

Modelos Baseline

Modelos baseline são abordagens de previsão simples que servem como um ponto de referência para avaliar o desempenho de modelos mais complexos. A sua principal função é estabelecer um limiar de performance: se um modelo sofisticado não conseguir superar um baseline simples, isso sugere que o modelo mais complexo não está capturando bem os padrões da série ou que a série é inherentemente difícil de prever. Para este trabalho, foram implementados cinco modelos baseline, cujos resultados foram comparados utilizando as métricas MAE, RMSE e MAPE.

Descrição dos Modelos

Mean Method: Este é o baseline mais simples, onde a previsão para qualquer período futuro é a média de todos os valores observados no conjunto de treino.

$$\hat{y}_{T+h|T} = \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

Naive Method: Neste método, a previsão para o próximo período é simplesmente o último valor observado.

$$\hat{y}_{T+h|T} = y_T$$

Seasonal Naive: Uma variação do Naive Method, útil para séries com forte sazonalidade. Ele prevê o valor futuro utilizando a observação do mesmo período no ciclo sazonal anterior. Utilizamos dados semanais considerando sazonalidade anual, portanto o período é $m = 52$.

$$\hat{y}_{T+h|T} = y_{T+h-m}$$

Drift Method: Este método é uma extensão do Naive e é adequado para séries com tendência. Ele extrapola uma linha traçada entre a primeira e a última observação do conjunto de treino. A previsão incorpora uma “inclinação” (drift) média.

$$\hat{y}_{T+h|T} = y_T + h \left(\frac{y_T - y_1}{T - 1} \right)$$

Média Móvel (Rolling Mean): A previsão é calculada como a média dos últimos K valores observados no conjunto de treino. Este método suaviza flutuações de curto prazo. Para este estudo, foi utilizado $K = 12$.

$$\hat{y}_{T+h|T} = \frac{1}{K} \sum_{t=T-K+1}^T y_t$$

Metodologia de Validação

Para avaliar o desempenho dos modelos de forma robusta, foi empregada a técnica de *Time Series Cross-Validation* com janela expansível. Esta abordagem consiste em:

- **Janela inicial de treino:** 120 observações
- **Horizonte de previsão:** 30 pontos por fold
- **Passo de deslocamento:** 1 observação entre folds consecutivos
- **Número de folds:** 8 folds

Em cada fold, o modelo é treinado com os dados até o ponto de corte e realiza previsões *one-step-ahead rolling forecast* para o horizonte de 30 pontos. Nesta estratégia, ao prever o ponto $h + 1$, utilizamos todos os dados observados até o ponto h (incluindo os valores reais do teste até esse ponto), garantindo uma avaliação realista da capacidade preditiva dos modelos.

A Figura 1 ilustra o comportamento de todos os modelos baseline no último fold, permitindo visualizar como cada método se comporta em relação aos valores observados.

