Introduction to time series analysis

Margherita Gerolimetto

November 3, 2010

1 What is a time series?

A time series is a collection of observations ordered following a parameter that for us is time. Examples occur in a variety of fields (economics, engineering, meteorology and so on...).

For instance, economic time series include share prices on successive days, export totals in successive months, average incomes in successive months, company profits in successive years (see the graph below). Marketing time series are sales figures in successive weeks or months.

2 Terminology

A series is said **continuous** when observations are made continuously in time (the term continuous is used also when the measured variable takes only a discrete set of values). A series is said **discrete** when observations are taken only at specific times, in general equally spaced (the term discrete is used for this kind of series even when the measured variable is a continuous one).

In general we will use discrete time series, that can arise in several ways:

1. Given a continuous time series we can read off (or digitize) the values in equal intervals of time to produce a discrete time series that is a sampled time series (for example, temperature measured every 15 minutes or the price of an asset every day). In this case the time series (especially the economic ones) are called also **flows**.

2. Suppose a variable does not have an instantaneous values, but the values can be aggregated over equal intervals of time (for example, exports
measured monthly, rainfall measured monthly). In this case the series (especially the economic ones) are called **stocks**.

Another introductory comment should be made. Time series analysis has the characteristic of not assuming that the sample consists of iid data (typical of statistical theory). Dependence is, on the contrary, maybe the most important feature of time series data and the analysis must be made taking into account the time order. Indeed, when successive observations are dependent, future values can be predicted by past observations\(^1\).

### 3 Objectives of time series analysis

The main objectives of time series analysis are:

- **Description**: the first step in the analysis is the plot of the data and obtain simple measures to have a look at the main properties of the series. Although much more sophisticated techniques are often used, the analysis of the graph must not be neglected since it can reveal the existence of trend, seasonality, outliers, turning points...

- **Explanation**: this means indentifying the random mechanism that generates the phenomenon of which a sequence of observations are available. In case observations are taken on two or more variables it is possible to explain the variation in one time series on the basis of the variation of another time series (linear systems).

- **Prediction**: given observed time series a typical aim is to predict future values. This is based on the principle that the behaviour of the phenomenon in the past is maintained in the future also.

### 4 Approaches to time series analysis

To practically conduct time series analysis bit further that simple description, it is often necessary to introduce a statistical model for the time series, whose

\(^1\)Of course, in case the time series is stochastic, as it usually is, exact predictions are not possible and are replaced by the idea that future values have a probability distribution, conditioned by the knowledge of past values
specification is arbitrary since it complies with convenience requirements. Stylized behaviours can be:

- **The Error Model:**
  
  \[ x_t = f(t) + \epsilon_t \quad t = 1, 2, \ldots, n \]

  In this case the series \( x_t \) is defined as the resultant of an explicit mathematical function plus a random error term which must comply with the typical assumptions of the regression model:

  \[
  E(\epsilon_t) = 0; \quad E(\epsilon_j^2) = \sigma^2 < +\infty; \quad E(\epsilon_t \epsilon_s) = 0, \forall t \neq s
  \]

  This model is particularly appropriate for regular phenomena. The stochastic component is only conceptually added to the dynamic (consider for example cyclical phenomena for which the function \( f \) is the combination of harmonic waves).

- **The stochastic model:**

  \[ X_t = g(\epsilon_t, \epsilon_{t-1}, \ldots) \]

  In this case the phenomenon is described as a function \( g \) of stochastic variables. This model is stochastic since the mechanism that generates the \( \epsilon \)'s (hence the \( x \)'s) is totally stochastic.

According to how the model is conceived, two broad approaches for the analysis of time series can be considered:

### 4.1 Classical approach to time series analysis

The classical (or traditional) approach is mirrored in the Error Model above presented:

\[ x_t = f(t) + \epsilon_t \quad t = 1, 2, \ldots, n \]

Every time series \( x_t \) can be considered as the result of a combination of different unobserved factors (stochastic and non-stochastic), called the **components** of a time series:

- **Trend** \( T_t \): can be defined as long-term change in the mean level. The meaning of “long term” is quite a crucial point. Granger (1966) writes that a trend in mean could be defined as comprising all cyclic components whose wave length exceeds the length of the observed time series.
• *Cycle* $C_t$: refers to variations exhibited by the time series at fixed period that cannot be considered as seasonality. For example, economic data are often affected by what is called business cycle, with a period varying between 5 and 7 years.

• *Seasonality* $S_t$: refers to variations that are annual in period. For example the consumption of some products.

• *Irregular fluctuations* $a_t$: after trend, cycle and seasonal variation have been removed something is left in the residuals that is not predictable from the past history. Making sure that the $a_t$ component is truly random is a good guarantee that the decomposition of the series into components is correct.

