

Forecasting overview

Rob J Hyndman*

1 What can be forecast?

Forecasting is required in many situations: deciding whether to build another power generation plant in the next five years requires forecasts of future demand; scheduling staff in a call centre next week requires forecasts of call volume; stocking an inventory requires forecasts of stock requirements. Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for telecommunication routing). Whatever the circumstances or time horizons involved, forecasting is an important aid in effective and efficient planning.

Some things are easier to forecast than others. The time of the sunrise tomorrow morning can be forecast very precisely. On the other hand, currency exchange rates are very difficult to forecast with any accuracy. The predictability of an event or a quantity depends on several factors (Hyndman and Athanasopoulos, 2021):

1. how well we understand the factors that contribute to it;
2. how much data is available;
3. how similar the future is to the past;
4. whether the forecasts can affect the thing we are trying to forecast.

Forecasting situations vary widely in their time horizons, factors determining actual outcomes, types of data patterns, and many other aspects. Forecasting methods can be very simple such as using the most recent observation as a forecast (which is called the “naïve method”), or highly complex such as neural networks and econometric systems of simultaneous equations. The choice of method depends on what data is available and the predictability of the quantity to be forecast.

2 Forecasting methods

Forecasting methods fall into two major categories: statistical (including machine learning) and judgemental methods.

Statistical forecasting can be applied when two conditions are satisfied:

1. numerical information about the past is available;
2. it is reasonable to assume that some aspects of the past patterns will continue into the future.

There is a wide range of statistical forecasting methods, often developed within specific disciplines for specific purposes. Each method has its own properties, accuracies, and costs that must be considered when choosing a specific method.

Judgemental forecasting methods are used when one or both of the above conditions does not hold.

They are also used to adjust statistical forecasts, taking account of information that was not able to be incorporated into the formal statistical model. These are not purely guesswork—there are well-developed structured approaches to obtaining good judgemental forecasts.

*Rob J Hyndman FAA FASSA is Professor of Statistics in the Department of Econometrics & Business Statistics at Monash University, Australia. He has published extensively in leading statistical and forecasting journals. He is co-author of the highly regarded online text on business forecasting, *Forecasting: principles and practice* (OTexts, 3rd edition, 2021, [OTexts.com/fpp3](https://otexts.com/fpp3)). Professor Hyndman was Editor-in-Chief of the *International Journal of Forecasting* from 2005 to 2018. He is an elected Fellow of the Australian Academy of Science, the Academy of the Social Sciences in Australia, and the International Institute of Forecasters. In 2007, he was awarded the prestigious Moran Medal from the Australian Academy of Science, for his contributions to statistical research. In 2021, he won the Pitman Medal from the Statistical Society of Australia.

As judgemental methods are non-statistical, they will not be considered further in this article.

Explanatory versus time series forecasting

Statistical forecasts can be largely divided into two classes: time series and explanatory models. Explanatory models assume that the variable to be forecasted exhibits an explanatory relationship with one or more other “predictor” variables. For example, we may model the electricity demand (ED) of a hot region during the summer period as

$$ED = f(\text{current temperature, strength of economy, population, time of day, day of week, error}). \quad (1)$$

The relationship is not exact—there will always be changes in electricity demand that can not be accounted for by the variables in the model. The “error” term on the right allows for random variation and the effects of relevant variables not included in the model. Models in this class include regression models, additive models, and some kinds of neural networks.

The purpose of the explanatory model is to describe the form of the relationship and use it to forecast future values of the forecast variable. Under this model, any change in inputs will affect the output of the system in a predictable way, assuming that the explanatory relationship does not change.

In contrast, time series forecasting uses only information on the variable to be forecast, and makes no attempt to discover the factors affecting its behaviour. For example,

$$ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, ED_{t-3}, \dots, \text{error}), \quad (2)$$

where t is the present hour, $t + 1$ is the next hour, $t - 1$ is the previous hour, $t - 2$ is two hours ago, and so on. Here, prediction of the future is based on past values of a variable and/or past errors, but not on predictor variables that may affect the system. Time series models used for forecasting include ARIMA models, exponential smoothing and structural models.

