# Business Forecasting Methods

Rob J Hyndman[*]

## 1 Forecasting, planning and goals

Forecasting is a common statistical task in business where it helps inform decisions about scheduling of production, transportation and personnel, and provides a guide to long-term strategic planning. However, business forecasting is often done poorly and is frequently confused with planning and goals. They are three different things.

**Forecasting** is about predicting the future as accurately as possible, given all the information available including historical data and knowledge of any future events that might impact the forecasts.

**Goals** are what you would like to happen. Goals should be linked to forecasts and plans, but this does not always occur. Too often, goals are set without any plan for how to achieve them, and no forecasts for whether they are realistic.

**Planning** is a response to forecasts and goals. Planning involves determining the appropriate actions that are required to make your forecasts match your goals.

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organisations require short-, medium-, and long-term forecasts, depending on the specific application.

**Short-term forecasts** are needed for scheduling of personnel, production and transportation. As part of the scheduling process, forecasts of demand are often also required.

**Medium-term forecasts** are needed to determine future resource requirements in order to purchase raw materials, hire personnel, or buy machinery and equipment.

**Long-term forecasts** are used in strategic planning. Such decisions must take account of market opportunities, environmental factors and internal resources.

An organisation needs to develop a forecasting system involving several approaches to predicting uncertain events. Such forecasting systems require the development of expertise in identifying forecasting problems, applying a range of forecasting methods, selecting appropriate methods for each problem, and evaluating and refining forecasting methods over time. It is also important to have strong organisational support for the use of formal forecasting methods if they are to be used successfully.

Throughout the 20th century, businesses tended to use relatively simple forecasting methods that were often not based on statistical modelling. However, the use of statistical forecasting has been growing over the last 20 years, and more recently, some machine learning methods have become popular.

Let the historical time series data be denoted by $y_1, \ldots, y_T$, and the forecast of $y_{T+h}$ be given by $\hat{y}_{T+h|T}, h > 0$.

---

[*]Professor of Statistics, Department of Econometrics & Business Statistics, Monash University, Australia. For biography, *see* the entry **Forecasting overview**.

## 2   Naïve methods

Naïve forecasting methods are very simple, but often provide better forecasts than more sophisticated methods.

**Naïve forecasting** is where the forecasts of all future values of a time series are set to be equal to the last observed value: $\hat{y}_{T+h|T} = y_T$, $h = 1, 2, \ldots$. If the data follow a random walk process ($y_t = y_{t-1} + e_t$ where $e_t$ is white noise[1]), then this is the optimal method of forecasting. Consequently, it is popular for stock price and stock index forecasting, and for other time series that measure the behaviour of a market that can be assumed to be efficient.

**Seasonal naïve forecasting** is a variation of naïve forecasting where the forecast is set to be equal to the last observed value from the same season (e.g., with monthly data, each December is forecast using the last observed December value): $\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$, where $m$ = the seasonal period and $k$ is the integer part of $(h-1)/m$, the number of complete years in the forecast period prior to time $T + h$. This is optimal if the data follow a seasonal random walk process, $y_t = y_{t-m} + e_t$ where $e_t$ is white noise.

## 3   Exponential smoothing methods

Exponential smoothing was proposed in the late 1950s (Brown, 1959; Holt, 1957), and has motivated some of the most successful forecasting methods.

In **simple exponential smoothing**, forecasts can be computed recursively as each new data point is observed:
$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1},$$
where $0 < \alpha < 1$. (Longer-term forecasts are constant: $\hat{y}_{t+h|t} = \hat{y}_{t+1|t}$, $h \geq 2$.) Consequently, only the most recent data point and most recent forecast need to be stored. This was an attractive feature of the method when computer storage was expensive. The method has proved remarkably robust to a wide range of time series, and is optimal for several processes including the ARIMA(0,1,1) process (Chatfield et al., 2001).

**Holt's linear method** (Holt, 1957) is an extension of simple exponential forecasting that allows a locally linear trend to be extrapolated. Forecasts are given by $\hat{y}_{t+h|t} = \ell_t + hb_t$ where

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}),$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1},$$

and the two parameters $\alpha$ and $\beta$ must lie in $[0, 1]$. Here $\ell_t$ denotes the level of the series and $b_t$ the slope of the trend at time $t$.