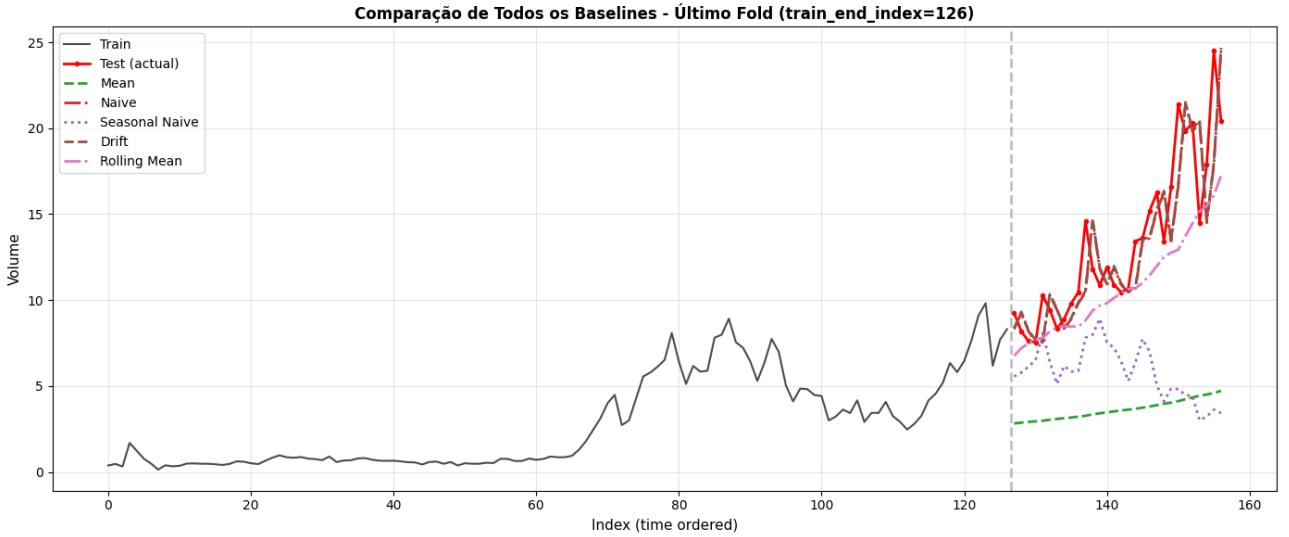


Figure 1: Comparação visual de todos os modelos baseline no último fold de validação. A linha preta representa os dados de treino, a linha vermelha os valores reais do teste, e as linhas tracejadas coloridas representam as previsões de cada modelo baseline.

Resultados

Os resultados obtidos para as métricas de erro, calculados como a média sobre os 8 folds de validação cruzada, estão consolidados na Tabela 1. O **Drift Method** apresentou o melhor desempenho geral, seguido de perto pelo Naive Method.

Modelo	MAE	RMSE	MAPE (%)
Drift Method	1.6139	2.1241	13.34
Naive Method	1.6258	2.1323	13.39
Rolling Mean (K=12)	2.3230	3.0989	17.68
Seasonal Naive	6.0535	7.3639	46.48
Mean Method	8.4744	9.1184	69.71

Table 1: Resultados médios dos modelos baseline na validação cruzada (8 folds). Modelos ordenados por RMSE crescente (melhor para pior).

O modelo com melhor desempenho (**Drift Method**) foi selecionado para uma análise de resíduos detalhada, apresentada na próxima subseção.

Para aplicações práticas, saber a margem de erro é tão importante quanto a previsão em si. Avaliamos os intervalos de confiança (IC) de 95% gerados pelo Drift Method utilizando métricas probabilísticas robustas. A Figura 2 apresenta a previsão visual juntamente com a faixa de incerteza estimada.

