Forecasting with Excel

Victor Grech[1,*]

ABSTRACT

Introduction: Time series analysis is used by statisticians to make predictions from time-ordered data. This is crucial for planning for the future. The inclusion of little-known forecasting function in Excel™ has brought this type of analysis within the ability of less mathematically sophisticated individuals, including doctors. There are two main models for time series analysis: ARIMA (Autoregressive Integrated Moving Average) and exponential smoothing. This paper will demonstrate how the ubiquitous Excel facilitates a little-known sophisticated forecasting technique that employs the latter and presents a facilitating spreadsheet.

Methods: Excel's FORECAST.ETS function was invoked with supporting macros.

Results: A bespoke spreadsheet was created that would prompt for data to be pasted in columns A and B, formatted as a valid date in A and data in B. After error trapping and a horizon date, the FORECAST.ETS function calculates forecasts with 95% CI and a line graph. The FORECAST.ETS.CONFINT was also invoked using a macro to obtain a 95, 96, 97, 98 and 99% confidence intervals table.

Discussion: Forecasting is vital in all fields, including the medical field, for innumerable reasons. Statisticians are capable of far more sophisticated time series analyses and techniques and may use multiple techniques that are beyond the competence of ordinary clinicians. However, the sophisticated Excel tool described in this paper allows simple forecasting by anyone with some knowledge of this ubiquitous software. It is hoped that the spreadsheet included with this paper helps to encourage colleagues to engage with this simple-to-use Excel function.

KEYWORDS

health services needs and demand; statistics & numerical data; trends; health Workforce; models; statistical; Excel; forecasting; exponential smoothing; holt-winters test

AUTHOR AFFILIATION

[1] Mater Dei Hospital, Malta

* Paediatric Dept, Mater Dei Hospital, Malta; victor.e.grech@gov.mt

Received: 13 April 2024
Accepted: 14 October 2024
Published online: 18 February 2025

Acta Medica (Hradec Králové) 2024; 67(3): 87–90
https://doi.org/10.14712/18059694.2025.3

INTRODUCTION

Time series analysis and forecasting techniques are used by statisticians and data scientists to make predictions from time-ordered data. This type of data consists of a series of observations over time, such as patient admissions, weather patterns, or sales figures. Forecasting is crucial for planning, allowing the anticipation and preparation for future trends, opportunities, and challenges and for providing benchmarks against which performance can be compared. Forecasting also allows correct resource allocation which may be financial, manpower, and materials, thereby minimising or possibly eliminating shortages or surpluses (1).

Forecasting is vital in medicine for many reasons. It is crucial in resource allocation, predicting admissions, disease outbreaks, or surgical procedures, allowing hospitals and healthcare facilities to efficiently allocate resources such as hospital beds, staff, medications, and equipment. By forecasting patient volumes and case types, hospitals and healthcare facilities can better plan staffing requirements at any given time. Forecasts also predict demand for medical supplies such as medications, vaccines, and medical equipment, managing inventory levels, reducing wastage, and ensuring availability during emergencies or spikes in demand. Forecasting may also help to predict disease outbreaks based on historical data, environmental factors, and population trends, allowing early intervention/prevention and resources allocation to control spread. This includes public health preparedness for pandemics, natural disasters, or bioterrorism events by estimating healthcare demands and required resources should such calamities occur. With regard to treatment, forecasting may assist in planning treatment options for patients based on predicted outcomes, thus optimizing care pathways, and ultimately improving patient outcomes. Forecasting may also predict readmission rates, complications, and disease progression, allowing for the preparation of personalized, timely interventions.

Forecasting may also be useful in clinical research, in clinical trials for example, by estimating patient recruitment rates and disease progression, thus helping to plan studies more effectively. Ultimately, forecasting can help to optimise cost management and reduce waste in healthcare services, a universal problem due to ever-increasing demands with finite budgets, a quandary that threatens the viability and sustainability of medical systems worldwide.

A little-known forecasting function in Excel™ has brought this type of analysis within the ability of mathematically unsophisticated individuals, including doctors.

Time series analysis entails the usual steps in data collection and data pre-processing to ensure that the data to be analysed is clean and with the least possible missing values, outliers, and other anomalies. It is also always wise to perform an initial exploratory data analysis to understand the dataset's underlying patterns, and plotting the data is almost a mandatory step (2). A time series analysis is characterised by the following patterns or components:

Level represents the baseline value of the data and is updated at each step by combining the observed value with the estimated level from the previous time step. It thus represents the long-term average of the dataset.

Trend represents the underlying long-term trend direction and rate of change in the data, without the inclusion of seasonal fluctuations (level/upward/downward).

Seasonality refers to regular, predictable patterns that repeat at fixed intervals within a year. These patterns may be daily, monthly, quarterly, or yearly.

Cyclicity refers to fluctuations that occur over an extended period, often more than a year, and are not tied to specific seasons. Unlike seasonal component/s, cyclical component/s do not have fixed periodicity.

