Applied Statistics 2024-25

Final Project

Statistics for Business Students - Chapters 10/11 Review

Electrical & Computer engineering
University of Thessally, Volos

Instructor: Elias Houstis

Kakepakis Georgios - 03198
Pisxos Vasileios - 03175
Team 4

1 Chapter 10: Introduction to time series data, long-term forecasts and seasonality

1.1 Introduction

Time series are time-based variables measured in equidistant units of time and listed in time order (stock prices, weather data, quarterly sales, ...). This chapter aims to apply various time series methods for long-term forecasting and extrapolation. In many ways this is an extension of the **Linear Regression**, using just a **single variable** measured in time as the dependent variable, and the depended variable is just sequential units representing the time. The objective of forecasting is essentially to extrapolate the variable into the future, based on its **history**.

1.2 Stationarity

A **stationary** time series is one whose **statistical properties** such as mean, variance, covariance, and standard deviation **do not vary with time**. Stationary time series oscillate around this fixed mean value. If that is not the case, the time series is **non-stationary**. Below we see an example of each of these 2 categories:

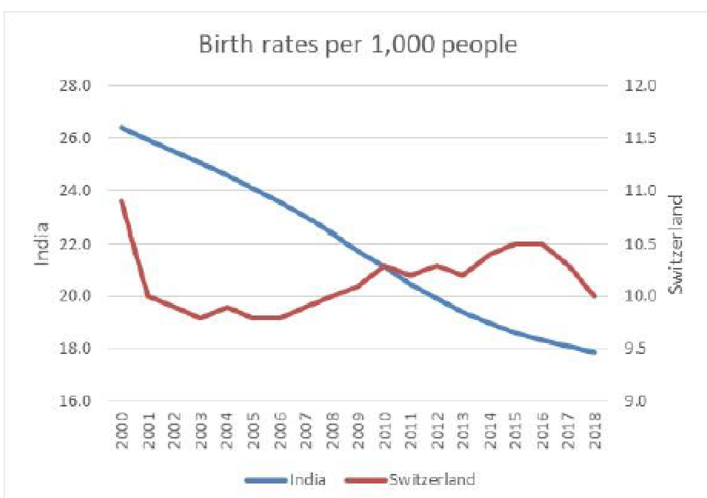

Figure 1: Stationary VS Non-Stationary Time Series

1.3 Seasonality

A **seasonal** time series is one shows some **repeated patterns** over a specific amount of time units (periodicity/season). If the season is a longer time interval (years) we call the time series **cyclical**. Below we see an example of a seasonal time series:

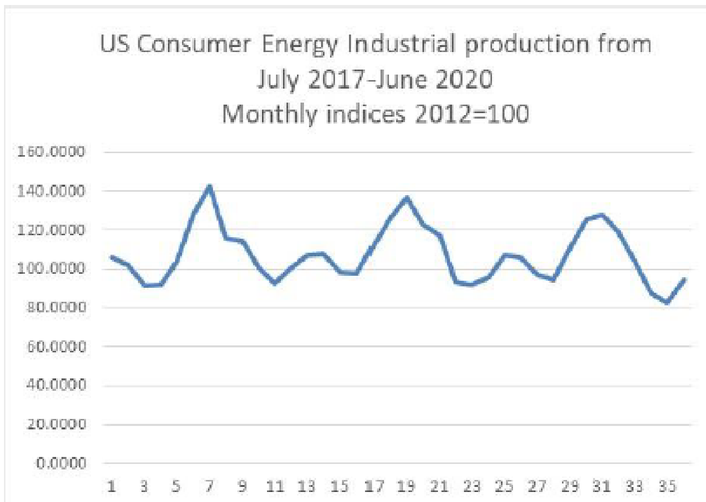

Figure 2: Seasonal Time Series

1.4 Trend Extrapolation

Trend extrapolation is essentially the **long-term forecasting** of the variable, meaning the prediction of the general direction and the speed at which the future values of the variables are going to happen, and not the necessarily every detail of the value. In this chapter we will use **linear** trends. By extrapolating a linear trend and use it for forecasting we aticipate that our forecast values will not be completely accurate, but it represents well the direction and how steep the movements of the variables are.