For seasonal data, a popular method is **Holt-Winters' method**, also introduced in Holt (1957), which extends Holt's method to include seasonal terms. Then $\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$, where

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}),$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1},$$
$$s_t = \gamma(y_t - \ell_t) + (1 - \gamma)s_{t-m},$$

$k$ is the integer part of $(h-1)/m$, and the three parameters $\alpha$, $\beta$ and $\gamma$ all lie in [0,1].

There is also a multiplicative version of the Holt-Winters' method, and damped trend versions of both Holt's linear method and Holt-Winters' method (Hyndman and Athanasopoulos, 2021). None of these methods were explicitly based on underlying time series models, and as a result, the estimation of parameters and the computation of prediction intervals was often not done, or was handled via heuristics, for the first 40–50 years after they were developed.

However, all the above methods have since been shown to be optimal for the **ETS** class of state space models (Hyndman et al., 2008), and maximum likelihood estimation of parameters, statistical model selection and computation of prediction intervals is now widespread.

Hyndman et al. (2002) used this state space framework to develop an automatic forecasting algorithm that could handle any series containing a mix of trend and seasonal components. It quickly became very

---

[1]White noise is a series of iid random variables with zero mean

widely used in business contexts, and variants of the algorithm are now implemented in most forecasting software.

# 4 ARIMA models

Another widely-used model class for forecasting is the class of **ARIMA** (AutoRegressive Integrated Moving Average) models. Rather than describing a series using unobserved level, slope and seasonal components, these models use lagged values of the observed data and the stochastic noise to model autocorrelations in the data. For example, a model can be of the form

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t$$

where $p$ determines the number of lagged response variables to include, $q$ determines the number of lagged noise terms to include, and $e_t$ is white noise. There are also variations that handle differenced data, and seasonal terms. These models were popularised by Box and Jenkins (1970), and so they are sometimes called "Box-Jenkins" models.

Historically, the fitting process for such models was manually-intensive and the resulting forecast were often worse than the simpler methods outlined above. But an automatic modelling algorithm proposed by Hyndman and Khandakar (2008) proved to be fast and effective for a wide range of time series, and led to these models becoming much more widely used in business contexts. Now most ARIMA modelling is done using the Hyndman-Khandakar algorithm, or a variant of it.

# 5 Explanatory variable models for forecasting

The use of explanatory models in business forecasting is also popular, particularly in data-rich contexts where predictor variables help explain the variation in the variable to be forecast.

For example, **linear regression** modelling is widely used (e.g., Pardoe, 2021) where a variable to be forecast is modelled as a linear combination of potential predictor variables:

$$y_t = c_0 + \sum_{j=1}^{J} c_j x_{j,t} + e_t,$$

and $e_t$ denotes a white noise process. Whenever a regression model is used for forecasting, the predictors themselves must also be forecast if they are not known in advance. This limits the value of such models for forecasting. Consequently, it is far more common for regression modelling to be used to explain historical variation than for it to be used for forecasting purposes.

In some domains, the use of **nonparametric additive** models for forecasting is growing (e.g., Hyndman and Fan, 2010). Here, the model is often of the form

$$y_t = \sum_{j=1}^{J} f_j(x_{j,t}) + e_t,$$

where $f_j$ is a smooth nonlinear function to be estimated nonparametrically. The Facebook Prophet model Taylor and Letham (2018) is of this form.

In advertising, there is a well-developed culture of using **distributed lag** regression models (e.g., Hanssens et al., 2001) such as

$$y_t = \sum_{j=1}^{J} \alpha \lambda^j x_{t-j} + e_t,$$

where $x_t$ denotes advertising expenditure in month $t$, $0 < \lambda < 1$ and $\alpha > 0$.

It is also possible to combine ARIMA models with linear regression models, to form **dynamic regression** models of the form

$$y_t = c_0 + \sum_{j=1}^{J} c_j x_{j,t} + n_t$$

where $n_t$ is an ARIMA process. See Hyndman and Athanasopoulos (2021) for further details.