The irregular(residual/noise) component consists of random fluctuations or noise in the data that cannot be explained by the above and includes unpredictable events, measurement errors, or other irregular influences. All these components interact to produce the eventual observations/dataset. The data can therefore also be decomposed from a time series into these individual constituents to better understand the structure and behaviour of the time series. and hence facilitate forecasting. There are two main models for time series analysis.

ARIMA

ARIMA (Autoregressive Integrated Moving Average) techniques are widely used in time series forecasting due to their ability to handle various types of data patterns. The method combines three critical components: the autoregressive (AR) component, the moving average (MA) component, and an integrated (I) aspect. The autoregressive element implies that the current data point is influenced by its preceding values, making it a powerful tool for identifying trends and patterns over time. Essentially, AR models explain the variability of the present value based on a weighted sum of prior data points. This relationship allows ARIMA to capture momentum and fluctuations within a time series. The moving average (MA) component relies on a sequence of past forecast errors to correct the model. By taking a linear combination of several points from the recent past, the model reduces noise, smoothing out short-term variations and allowing for a more accurate prediction of future values. This method captures short-term shocks and fluctuations by adjusting predictions based on previous errors, making the model more responsive to sudden changes in data. The "Integrated" part of ARIMA refers to the process of differencing the data to make it stationary. Time series data often exhibit trends or seasonality that ARIMA models need to eliminate through differencing. This mathematical transformation ensures that any trend is removed, and the data's mean and variance remain constant over time. By applying this process, ARIMA models can adapt to datasets with a wide variety of behaviors, including those with long-term trends or seasonal effects. ARIMA's versatility stems from its capacity to combine these three elements – autoregression, moving averages, and integration – into a single, cohesive framework. As a result, ARIMA models are highly adaptable to different types of data, making them suitable for applications ranging from economic forecasting to environmental data analysis. They are flexible enough to handle datasets with missing values or irregular intervals, and they exhibit robustness in the face of outliers, allowing for reliable forecasting even when dealing with complex real-world data. Although Excel does not natively support ARIMA modeling, it is possible to apply these techniques using specialized add-ins (3, 4).

EXPONENTIAL SMOOTHING

Exponential methods smooth datasets by using exponential functions such that the weighting of individual datapoints on subsequent data exponentially decreases over time, giving more weight to more recent observations. Techniques include simple exponential smoothing, double exponential smoothing (Holt's method), and triple exponential smoothing (Holt-Winters method) with the latter taking into account level, trend and seasonality (5–8). In brief, this technique identifies the level, trend, and seasonal components, updates each of these at each time step by exponential smoothing and uses these updated elements to forecast future steps. The method requires the entry of parameters such as smoothing coefficients and the lengths of seasonal periods. However, the software used to perform this test may be able to self-optimise these parameters for best fit.

PROS AND CONS

The Holt-Winters technique is arguably better at capturing and forecasting seasonality, and is simpler to use, with fewer parameters needing to be input. It is also generally more accurate for short-term predictions. Furthermore, an ARIMA forecast is often a straight line devoid of patterns, showing the mathematical model of best fit unless a more complex SARIMA (Seasonal Autoregressive Integrated Moving Average) is used (an extension of the non-seasonal ARIMA model, designed to handle data with seasonal patterns), while a Holt-Winters analysis will create forecasts that closely resemble the original data. In practice, the choice between the two depends on the properties of the dataset and the horizon required to forecast to (9, 10). Vis-à-vis implementing ARIMA analysis in Excel, the utilisation of an add-in to Excel may slow down Excel during daily use.

For complex analyses, it may be necessary to use multiple methods including hybrid approaches, but this is beyond the scope of this paper. Machine learning models and artificial intelligence are also being harnessed, but these are also beyond the scope of this paper.

This paper will demonstrate how the ubiquitous Excel contains a little known and little appreciated but quite sophisticated forecasting technique that employs the Holt-Winters method. This paper also presents a spreadsheet that facilitates the software's performance of this test.

METHODS

EXCEL FORECAST.ETS FUNCTION

Excel's FORECAST.ETS function (version 2016 and above) forecasts using the so-called AAA (additive error, additive trend, and additive seasonality) version of the Exponential Triple Smoothing (ETS) algorithm.

The syntax is: FORECAST.ETS(target_date, values, timeline, [seasonality], [data_completion], [aggregation])

Alternatively, the function can be accessed without invoking the function directly from the Data ribbon, in the Forecast group: activate Forecast Sheet. The user is asked to set the forecast date horizon and choose between a line chart or a column chart for graphical output which is also produced with 95% confidence intervals. The latter can be varied at this point as well (11).

This function requires data at regular intervals (hourly, daily, monthly, quarterly, yearly) and is best suited for non-linear data with seasonal or other repetitive pattern/s. When the function fails to detect a pattern, the forecast output is linear. The function is robust up to a point with incomplete datasets, where up to 30% of data points may not available.