In order to extract the linear trend we just apply a **linear regression** to the data. This means that the trend can be described as: $y = a \cdot x + b$, and the formulas to find the coefficients are:

$$\begin{cases} a = \bar{y} - b \cdot \bar{x}, \\ b = \dfrac{\Sigma x_i \cdot y_i - \bar{x} \cdot \bar{y} \cdot n}{\Sigma x_i^2 - \bar{x}^2 \cdot n} \end{cases}$$

Below we see the extrapolated linear trend of a time series:

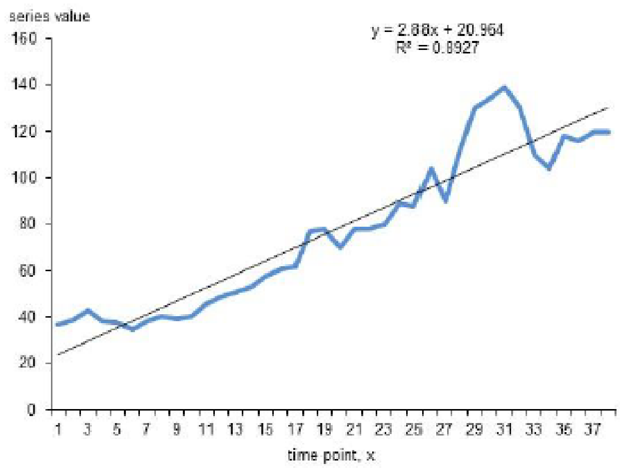

Figure 3: Time Series Trend Extrapolation

1.5 Error Measurements

An error, or **forecasting error**, is the difference between the actual data and the data produced by a model: $e_t = y_t - \hat{y}_t$. They tell us how good our model is. We want the errors to resemble white noise and to look like they have a **normal distribution** for our model to be considered good.

A variety of error measurements can be used to assess how good the forecasts are. Below we present 6 frequent error statistics: ($e_t = y_t - \hat{y}_t$)

- **Mean Error**: $ME = \frac{1}{n} \cdot \Sigma e_t$
- **Mean Absoloute Error**: $MAE = \frac{1}{n} \cdot \Sigma |e_t|$
- **Mean Squared Error**: $MSE = \frac{1}{n} \cdot \Sigma e_t^2$
- **Root Mean Squared Error**: $RMS = \sqrt{MSE} = \sqrt{\frac{1}{n} \cdot \Sigma e_t^2}$
- **Mean Percentage Error**: $MPE = \frac{1}{n} \cdot \Sigma \frac{e_t}{y_i}$
- **Mean Absolout Percentage Error**: $MAPE = \frac{1}{n} \cdot \Sigma \frac{|e_t|}{y_i}$

1.6 Prediction Interval

In previous chapters, in order to make an estimate that the true mean is somewhere in a given interval, we used a **confidence interval** of the structure: $CI : \bar{x} \pm z \cdot SE$. The value of z depends on the confidence level. If the sample was relatively small we used a t-value instead of a z-value: $CI : \bar{x} \pm t \cdot SE$. The value of t depends both on the confidence level and the degrees of freedom.

We use the same concept on time series to predict the interval of the forecasted value. In order to do that, we need to define the **standard error** of the estimate of the predicted values: $SE_{\hat{y},y} = \sqrt{\frac{1}{n-2} \cdot \Sigma(y_i - \hat{y}_i)^2}$. We can modify the equations of the confidence intervals as: $CI : \bar{x} \pm z \cdot SE_{\hat{y},y}$ / $CI : \bar{x} \pm t \cdot SE_{\hat{y},y}$. Below we see a prediction and its corresponding prediction interval:

