Financial Time-Series Forecasting Archive

Jeremy Bejarano[*], Fernando Rocha Urbano[†]

March 24, 2025

---

[*]jbejarano@uchicago.edu
[†]fernandourbano@uchicago.edu

model	seasonality	mean_mase	median_mase	entity_count
Simple Exponential Smoothing	5	9.200310	8.953235	30
ARIMA(1,1,1)	5	9.218529	8.953667	30

1 Abstract

2 Methodology

TEMP DEMO:

For each of the series forecasted, we calculate multiple error metrics: MSE, RMSE, MAE, MASE, and SMAPE. A detailed explanation of the metrics is given in Appendix 3.1.

The error is calculated based on all the available forecasts for each series.

Different models perform distinctly for different forecasting horizons; according to [REFERENCE], SARIMA models are better for short-term forecasts, while LSTM models are better for long-term forecasts. Viewing those differences, we provide predictions for each time-series and model using different options of:

- Number of forecasting (n)

- Forecasting horizons (h)

- Intersecting forecasting

- Rolling and expanding training sample

Each of those parameters is explained below.

2.1 Number of Forecasts

The number of forecasts refers to the number of times a model trains and predicts. Each time it predicts, it goes forward to use the previous prediction data as part of the training data, then aiming to predict the next date not yet predict.

When the number of forecasts is defined, we organize divisions of training and prediction dates in a way that the last division predicts the last chunck of dates available in the series.

After all are predicted, the forecasts are compared with the actual values to generate the error metrics.

2.2 Forecasting Horizon (Step Size)

The forecasting horizon is the number of periods ahead that the model is predicting. For each of the forecasts made, we define the prediction period based on the forecasting horizon.

For example, consider the forecasting horizon equals to 1, the number of forecasts equals to 12, monthly data and December 2025 being the last date available. In this case, we create 12 divisions.

For the first one, the training data goes until December 2024, and it predicts the value for January 2025. The second division has the training data until January 2025 and predicts the value for February 2025, and so on until the last division predicts December 2025 using data until November 2025 to train the model. The predictions are gathered and compared with the actual values to generate the error metrics.

The following graphic illustrates a forecast with number of forecasts equal to 6 and forecast horizon equals to 1.

[Forecasting Example: 6 forecasts with step size 1]
Image placeholder - file will be generated later

Using the same example, if the forecasting horizon is 3, we must start predicting from earlier dates. The first division uses training data until December 2022 to predict January, February and March 2023. The second division trains its model using data until March 2023 to predict April, May and June 2023, and so on until the last division predicts October, November and December 2025 using data until September 2025 to train the model.

The following graphic illustrates a forecast with number of forecasts equal to 6 and forecast horizon equals to 3.

[Forecasting Example: 6 forecasts with step size 3]
Image placeholder - file will be generated later

For larger forecasting horizons, the first division will have considerably less training data. The following illustrates 4 forecasts with 12 months forecasting horizon.

[Forecasting Example: 4 forecasts with step size 12]
Image placeholder - file will be generated later

2.3   Intersecting Forecasts

By default, when the forecasting horizon is bigger than 1, the predictions are made in a non-intersecting way, meaning that each division has prediction dates without any overlap with other division.

However, we also provide the option to make the predictions in an intersecting way. In this case, the training and prediction dates always advance one step at a time.

For instance, if the forecasting horizon is 3, the number of forecasts is equal to 12, and the last date available is December 2025, the first division uses data until October 2024 to predict November, December 2024 and January 2025. The second division uses data until November 2024 to predict December 2024, January and February 2025 (different from the example above). The last division predicts October, November and December 2025 using data until September 2025 to train the model.

Jeremy Bejarano, Fernando Rocha Urbano

After the forecasts are completed, we only use the last prediction of each division to calculate the error metrics. This way prioritizes exclusively the long-term predictions. The non-intersecting way of forecasting analyzes the short and long-term predictions equally by comparing predictions and actual values for dates that are one step ahead of the training data and dates that are the furthest from the training data. The intersecting way provides a more truthful analysis of the performance only in the longer horizon.

The graphic below provides 6 forecasts with step size of 3 similarly to graphic XX. The difference with the previous one occurs due to the use of intersecting dates of the divisions.
CHANGE GRAPHIC

[Forecasting Example: Intersecting forecasts]
Image placeholder - file will be generated later

2.4   Rolling vs Expanding Training Sample

By default, the expanding training sample is used, meaning that the training data always starts from the same date for all the divisions and expands until the last date before the prediction dates. In the rolling sample, after running the first division, we move the training start date by the forecasting horizon to run the next division. In this way, the data training always has the same size for all the divisions.

3 Appendix

3.1 Error Metrics

Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3.1}$$

Penalizes large errors more than small ones due to squaring, making it useful for detecting significant deviations. Sensitive to outliers, which can dominate the error measure.

Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3.2}$$

Maintains the same unit as the target variable, making it more interpretable than MSE. Still sensitive to outliers due to squaring before averaging.

Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.3}$$

Less sensitive to outliers than MSE and RMSE. Does not differentiate between small and large errors as effectively as squared errors.

Mean Absolute Scaled Error (MASE)

$$\text{MASE} = \frac{\frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|}{\frac{1}{n} \sum_{i=2}^{n} |y_i - y_{i-1}|} \tag{3.4}$$

Scale-independent, making it useful for comparing forecast errors across different datasets. Requires a meaningful benchmark (like a naive forecast) to be interpreted correctly.

Symmetric Mean Absolute Percentage Error (SMAPE)

$$\text{SMAPE} = \frac{2}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} \tag{3.5}$$

Bounded between 0 and 200%, preventing extreme values. Can be unstable when the denominator is close to zero, leading to high error values.