma série temporal não estacionária em estacionária.

Uma série é considerada estacionária quando suas propriedades estatísticas: como média, variância e autocorrelação permanecem constantes ao longo do tempo. Isso é fundamental porque:

1. **Estabilidade Preditiva:** Modelos estatísticos clássicos assumem que o comportamento passado (média e variância) se repetirá no futuro. Tendências e sazonais violam essa premissa.
2. **Remoção de Tendência e Sazonalidade:** A diferenciação remove componentes determinísticos (como o crescimento contínuo de vendas) e ciclos repetitivos (padrões anuais), isolando o componente estocástico (ruído e sinal de curto prazo) que o modelo deve prever.
3. **Evitar Correlações Espúrias:** Trabalhar com dados em nível (sem diferenciar) pode levar a regressões espúrias, onde duas variáveis parecem correlacionadas apenas porque ambas crescem ao longo do tempo, sem haver causalidade real.

3.1 Metodologia

Para esta análise, aplicou-se inicialmente uma **transformação logarítmica** nas variáveis (volume, inv, users) para estabilizar a variância (heterocedasticidade) e linearizar relações exponenciais.

Posteriormente, a estacionariedade foi verificada utilizando o teste **Augmented Dickey-Fuller (ADF)**.

- **Hipótese Nula (H_0):** A série possui uma raiz unitária (é não estacionária).
- **Hipótese Alternativa (H_1):** A série é estacionária.
- **Critério:** Rejeita-se H_0 se o p-valor for menor que 5%.

3.2 Resultados

Abaixo estão os resultados obtidos para a variável alvo (`log_volume`) em diferentes ordens de diferenciação:

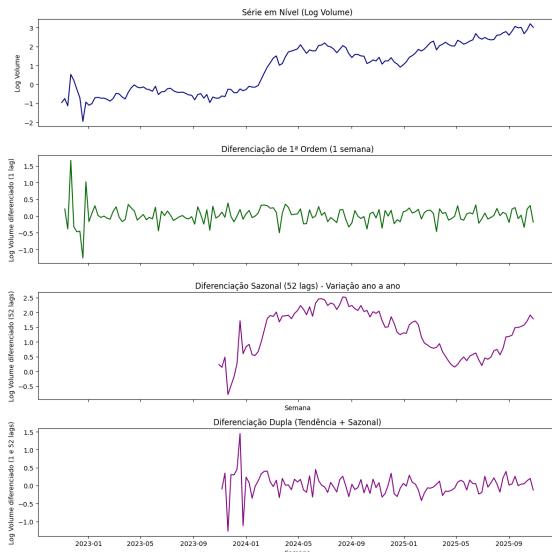


Figura 8: Line plot com as diferentes diferenciações avaliadas

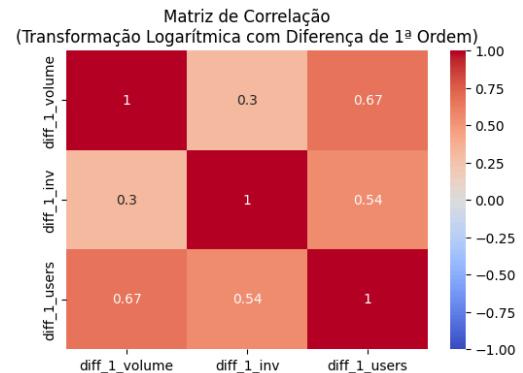


Figura 9: Matriz de correlação entre as variáveis com diferenciação de primeira ordem

Série em Nível (Log Volume)

- Estatística ADF: -1.1603
- P-valor: 0.6904
- Conclusão: A série original não é estacionária. O p-valor alto indica a presença clara de uma tendência ou raiz unitária, confirmando a necessidade de diferenciação.

Primeira Diferença no Log Volume ($\Delta y_t = y_t - y_{t-1}$)

- Estatística ADF: -5.9177
- P-valor: 0.0000
- Conclusão: A série tornou-se estacionária com alta significância estatística. A primeira diferença foi eficaz em remover a tendência linear dos dados.
- Análise Visual (ACF/PACF): Os gráficos de autocorrelação mostram um decaimento rápido, característico de séries estacionárias, embora ainda possam existir correlações sazonais remanescentes.

Diferença Sazonal ($\Delta y_t = y_t - y_{t-52}$)

- Estatística ADF: -2.1894
- P-valor: 0.2101
- Conclusão: A diferenciação apenas sazonal (lag 52, correspondente a um ano em dados semanais) não foi suficiente para tornar a série estacionária. Isso sugere que, embora haja sazonalidade, a tendência (nível da série) ainda domina a estrutura dos dados.