There are several reasons for using a time series forecast rather than an explanatory model for forecasting. First, it may be difficult to measure the relevant predictor variables in order to obtain the necessary data. Second, even if the data is available, it is necessary to forecast the various predictor variables in order to be able to forecast the variable of interest, and this may be too difficult. Third, even if the predictor variables can be measured and forecasted, their forecasts will introduce additional uncertainty into the process, and the resulting forecasts of the variable of interest may be worse than simply using a time series model.

A third type of forecasting model uses both time series and predictor variables. For example,

$$ED_{t+1} = f(ED_t, \text{current temperature, time of day, day of week, error}). \quad (3)$$

These types of models have been given various names in different disciplines. They are known as dynamic regression models, ARMAX models, transfer function models, and linear system models (assuming f is linear).

3 The basic steps in a forecasting task

There are usually five basic steps in any forecasting task.

Step 1: Problem definition Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of how the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organisation requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning.

Step 2: Gathering information There are always at least two kinds of information required: (a) statistical data, and (b) the accumulated expertise of the people who collect the data and use the forecasts. Often, a difficulty will be obtaining enough historical data to be able to fit a good statistical model. However, occasionally, very old data will not be so useful due to changes in the system being forecast.

Step 3: Preliminary (exploratory) analysis Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis?

Step 4: Choosing and fitting models Which model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way the forecasts are to be used. It is common to compare two or three potential models.

Step 5: Using and evaluating a forecasting model Once a model has been selected and its parameters estimated, the model is to be used to make forecasts. The performance of the model can only be properly evaluated after the data for the forecast period have become available. A number of methods have been developed to help in assessing the accuracy of forecasts as discussed in the next section.

4 Forecast distributions

All forecasting is about estimating some aspects of the conditional distribution of a random variable. For example, if we are interested in monthly sales denoted by y_t for month t , then forecasting concerns the distribution of y_{t+h} conditional on the values of y_1, \dots, y_t along with any other information available. Let \mathcal{I}_t denote all other information available at time t . Then we call the distribution of $(y_{t+h} \mid y_1, \dots, y_t, \mathcal{I}_t)$ the *forecast distribution*.

Typically, a forecast is presented as a single number (known as a “point forecast”). This can be considered an estimate of the mean or median of the forecast distribution. It is often useful to provide information about forecast uncertainty as well in the form of a prediction interval. For example, if the forecast distribution is normal with mean \hat{y}_{t+h} and variance σ_{t+h}^2 , then a 95% prediction interval for y_{t+h} is $\hat{y}_{t+h} \pm 1.96\sigma_{t+h}$. Prediction intervals in forecasting are sometimes called “interval forecasts”

For some problems, it is also useful to estimate the forecast distribution rather than assume normality or some other parametric form. This is called “density forecasting” or “probabilistic forecasting”. When only certain quantiles of the forecast distribution are computed, it is often called “quantile forecasting”.

5 Evaluating forecast accuracy

It is important to evaluate forecast accuracy using genuine forecasts. That is, it is invalid to look at how well a model fits the historical data; the accuracy of forecasts can only be determined by considering how well a model performs on new data that were not used when developing the model. When choosing models, it is common to split the data into a training set, $\{y_1, \dots, y_t\}$, and a test set $\{y_{t+1}, \dots, y_T\}$, where T is the time of the last observation. Only the training set can be used for fitting or training the model. Then the test data can be used to measure how well the model is likely to forecast on new data.

A more sophisticated variation of a training/test split is time series cross-validation, where the training/test set split above is repeated for $t = T_1, T_1 + 1, \dots, T - 1$; here T_1 is the minimal sample size needed to train a model. This is sometimes called “forecasting with a rolling origin”.