# 6 Machine learning methods for business forecasting

Outside of traditional statistical modelling, an enormous amount of forecasting is now done using machine learning methods. Most of these methods have no formal statistical model, prediction intervals are not computed, and there is limited model checking.

Until the last few years, machine learning methods have performed quite poorly in forecasting competitions. For example, in the M3 competition (Makridakis and Hibon, 2000), the neural network submission didn't beat the simple benchmark methods. But in the M4 competition, the two top-performing methods combined machine learning and statistical modelling in a hybrid fashion (Makridakis et al., 2020). Then, in the M5 competition (Makridakis, Spiliotis and Assimakopoulos, 2022), almost all of the top methods used LightGBM for performing nonlinear regression with gradient boosted trees (Ke et al., 2017).

A recent review of the methods used in forecasting competitions (Makridakis, Spiliotis, Assimakopoulos, Semenoglou, Mulder and Nikolopoulos, 2022) concluded:

- Combining forecasting methods improves forecast accuracy (see also Wang et al., 2022);

- With enough data, deep learning methods outperform statistical and other machine learning methods by about 6%;

- Statistical models perform better at shorter forecast horizons, while deep learning performs better at longer horizons.

# References

Box, G. E. P. and Jenkins, G. M. (1970), *Time Series Analysis: Forecasting and Control*, 1st edn, Holden-Day, San Francisco.

Brown, R. G. (1959), *Statistical forecasting for inventory control*, McGraw-Hill, New York.

Chatfield, C., Koehler, A. B., Keith Ord, J. and Snyder, R. D. (2001), 'A new look at models for exponential smoothing', *Journal of the Royal Statistical Society: Series D* **50**(2), 147–159.

Hanssens, D. M., Parsons, L. J. and Schultz, R. L. (2001), *Market Response Models: Econometric and Time Series Analysis*, 2nd edn, Kluwer Academic Publishers, Boston, MA.

Holt, C. C. (1957), Forecasting trends and seasonals by exponentially weighted averages, O.N.R. Memorandum 52/1957, Carnegie Institute of Technology.

Hyndman, R. J. and Athanasopoulos, G. (2021), *Forecasting: principles and practice*, 3rd edn, OTexts, Melbourne, Australia.
**URL:** *OTexts.org/fpp3*

Hyndman, R. J. and Fan, S. (2010), 'Density forecasting for long-term peak electricity demand', *IEEE Transactions on Power Systems* **25**(2), 1142–1153.

Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *J Statistical Software* **26**(3), 1–22.

Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2008), *Forecasting with exponential smoothing: the state space approach*, Springer-Verlag, Berlin.
**URL:** *www.exponentialsmoothing.net*

Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S. (2002), 'A state space framework for automatic forecasting using exponential smoothing methods', *International J Forecasting* **18**(3), 439–454.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017), LightGBM: A highly efficient gradient boosting decision tree, *in* I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds, 'Advances in Neural Information Processing Systems', Vol. 30, Curran Associates, Inc.
**URL:** *https://bit.ly/lightgbm*

Makridakis, S. G. and Hibon, M. (2000), 'The M3-Competition: results, conclusions and implications', *International Journal of Forecasting* **16**(4), 451–476.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2020), 'The M4 competition: 100,000 time series and 61 forecasting methods', *International Journal of Forecasting* **36**, 54–74.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2022), 'M5 accuracy competition: Results, findings, and conclusions', *International Journal of Forecasting* **38**(4), 1346–1364.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A.-A., Mulder, G. and Nikolopoulos, K. (2022), 'Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward', *Journal of the Operational Research Society* . in press.
**URL:** *https://doi.org/10.1080/01605682.2022.2118629*

Pardoe, I. (2021), *Applied regression modeling*, 3rd edn, John Wiley & Sons, Hoboken, NY.

Taylor, S. J. and Letham, B. (2018), 'Forecasting at scale', *The American Statistician* **72**(1), 37–45.

Wang, X., Hyndman, R. J., Li, F. and Kang, Y. (2022), 'Forecast combinations: an over 50-year review', *International J Forecasting* . in press.
**URL:** *https://doi.org/10.1016/j.ijforecast.2022.11.005*