Diferença Combinada (1ª Ordem + Sazonal)

- Estatística ADF: -12.0638
- P-valor: 0.0000
- Conclusão: A aplicação de ambas as diferenças tornou a série estacionária. No entanto, é necessário cautela: a sobre-diferenciação pode introduzir ruído desnecessário e correlações negativas artificiais.

Foi analisada a matriz de correlação entre a variável alvo (`volume`) e as covariáveis (`inv`, `users`) após a transformação logarítmica e a primeira diferença.

X	Y	Correlação (Log)	Correlação (1ª Diferença)
Volume	Users	0.67	0.67
Volume	Inv	0.31	0.30
Inv	Users	0.80	0.54

Tabela 3: Comparação do Log x Log com diferenciação

Observações:

1. **Robustez da Relação Volume-Users:** A correlação entre `volume` e `users` manteve-se forte (0.67) mesmo após a diferenciação. Isso indica que o crescimento de usuários está genuinamente associado ao volume vendido.

mente ligado ao crescimento do volume, e não é apenas uma tendência coincidente.

- Multicolinearidade:** A correlação entre as covariáveis (*inv* e *users*) caiu de 0.80 (nível) para 0.54 (diferença). Isso é positivo para a modelagem, pois reduz (embora não elimine) o risco de multicolinearidade que poderia inflar a variância dos coeficientes no modelo SARIMAX.

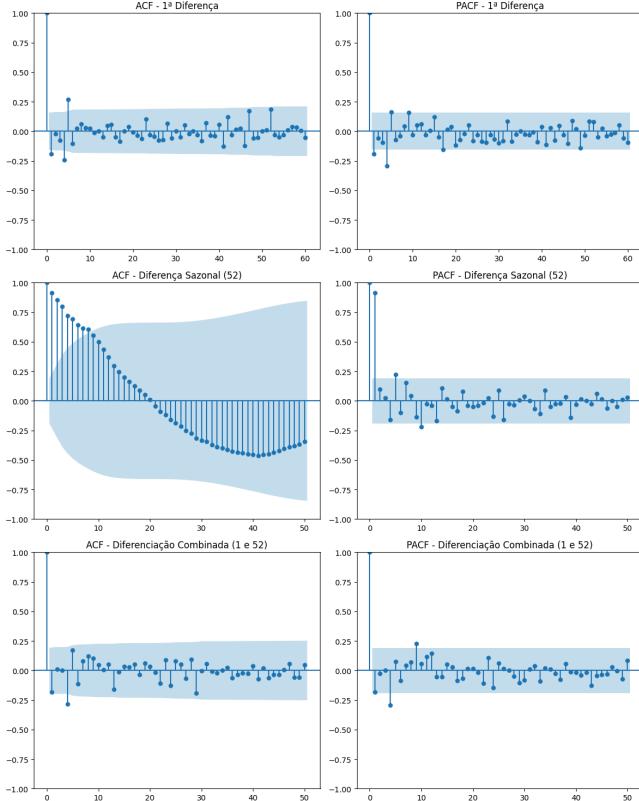


Figura 10: ACF / PACF dos 3 tipos de diferenciação testados.

3.3 Conclusão e aprendizados

A análise indica que o modelo preditivo deve operar sobre a série transformada.

- Para modelos ARIMA/SARIMA, recomenda-se utilizar $d = 1$ (uma diferenciação de primeira ordem) para tratar a tendência.
- O uso de $D = 1$ (diferenciação sazonal) deve ser validado via critério de informação (AIC), pois o teste ADF da primeira diferença já indicou estacionariedade suficiente, e a diferenciação dupla pode ser excessiva dado o tamanho da amostra (3 ciclos anuais).
- A variável *users* demonstrou ser um preditor robusto e deve ser priorizada como variável exógena no modelo.

4 Regressão Linear Simples

Antes de partir para a análise e entendimento de modelos pensados para séries temporais, é um exercício interessante estudar o modelo estatístico de regressão linear padrão, sem a adição de termos autoregressivos ou quaisquer outros que levem a modelagem temporal em consideração. Isso é útil para comparação posterior e para entender o quanto a estrutura temporal ajuda ou não na previsão da série que estamos trabalhando.

		MAE (h=1)	RMSE (h=1)	MASE (h=1)
Regressão	com- pleta	5.0444	6.5153	1.4160
Regressão	com <i>users</i>	5.5627	6.7705	1.5697
Regressão	com covariáveis log	4.6223	6.0683	1.2964
Regressão	com log <i>users</i>	4.9678	6.0963	1.4004

Tabela 4: Métricas de desempenho dos modelos sem modelagem temporal

Observevamos um desempenho consideravelmente pior dos modelos tradicionais do que os modelos baseline feitos para séries temporais. Essa piora indica que a modelagem temporal é essencial para predizer a série de interesse.