The issue of measuring the accuracy of forecasts from different methods has been the subject of much attention. We summarize some of the approaches here. A more thorough discussion is given by Hyndman and Koehler (2006). In the following discussion, \hat{y}_t denotes a forecast of y_t . We only consider the evaluation of point forecasts. There are also methods available for evaluating interval forecasts and density forecasts (Hyndman and Athanasopoulos, 2021).

Scale-dependent errors

The forecast error is simply $e_t = y_t - \hat{y}_t$ which is on the same scale as the data. Accuracy measures that are based on e_t are therefore scale-dependent and cannot be used to make comparisons between series that are on different scales.

The two most commonly used scale-dependent measures are based on the absolute error or squared errors:

$$\text{Mean absolute error (MAE)} = \text{mean}(|e_t|),$$

$$\text{Mean squared error (MSE)} = \text{mean}(e_t^2).$$

When comparing forecast methods on a single series, the MAE is popular as it is easy to understand and compute. It is optimised when the point forecast is the median of the forecast distribution. The MSE is optimised when the point forecast is the mean of the forecast distribution.

Percentage errors

The percentage error is given by $p_t = 100e_t/y_t$. Percentage errors have the advantage of being scale-independent, and so are frequently used to compare forecast performance between different data sets. The most commonly used measure is:

$$\text{Mean absolute percentage error (MAPE)} = \text{mean}(|p_t|)$$

Measures based on percentage errors have the disadvantage of being infinite or undefined if $y_t = 0$ for any t in the period of interest, and having an extremely skewed distribution when any y_t is close to zero. Another problem with percentage errors which is often overlooked is that they assume the data are on a ratio scale, so that percentages are unchanged if the units of measurement are changed. For example, a percentage error makes no sense when measuring the accuracy of temperature forecasts on the Fahrenheit or Celsius scales. They also have the disadvantage that they put a heavier penalty on negative errors than on positive errors.

Scaled errors

The MASE was proposed by Hyndman and Koehler (2006) as an alternative to the MAPE when comparing forecast accuracy across series on different scales. They proposed scaling the errors based on the *training set* MAE from a naïve forecast method. A scaled error can be defined as

$$q_t = \frac{e_t}{\frac{1}{T-m} \sum_{i=m+1}^T |y_i - y_{i-m}|},$$

where m is the length of the seasonal period of the data (e.g., $m = 12$ for monthly data, and $m = 1$ for non-seasonal data). Because the numerator and denominator both involve values on the scale of the original data, q_t is independent of the scale of the data. A scaled error is less than one if it arises from a better forecast than the average one-step naïve forecast computed on the training set. Conversely, it is greater than one if the forecast is worse than the average one-step naïve forecast computed on the training set.

The *mean absolute scaled error* is simply $\text{MASE} = \text{mean}(|q_t|)$. Similarly, the *mean squared scaled error* is given by $\text{MSSE} = \text{mean}(q_t^2)$ where

$$q_t^2 = \frac{e_t^2}{\frac{1}{T-m} \sum_{i=m+1}^T (y_i - y_{i-m})^2},$$

The latter has the advantage of being both scale-free, and optimal for the mean of the forecast distribution. The recent M5 forecasting competition used the MSSE (Makridakis et al., 2022) to measure point forecast accuracy.

References

- Hyndman, R. J. and Athanasopoulos, G. (2021), *Forecasting: principles and practice*, 3rd edn, OTexts, Melbourne, Australia.
URL: [OTexts.org/fpp3](https://otexts.org/fpp3)
- Hyndman, R. J. and Koehler, A. B. (2006), 'Another look at measures of forecast accuracy', *International J Forecasting* **22**(4), 679–688.
- Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2022), 'M5 accuracy competition: Results, findings, and conclusions', *International Journal of Forecasting* **38**(4), 1346–1364.