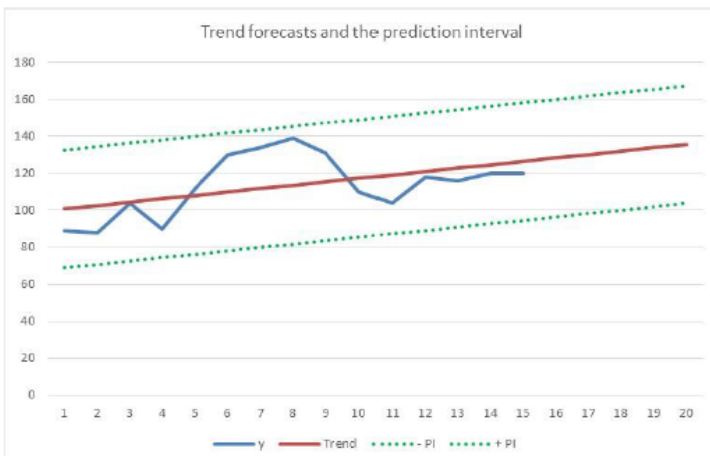

Figure 4: Forecasting with Prediction Interval

The problem of this approach is that it does not comply with the assumption that the **interval should get wider the further we get to the future**. In order to achive that, we need to replace the standard error equation with: $SE_{\hat{y},x} = SE_{\hat{y},y} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$. Below we see the same prediction and its updated prediction interval:

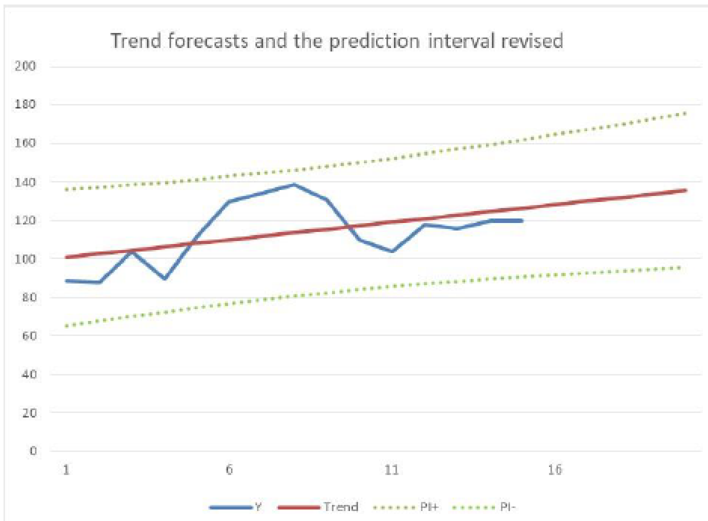

Figure 5: Forecasting with Prediction Interval

1.7 Time Series Decomposition

Until now we assumed that the data consists of only the trend and the residuals (everything else), that need to be randomly fluctuating around the trend line. Sometimes however no matter what we do the residuals are not random and they show a regularity in movements. This is why we also introduce the **seasonal component** for these patterns, since this simplistic approach is not enough. The classical **time series decomposition** method assumes that **every time series can be decomposed** to the following elements:

- Trend (T)
- Cyclical variations (C)
- Seasonal variations (S)
- Irregular variations (R)

We can assume either an **additive model**: $Y = T + C + S + I$ (better for stationary data), or a **multiplicative model**: $Y = T \cdot C \cdot S \cdot I$ (better for non-stationary data). For more complex data we can also use a **mixed model**, for example: $Y = (T \cdot C \cdot S) + I$.

The process of isolating these different components is called the **classical time series decomposition method**. Assuming a multiplicative model ($Y = T \cdot C \cdot S \cdot I$), we can find the trend using linear regression as we did above, so if we divide with the trend we get: $C \cdot S \cdot I$. Also, we assume that the seasonal component is either cyclical or seasonal based on the value of the periodicity. If we assume that it is seasonal, we are left with: $S \cdot I$, which is essentially the seasonal component "polluted" with the residuals.

To get the seasonal component we can find the **average** value of every corresponding cycle of different periods. This removes the irregular variations. What is left are the irregular variations / residuals.

In order to find the **prediction interval** for the seasonal time series, we will use a modified version of Root Mean Squared Error as the **standard error**. Specifically, we use: $RMSE_h = \sqrt{h \cdot MSE}$, in order to assure again the assumption that the interval should get wider the further we get to the future. Below we see an example of a seasonal prediction and its corresponding prediction interval:

