EFRAIN CASTILLO

Merchant Revenue Forecasting

Benchmarking forecasting methods across a heterogeneous merchant portfolio

Self-directed portfolio project using synthetic data

BUSINESS QUESTION

How can Finance improve monthly merchant revenue forecasting across a diverse merchant portfolio while balancing accuracy, stability, and interpretability?

PORTFOLIO	HISTORY	OBSERVATIONS	EVALUATION
50 synthetic merchants	**60 months (2020-2024)**	**3,000 merchant-months**	**MAPE, RMSE, MAE**

HEADLINE RESULT

Holt achieved the lowest mean MAPE at approximately **8.6%**, versus approximately **12.5%** for the Seasonal Naive baseline.

~31%

relative reduction in mean MAPE versus Seasonal Naive

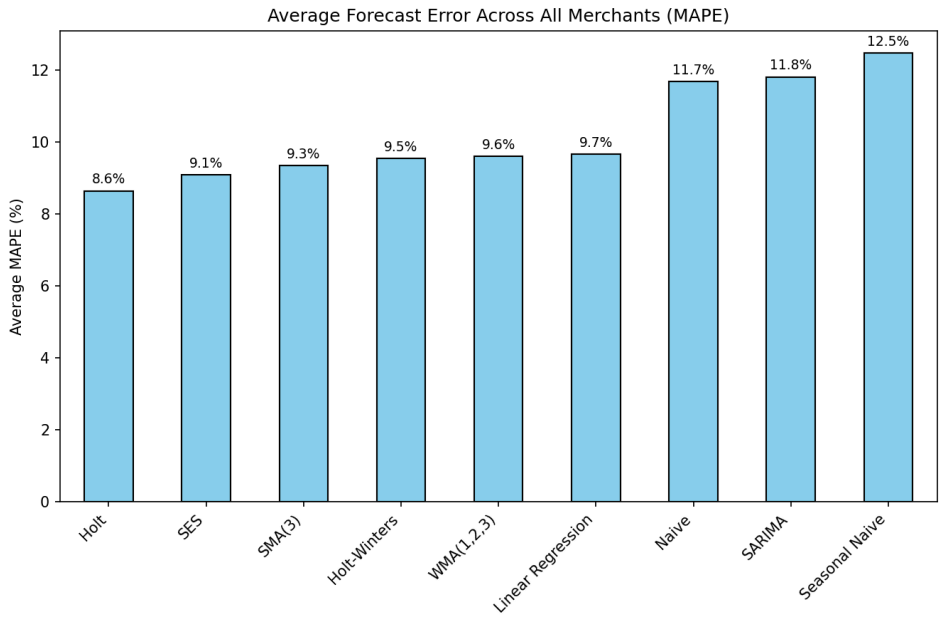

Figure 1. Average forecast error (MAPE) across all 50 merchants. Nine methods were benchmarked on a common 12-month holdout.

Takeaway: More complex models did not automatically outperform simpler, interpretable approaches. Holt provided the strongest average accuracy across this portfolio. This does not mean Holt will always be best for every merchant or future period.

Model Performance & Forecast Variability

Portfolio averages and merchant-level error distributions

Methods Benchmarked

Naive | Seasonal Naive | SMA(3) | WMA | SES | Holt | Holt-Winters | SARIMA | Linear Regression

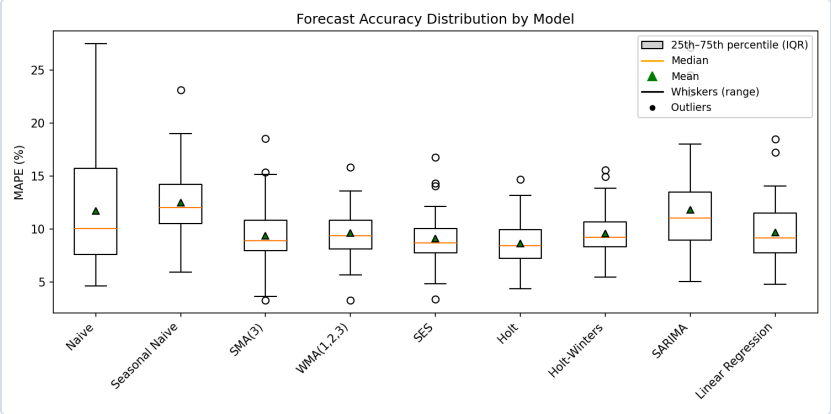

Figure 2. Distribution of merchant-level MAPE by model. Means (green triangles) and medians (orange lines) show both central tendency and spread.