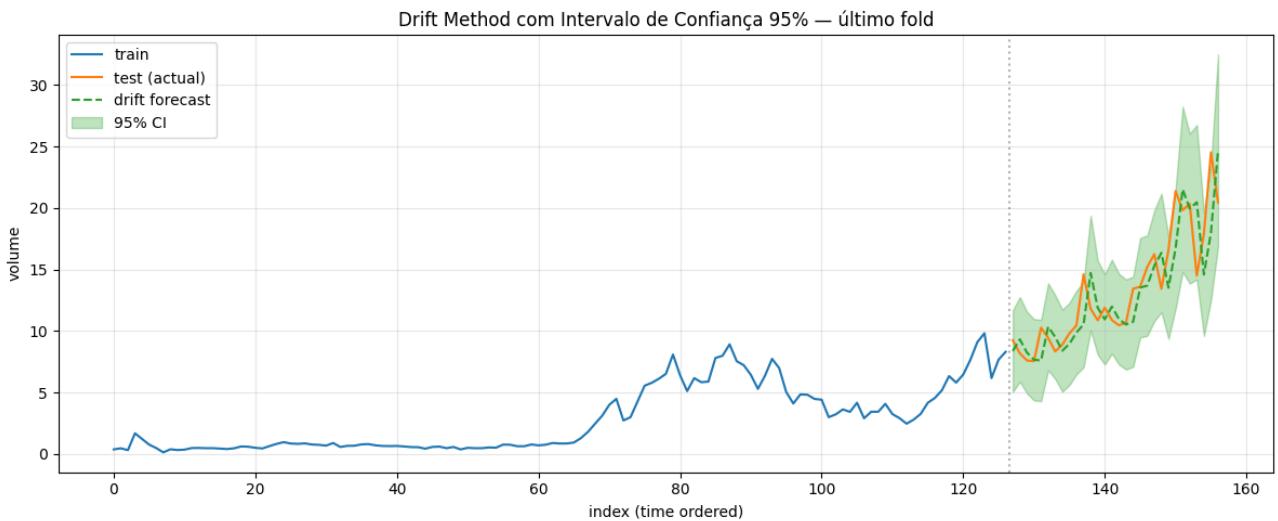


Figure 2: Previsão *one-step-ahead* do Drift Method no último fold, exibindo o Intervalo de Confiança de 95% (área sombreada). A caixa de texto destaca as métricas de calibração do intervalo.

As métricas obtidas demonstram que o estimador de incerteza do Drift Method é confiável:

- **Cobertura (Coverage):** O modelo obteve **93.33%** de cobertura, valor muito próximo do ideal teórico de 95%. Isso significa que, em 93% das semanas, o valor real de volume caiu dentro da faixa prevista, indicando um modelo bem calibrado (nem excessivamente confiante, nem conservador demais).
- **Winkler Score (10.69):** Este score penaliza intervalos largos que não capturam o valor real. O valor baixo obtido, em conjunto com a alta cobertura, sugere que o modelo consegue manter o intervalo "estreito o suficiente" para ser útil (Largura Média ≈ 8.53) sem sacrificar a precisão.
- **Violacão (6.67%):** A taxa de violação está marginalmente acima dos 5% esperados, o que é aceitável para um modelo baseline.

Esta validação da incerteza reforça a escolha do Drift Method como um benchmark sólido: ele não apenas prevê a tendência corretamente, mas também fornece uma estimativa de risco realista para o negócio.

Análise de Resíduos do Melhor Modelo (Drift)

A análise de resíduos do melhor modelo baseline foi realizada em três etapas:

1. **Série Temporal dos Resíduos:** A série temporal dos resíduos parece apresentar algum padrão temporal, podemos ver alguns intervalos que se repetem em direção e diferem pouco em amplitude. A maioria dos resíduos concentrados em torno de zero. No entanto, a variância pareceu aumentar ligeiramente em alguns períodos, sugerindo possível heteroscedasticidade.