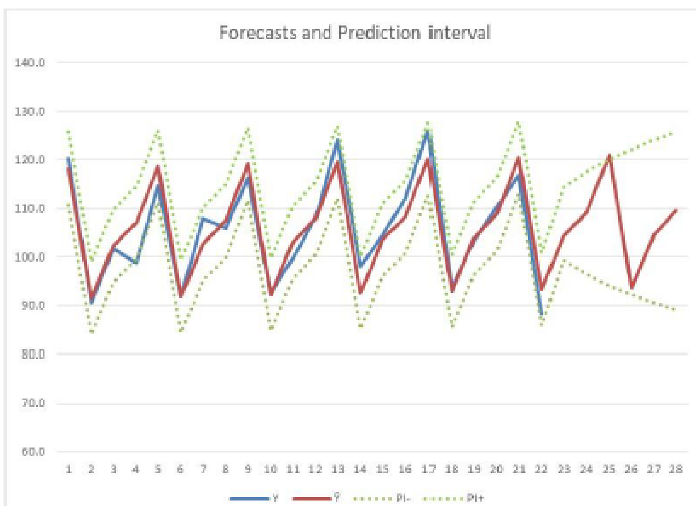

Figure 6: Forecasting with Prediction Interval

2 Chapter 11: Short and medium-term forecasts

2.1 Introduction

This chapters focuses on **short-term** (1 period ahead) and **medium-term** (2-6 periods ahead) forecasts, meaning that now we care about the accurcy of the forecasts and not just the direction as before.

2.2 Simple Moving Average - Short-Term Prediction

Moving average is technique that essentially "smooths" out the orignal time series. It is a **dynamic averge** that change depending on the number of periods for which they are calculated. We use it because when we have a stationary time series the average value is a good predictor, but this is not the case for a non-stationary time series.

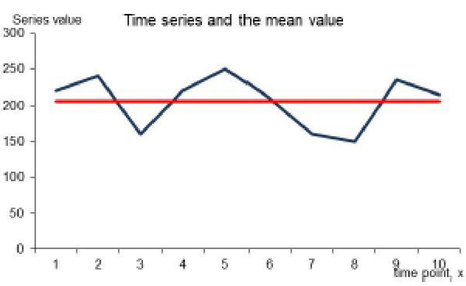

Figure 7: Stationary Time Series

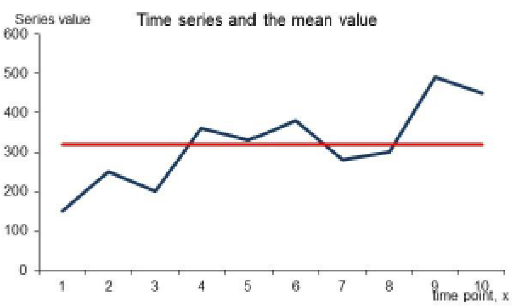

Figure 8: Non-Stationary Time Series

A general formula for the moving average is: $M_t = \frac{\Sigma x_i}{N}$. Another useful formula is: $M_t = M_{t-1} + \frac{x_t - x_{t-N}}{N}$. As we use a larger value for N, the moving average tends to the overall average and it becomes smoother. We can see that on the example below:

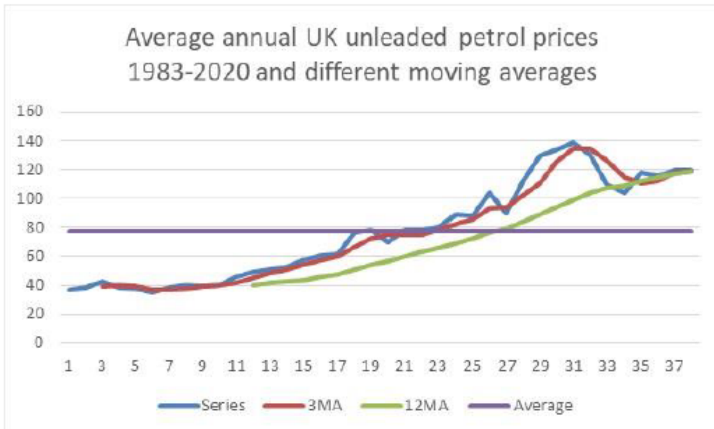

Figure 9: Moving Average

We can use the moving average as a **forecasting tool** by shifting it by one period in the future. This means that we perform a **short-term** forecast since moving average cannot extend the forecast beyond 1 period.

2.3 Double Moving Average - Medium-Term Prediction

Double moving average is essentially a moving average on the values of the moving average: $M''_t = \frac{\Sigma M'_t}{N}$. Using a double moving average we can construct dynamic coefficients (slope/intercept) which will move and fluctuate along with the time series as: $\hat{y}_{t+1} = a_t + b_t \cdot m$, where:

- intercept: $a_t = 2M'_t - M''_t$
- slope: $b_t = \frac{2}{N-1}(M'_t - M''_t)$

These values remind us of the linear regression / trend extrapolation method, but they differ because they are **dynamic** now.