Verified Mean MAPE

MODEL	MEAN MAPE
Holt	**8.64%**
SES	9.08%
SMA(3)	9.34%
Holt-Winters	9.54%
WMA	9.60%
Linear Regression	9.66%
Naive	11.68%
SARIMA	11.81%
Seasonal Naive	12.47%

Why Distribution Matters

- Portfolio averages alone can hide merchant-level variation.
- Model performance differed across merchants.
- Finance should evaluate both average forecast error and the distribution of error.
- Forecasting governance should identify merchants with greater uncertainty rather than relying solely on one portfolio-wide score.

Governance note: Greater model complexity did not guarantee better average performance. SARIMA (~11.81% mean MAPE) was not among the strongest average performers in this benchmark.

Merchant Example & FP&A Implications

Illustrative forecast path and planning takeaways

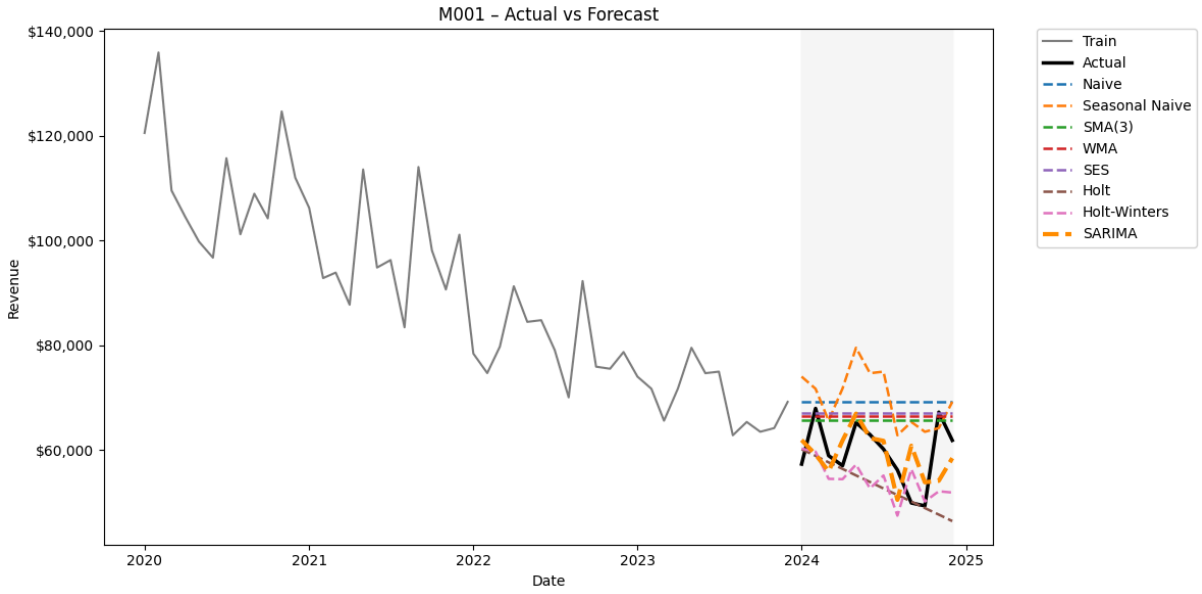

Figure 3. Illustrative single-merchant example (M001): held-out actuals versus competing forecasts. This view complements portfolio averages with a merchant-level path comparison.

FP&A Implications

- More reliable merchant-level forecasts can improve revenue planning.
- Forecast-error monitoring can help Finance distinguish predictable merchants from higher-uncertainty merchants.
- Model benchmarking provides a disciplined baseline before introducing additional complexity.
- Forecast performance should be refreshed as new actuals arrive.
- Different merchant behaviors may justify different models rather than forcing one methodology across every merchant.

Limitations

- Synthetic data — designed to mimic FP&A challenges, not represent a real portfolio.
- Results may not generalize to real merchant businesses.
- No causal interpretation of drivers or interventions.
- No production deployment claim.
- Model rankings can change as new observations arrive.
- External drivers may improve forecasts but were not the focus of this benchmark.

Tools: Python · pandas · statsmodels · scikit-learn · matplotlib
Self-directed portfolio project using synthetic data · efrainfinance.com

View technical methodology and code
https://github.com/ecastillo081/Merchant_Revenue_Forecasting