5 Modelos de Suavização Exponencial

Os modelos de suavização exponencial são uma classe de modelos temporais que não requerem estacionariedade. Nesta análise, foram testadas duas variantes ETS na série *log_volume* para estabilizar a variância (heterocedasticidade), conforme a metodologia:

1. Holt (Trend Only): Modela o nível e a tendência (aditiva e amortecida)
2. Holt-Winters (Seasonal): Adiciona o componente sazonal com período de 52 semanas.

5.1 Avaliação do desempenho

Modelo	AIC	BIC	MASE (h=1)	MASE (h=2)	MASE (h=3)	MASE (h=4)	RMSE (h=4)
Holt	-243.0	-230.9	0.35	0.49	0.56	0.54	2.49
Holt Winters	-79.5	-89.6	0.37	0.57	0.74	0.86	3.98

Tabela 5: Métricas de desempenho (walk-forward) dos modelos ETS aplicados à série *log_volume*

1. O modelo Holt (Trend Only) obteve as melhores métricas internas (AICc e BIC) e o melhor desempenho na previsão walk-forward. Este modelo alcançou um MASE (h=1) de 0.35, indicando que o erro médio de previsão é significativamente menor do que o erro do modelo Seasonal Naive.
2. O modelo Holt-Winters, que tenta capturar a sazonalidade de 52 semanas, teve um desempenho pior nas métricas internas e externas (RMSE h=4 de 3.98).

6 Modelos SARIMAX

O modelo SARIMAX foi aplicado à série *log_volume* com a diferenciação de primeira ordem ($d=1$) já definida para tratar a tendência. A ordem do modelo sazonal ($D=0$) foi mantida em zero, reforçando a conclusão da seção 4.3 de que a sazonalidade não é forte o suficiente para justificar a diferenciação sazonal.

6.1 Tratamento da Multicolinearidade

Uma etapa importante foi o tratamento da multicolinearidade entre as covariáveis *log_users* e *log_inv*:

1. Correlação Original: A correlação entre *inv* e *users* na escala original era alta (aproximadamente 0.83).

2. Correlação na Série Diferenciada: Após a transformação logarítmica e a diferenciação de primeira ordem ($d=1$), a correlação foi reduzida para 0.54.

Apesar da redução, a inclusão de ambas covariáveis no SARIMAX resultou em um desempenho subótimo devido à redundância informacional e ao aumento da variância dos coeficientes (multicolinearidade).

Modelo	MASE (h=1)	MASE (h=4)	RMSE (h=4)
ETS (Holt Trend)	0.3528	0.5413	2.4907
Sarimax (Exog)	0.2720	0.4375	2.0069

Tabela 6: Comparativo de desempenho entre o melhor ETS e o modelo SARIMAX final.

7 Comparação entre Modelos Profundos: TiDE x PatchTST

Os modelos **TiDE** e **PatchTST** emergiram como alternativas modernas e altamente competitivas para previsão em séries temporais. Embora ambos pertençam ao ecossistema de modelos neurais especializados em forecasting, suas filosofias estruturais são bastante diferentes:

- **TiDE**: modelo baseado em MLPs com **skip connections**, capaz de usar **covariáveis futuras (futr_exog)**, o que o torna extremamente flexível quando temos variáveis de calendário ou variáveis externas previstas.
- **PatchTST**: modelo Transformer que interpreta a série como **patches** e aplica **multi-head self-attention** apenas sobre entradas passadas. É excelente para tarefas **univariadas** e captura padrões sazonais/lokais de forma eficiente.

Na prática, observamos um “duelismo natural”:

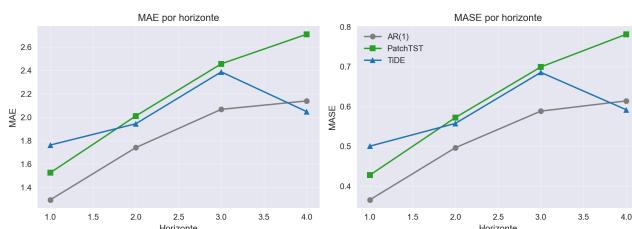
- O **PatchTST** domina quando a série é univariada ou com poucas (ou nenhuma) covariáveis futuras.
- O **TiDE** tende a ganhar força quando há diversas covariáveis preditivas disponíveis.

No nosso caso:

- PatchTST foi treinado **univariado**, pois a versão atual do NeuralForecast não suporta futr_exog nesse modelo.
- TiDE utilizou **month**, **weekofyear**, **inv**, **users** como exógenas futuras.

Abaixo os resultados walk-forward (h = 1 a 4), incluindo um baseline **AR(1)**:

7.1 Resultados (MAE e MASE por horizonte)



7.2 Análise Comparativa dos Resultados

A partir dos valores de MAE e MASE nos horizontes de previsão (h = 1 a 4), podemos destacar três conclusões centrais sobre o desempenho dos modelos AR(1), PatchTST e TiDE.