2.4 Simple Exponential Smoothing - Short-Term Prediction

In order to introduce exponential smoothing we make the assumption that: the best predictor of the current value of the variable is the **previous value** plus some **error element**: $\hat{y}_t = y_{t-1} + e_t$.

We use a variable $a \in [0, 1]$ to take a **fraction** of the error, by using the equation $\hat{y}_t = y_{t-1} + a \cdot e_t = y_{t-1} + a \cdot (y_{t-1} - \hat{y}_{t-1})$. By taking a fraction of the error we are discounting the influence that every previous observation and its associated error has on the current observation. If we expand this formula for previous values of the variable we see that we are effectively assigning an exponentially dropping weight on the previous observations. We can find the fraction value by the formula $a = \frac{2}{M+1}$.

As we did with the moving average method, we can use the exponential smoothing as a **forecasting tool** by shifting it by one period in the future. This means again that we perform a **short-term** forecast since moving average cannot extend the forecast beyond 1 period.

2.5 Double Exponential Smoothing - Mid-Term Prediction

Again following the logic of the moving average model, we seek to use the exponential smoothing for mid-term forecasting. We use a **double exponential smoothing**, which is essentially an exponential smoothing on the values of the exponential smoothing: $S''_t = s \cdot S'_t + (1-a) \cdot S''_{t_1}$. Using a double exponential smoothing we can construct dynamic coefficients (slope/intercept) which will move and fluctuate along with the time series as: $\hat{y}_{t+1} = a_t + b_t \cdot m$, where:

- intercept: $a_t = 2S'_t - S''_t$
- slope: $b_t = \frac{2}{N-1}(S'_t - S''_t)$

The formulas of the coefficients are the same as we used on double moving average, but use the double exponential smoothing instead. They also remind us of the linear regression / trend extrapolation method, but they differ because they are **dynamic** now again.

In general we can find the optimal value of a by estimating the MSE value (or another error statistic) for many values of a and keeping the one with the smallest one. This can be done automatically through functions of the statistics software we use (Excel/SPSS/...).

2.6 Holt-Winters' Seasonal Exponential Smoothing

Holt-Winters' method is one of the most effective forecasting methods fot **seasonal data** based on exponential smoothing. It uses 3 smoothing equations and 3 smoothing constans. It uses these components:

- level (l_t)
- trend (b_t)
- seasonality (S_t)

The smoothing constants are: α, β and γ, and they are used in the following equations for the **additive model**:

- $l_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$

- $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$
- $S_t = \gamma(y_t - l_t) + (1-\gamma)S_{t-s}$

The value of 's' denotes the periodicity. The forecasts are then produced as: $F_{t_m} = l_t + b_t \cdot m + S_{t-s+m}$, for 'm' forecasts ahead. For $\beta = 0$ and $\gamma = 0$ the Holt-Winters' method is equal to the exponential smoothing method. Below we see an example of the Holt-Winters' method forecasting and its corresponding prediction interval:

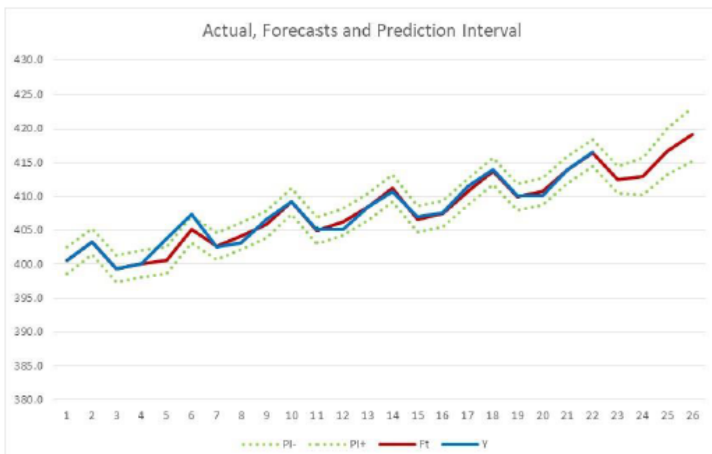

Figure 10: Holt-Winters' method