RESULTS

A spreadsheet was created that would prompt for data to be pasted in columns A and B, formatted as a valid date in column A and data in column B. The sheet also finds and displays the date in the last row of column A, and the user is expected to input a valid date to forecast to. A macro was created that performed error trapping and gives the user appropriate prompts (e.g., invalid forecast date format or date to forecast is less than the last date in the dataset). The macro invokes the FORECAST.ETS function with the following (default) parameters, with the calculation of 95% CI for the forecast dates and the creation of a line graph with 95% CI.

ActiveWorkbook.CreateForecastSheet Timeline:= Sheets("Forecast sheet").Range("A1:A" & lr), Values:=Sheets("Forecast sheet").Range("B1:B" & lr), ForecastEnd:=lrend, ConfInt:=0.95, Seasonality:=1, ChartType:=xlForecastChartTypeLine, Aggregation:= xlForecastAggregationAverage, DataCompletion:=xlForecastDataCompletionInterpolate, ShowStatsTable:=True

In this macro, lr is the last column row (for columns A and B), a variable which is obtained by using another function, and this therefore ensures that the entire data range is captured by the macro and used for analysis, whatever the column length.

The calculation of confidence intervals for the predicted forecast at each future date is also possible at any desired level with this function. For monthly data, the FORECAST.ETS function is therefore used for the actual prediction for a 10 year period and the FORECAST.ETS. CONFINT(range, CI) is used to obtain 95, 96, 97, 98 and 99% confidence intervals. These are, individually (at every next predicted value), added to and subtracted from the FORECAST.ETS output, allowing the calculation of confidence intervals for each of these five levels. A macro was created to import the dataset into this sheet so as not only to use the Excel standard function and output (table and graph) but also to tabulate these five confidence levels. This allows users to identify any outlier/s in the dataset by comparing to the sheet's output, and by locating the divergence away from the estimated average, be able to calculate p values based on how many percentiles the outlier/s are from the estimated mean.

The sheet is available for download from this location: https://tinyurl.com/3nbf9rba

DISCUSSION

It must be reiterated that statisticians and data scientists are capable of far more sophisticated time series analyses and techniques and indeed, may use multiple techniques that are beyond the competence of ordinary clinicians (12–14). Excel natively performs forecasting using the Holt-Winters test and the Excel tool described in this paper allows simple forecasting by anyone with some knowledge in using Excel. It is hoped that the spreadsheet included with this paper helps to encourage colleagues to engage with this simple-to-use Excel function.

In summary, forecasting provides valuable insights into the future, enabling better decision-making, risk management, and planning across various domains. It allows individuals and organizations to anticipate change, adapt to new circumstances, and thrive in an uncertain world, and this software may be of help.

CONFLICTS OF INTEREST

The author runs the Write a Scientific Paper course (WASP), for which the spreadsheet presented in this paper was created.

REFERENCES

1. Montgomery DC, Jennings CL, Kulahci M. Introduction to Time Series Analysis and Forecasting. New York: John Wiley & Sons, Ltd, 2015.
2. Grech V. Write a Scientific Paper (WASP): Effective graphs and tables. Early Hum Dev; 134. Early Hum Dev. 2019 Jul: 134: 51–54.
3. Box GEP, Jenkins GM, Reinsel GC, et al. Time Series Analysis: Forecasting and Control. New York: John Wiley & Sons, Ltd, 2015.
4. Zaiontz C. Real Statistics Using Excel, www.real-statistics.com (2024).
5. Brown RG. Exponential Smoothing for Predicting Demand. Cambridge (MA): Little, 1956.
6. Holt CC. Forecasting trends and seasonals by exponentially weighted averages. O.N.R. Research Memorandum, 1957.
7. Winters PR. Forecasting sales by exponentially weighted moving averages. Manage Sci. 1960; 6: 324–342.
8. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast. 2004; 20: 5–10.
9. Chaurasia V, Pal S. Application of machine learning time series analysis for prediction COVID-19 pandemic. Res Biomed Eng. 2020; 1–13.
10. Mgale YJ, Yan Y, Timothy S. A comparative study of ARIMA and Holt-Winters exponential smoothing models for rice price forecasting in Tanzania. Open Access Libr J. 2021; 8: 1–9.
11. Rahardja D. Statistical Time-Series Forecast via Microsoft Excel (FORECAST. ETS) Built-In Function. J Res Appl Math. 2021; 7: 69–73.
12. Vieira A, Sousa I, Dória-Nóbrega S. Forecasting daily admissions to an emergency department considering single and multiple seasonal patterns. Healthc Anal. 2023; 3: 100146.
13. Klein B, Zenteno AC, Joseph D, et al. Forecasting hospital-level COVID-19 admissions using real-time mobility data. Commun Med. 2023; 3: 25.
14. Zhou L, Zhao P, Wu D, et al. Time series model for forecasting the number of new admission inpatients. BMC Med Inform Decis Mak. 2018; 18: 39.