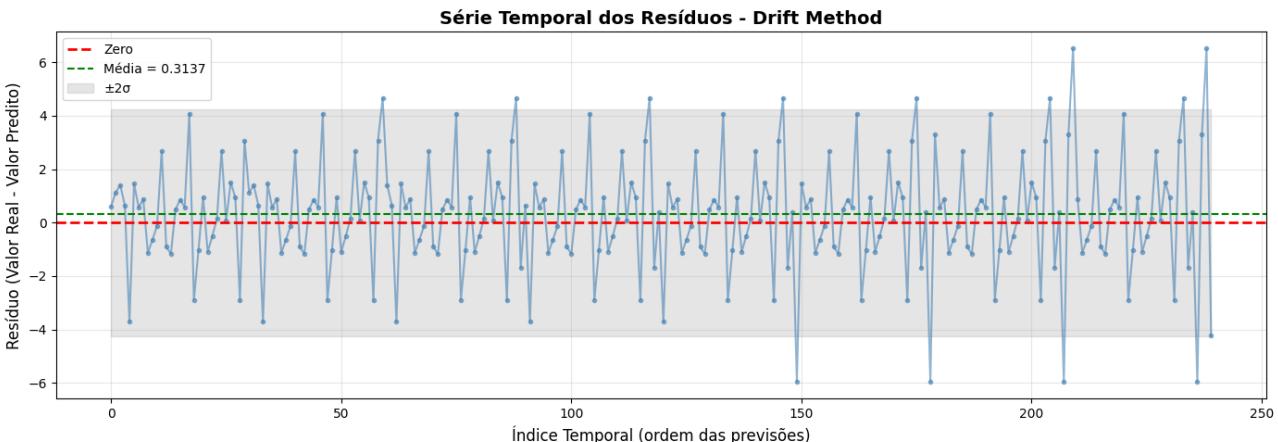


Figure 3: Série Temporal dos Resíduos

2. Normalidade dos Resíduos: A análise de normalidade através do histograma e gráfico Q-Q revelou desvios em relação à distribuição normal, especialmente nas caudas. Testes estatísticos confirmaram a rejeição da hipótese de normalidade, indicando que, embora o modelo capture a tendência central (média), ele pode não capturar perfeitamente os eventos extremos.