7.2.1 1. AR(1) foi superior em quase todos os horizontes curtos

O modelo AR(1) apresentou o melhor desempenho em h = 1 e manteve resultados competitivos em h = 2, superando tanto o PatchTST quanto o TiDE em praticamente todos os cenários de curto prazo.

Isso revela que a série possui uma forte persistência temporal, algo que um modelo linear simples consegue capturar com alta eficiência. Os modelos profundos, por sua vez, tendem a sofrer mais no curto prazo devido ao custo de parametrização e à necessidade de interpretar padrões mais amplos.

Conclusão: → Em horizontes muito curtos, AR(1) foi consistentemente melhor que PatchTST e TiDE.

7.2.2 2. PatchTST superou TiDE apenas nos horizontes muito curtos

O PatchTST apresentou desempenho superior ao TiDE nos horizontes:

- h = 1
- h = 2
- h = 3

Esse comportamento ocorre porque o PatchTST captura com muita eficiência padrões locais devido à operação em patches e ao self-attention restrito à janela de histórico. Assim, ele responde rapidamente às variações recentes da série.

No entanto, mesmo tendo vencido o TiDE nesses horizontes, é importante notar que:

→ PatchTST também ficou abaixo do AR(1) na maior parte dos casos.

Isso reforça que, na nossa série, a estrutura local de curto prazo é mais simples e bem explicada por um modelo autoregressivo linear.

7.2.3 3. TiDE foi melhor que o PatchTST nos horizontes longos e só superou o AR(1) no fim

O modelo TiDE superou o PatchTST em h = 4, mostrando que:

- seu uso de covariáveis futuras (futr-exog) se torna informativo nos horizontes mais longos;
- sua arquitetura MLP com skip connections favorece a captura de tendências e estruturas globais.

Entretanto:

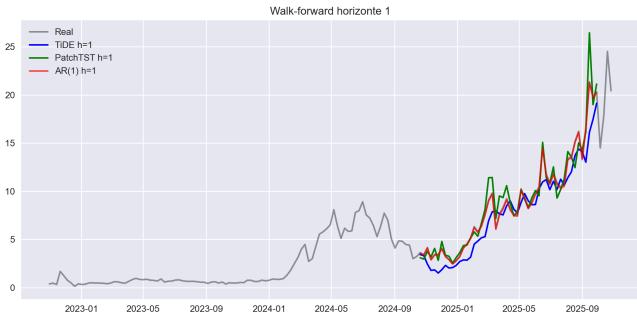
→ TiDE só superou o AR(1) nos horizontes mais longos, particularmente em h = 4. → Nos horizontes curtos, TiDE ficou em algumas janelas atrás do AR(1) e do PatchTST.

Conclusão: → TiDE domina no longo prazo; PatchTST domina no curto prazo; AR(1) domina quase sempre nos horizontes iniciais.

7.2.4 Síntese Geral

- AR(1): melhor desempenho geral em horizontes curtos (h = 1, 2).
- PatchTST: melhor que TiDE apenas nas previsões muito curtas (h < 2), mas ainda inferior ao AR(1).
- TiDE: supera o PatchTST nos horizontes longos e supera o AR(1) somente nas previsões mais distantes (h = 4).

Esses resultados reforçam a importância de escolher o modelo com base no horizonte de previsão – uma decisão que altera completamente qual modelo se torna dominante.



8 Conclusão

A inclusão das variáveis foi benéfica para a modelagem da série. Elas permitiram com que mais técnicas de modelagem e predição fossem aplicadas, e a dinâmica temporal permitiu a obtenção de bons resultados na predição.

Como foi possível observar, apenas utilizar as covariáveis para predição sem levar em conta a dinâmica temporal trás resultados piores do que utilizar técnicas de modelagem temporal adequadas (desde as baselines até as técnicas de suavização exponencial). Isso demonstrou a importância, para esse conjunto de dados, de modelar essa dinâmica, e motivou a exploração de outras técnicas.

Observamos que a diferenciação da série foi uma técnica importante e efetiva para a modelagem, e permitiu o bom desempenho de modelos como o SARIMAX, que alcançaram resultados equiparáveis ou até melhores do que os vistos em modelos mais complexos (que envolviam deep learning). Podemos atribuir isso à natureza da base de dados: com poucos registros e poucas covariáveis, modelos que não envolvem muitas camadas e treinos complexos podem evitar overfitting e alcançar resultados melhores.

Possíveis direções futuras para uma melhoria melhor dos resultados poderiam envolver a transformação das covariáveis, incorporando conhecimentos do negócio para gerar colunas que não sofram de multicolinearidade/alta correlação, além da exploração de mais modelos, desde que estes não sejam muito complexos.