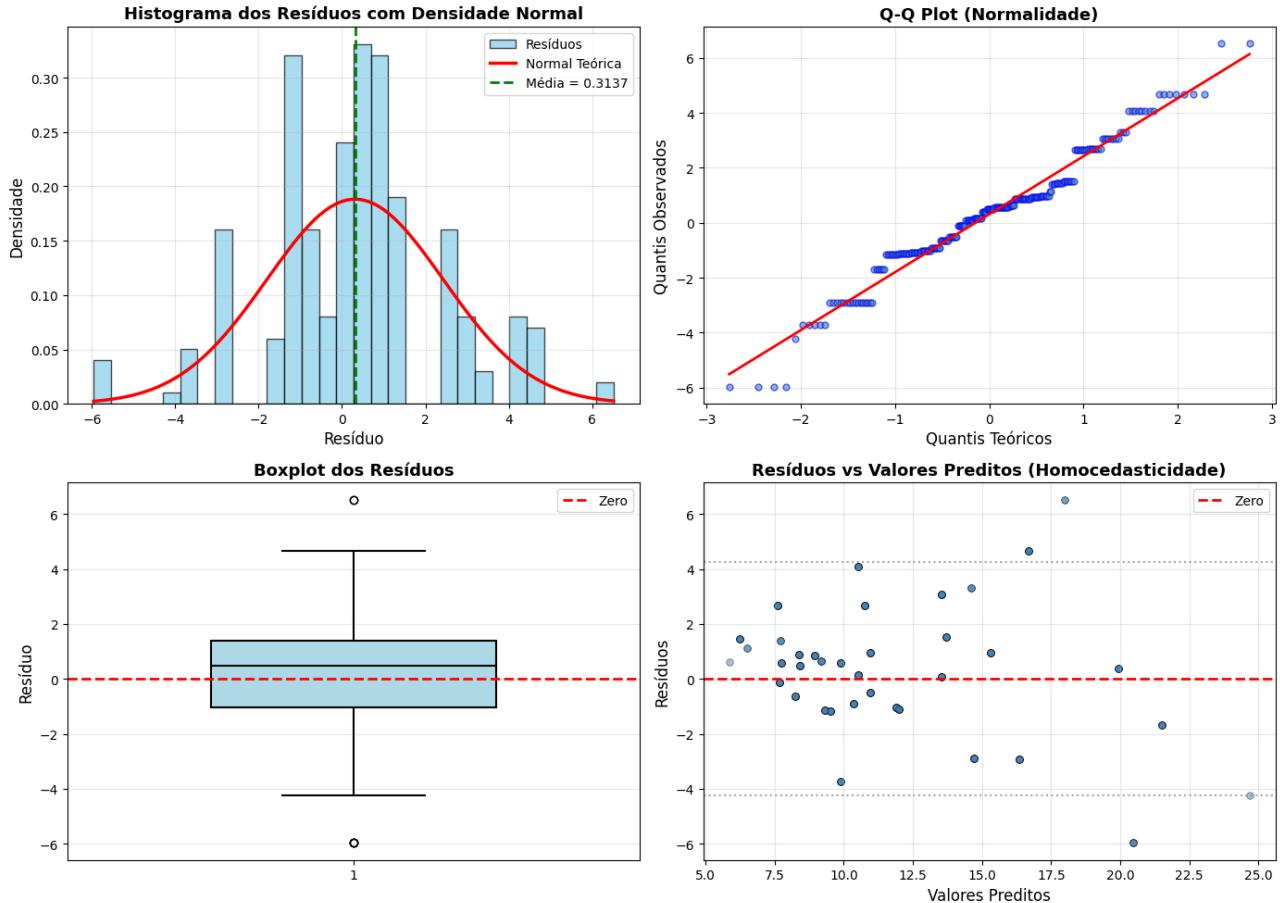


Figure 4: Normalidade dos Resíduos

3. Autocorrelação dos Resíduos: A análise de autocorrelação (ACF) mostrou correlação serial significativa em alguns lags iniciais. Isso era esperado para um modelo baseline simples e indica que há estrutura temporal de curto prazo remanescente que não foi capturada apenas pela tendência linear do Drift.

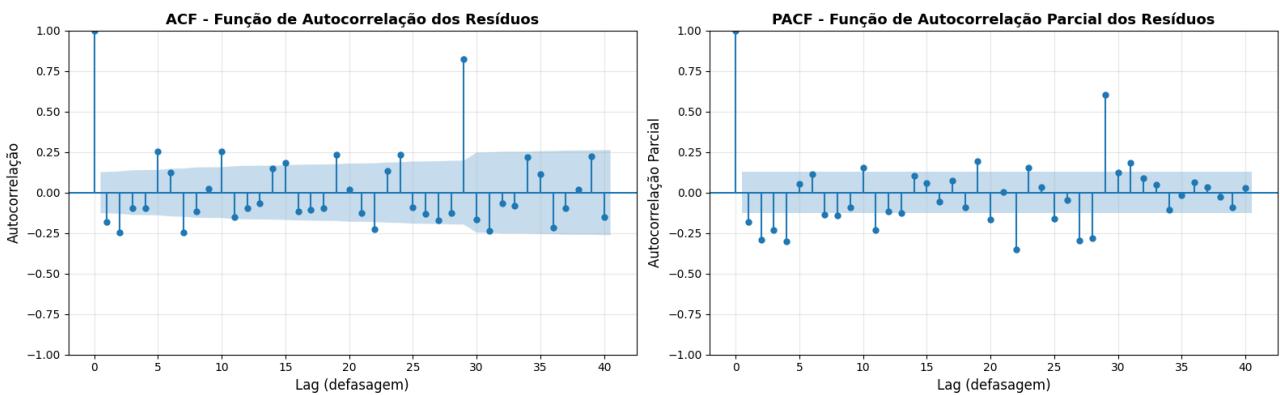


Figure 5: Autocorrelação dos Resíduos

Transformação Box-Cox e Heteroscedasticidade

A análise inicial dos dados sugeriu heteroscedasticidade. Para mitigar este problema, aplicou-se a transformação de Box-Cox.

Após aplicar a transformação, os modelos foram reavaliados. Os resultados mostraram diferenças marginais nas métricas de erro. Por exemplo, o MAE do Drift Method variou apenas de **1.614** para **1.608** com a transformação.

Esta pequena variação pode ser explicada por dois fatores principais:

- **Dominância da Tendência:** A série possui uma tendência linear forte que é bem capturada pelo Drift Method independentemente da escala de variância.
- **Horizonte Curto:** Com previsão *one-step-ahead*, o modelo é re-calibrado a cada passo, reduzindo o impacto negativo da heteroscedasticidade que seria mais visível em horizontes longos.

Conclusão

Os modelos baseline estabelecem um ponto de referência importante. O **Drift Method** ($MAE \approx 1.61$) provou ser um benchmark robusto, superando significativamente abordagens sazonais ou de média simples. A análise de resíduos evidenciou que, embora a tendência tenha sido removida, ainda existe autocorrelação a ser explorada por modelos estatísticos mais sofisticados.