The classical (or traditional methods) of analysis of time series are concerned with the decomposition of the series into Trend, Cycle, Seasonality, Irregular component. In other words, one could say that the objective is to comprehend how function $f(t)$ can be approximated. It is worth noting that the decomposition is, in general, not unique unless certain assumptions are made. The most common choice is to use an additive decomposition:

$$x_t = T_t + C_t + S_t + a_t$$

but the multiplicative decomposition is also used:

$$x_t = T_t C_t S_t a_t$$

The classical approach to time series has a number of advantages:

1. It is based on rather intuitive concepts
2. It is exploitable also when the series length is rather short
3. It is very useful when the series plot reveals that the variability is strongly dominated by a component, for example trend or seasonality

It also has a number of disadvantages:

1. The decomposition is not unique
2. The stochastic part is only in the error term

\footnote{The passage from the additive to the multiplicative is quite easily obtained by applying the logarithm function}
4.2 The modern approach to time series

In the modern approach to time series, the observed time series is conceived as a finite realization of a stochastic process. The stochastic part coincides with the systematic part of the DGP and not simply the error term. This is formally described by the above presented Stochastic Model:

\[ X_t = g(\epsilon_t, \epsilon_{t-1}, ...) \]

Compared to the Error Model, here the \( f(t) \) component is assumed to be absent or preliminarily removed and the focus is on \( g(\epsilon_t, \epsilon_{t-1}, ...) \), a correlated component process that governs the data generating mechanism. An inferential scheme is derived, where the observed data guides the researcher toward the model that looks more appropriate. The disadvantages, typical of the classical approach are here overcome.

The cornerstone of this approach is the Box-Jenkins procedure (Box and Jenkins, 1970).

5 Classical approach: estimate with mathematical functions

A first method to estimate components of an observed time series \( x_t \) is based on mathematical functions. The main assumption is that the hypothesized function governs the dynamic over the entire period.

5.1 Trend

The trend component can be usually well represented by a regular function \( f(t) \) (it could be for example a polynomial function). Starting from the model \( x_t = f(t) + \epsilon_t \), let us distinguish three operative scenarios:

1. If \( f(t) \) is known the issue is only in the estimation of the coefficients. This can be extremely easy in case \( f(t) \) is linear, more complicated in case \( f(t) \) is not linear. Still, a part from the computational difficulties in nonlinear estimation problems, it is only a matter of parameter estimations from the data since the functional form is already specified.

2. If \( f(t) \) is unknown, before estimating the coefficients the issue is to identify a functional that can be a good approximation for the unknown
For example, if the trend appears to fluctuate with wide movements it can be approximated by a polynomial, in case it has a more periodic appearance a combination of trigonometric terms can be an appealing option.

3. If \( f(t) \) is unknown and it is not even possible to derive a acceptable approximation, this is the case of using a different approach from mathematical function, which is based on smoothing techniques (we will see them later on) that are not based on the assumption (very strong!) that a certain functional form holds over the entire sample period.

Suppose to have a trend that can be classified as polynomial, hence we assume that the trend dynamic follows this form:

\[
f(t) = \alpha_0 + \alpha_1 t + ... + \alpha_q t^q
\]

hence

\[
x_t = \alpha_0 + \alpha_1 t + ... + \alpha_q t^q + \epsilon_t
\]

Using least squares estimation methods, it is possible to obtain the estimates of the coefficients (solving the normal equations system). The polynomial that results can be a good guide to interpret the phenomenon. Much more caution must be used, instead, when the aim is to make prediction since there is no warrant that the functional form, supposed to hold over the sample period, does so also outside the time span of the observed time series.

Another interesting case is when \( f(t) \) is supposed to follow an exponential trend

\[
f(t) = \alpha_0 e^{\alpha_1 t}
\]

The logarithm allows the linearization:

\[
\log f(t) = \alpha_0 + \alpha_1 t
\]

Often trend are not representable as linearizable functions. This happens when the phenomenon exhibits phases of extremely fast growth, followed by phases of slow dynamics. This kind of behaviours can be represented by growth curves that are non linear in parameters, a possible example of this category of functions is the logistic function.

\[
f(t) = \frac{e^{\alpha_0 + \alpha_1 t}}{1 + e^{\alpha_0 + \alpha_1 t}}
\]

\[3\]A possible formalization is:
5.2 Seasonality

The regression scheme presented for the trend component can be adopted also for the seasonal component, representable by a periodic function \( g(t) \). Remember that periodic functions are those functions such that the value taken at time \( t \) repeats at regular intervals of period \( s \):

\[
g(t) = g(t + s) = g(t + 2s)
\]

\( s = 4 \) for quarterly data, \( s = 12 \) for monthly data. So the idea is that the data generating process is

\[
x_t = g(t) + \epsilon_t
\]

- A first approach to study \( g \) is to assume that \( g \) is representable as a sum of dummy variables

\[
g(t) = \sum_{j=1}^{S} \gamma_j d_{jt}
\]

where \( d_{jt} \) is 1 in the \( j \)-th period and 0 otherwise. Then \( \gamma_j \) indicates the level of the phenomenon at the \( j \)-th period of the year.

Using least squares it is possible to estimate the \( \gamma_j \) coefficients that are denoted by \( \gamma_j^* \) and called raw seasonal coefficients. Generally, from these coefficients their mean is subtracted to obtain \( \gamma_j^* - \bar{\gamma}_j^* \) since the latter possess the property of summing to 0 at the end of the year, which is compatible with the concept of seasonality (the effect should not last more than one year).

- A second possibility is to assume that the seasonal component is given by a sum of harmonic functions. Generally a sum of 2,3 harmonic waves can be a reasonable option to represent seasonal dynamics sufficiently complex.

6 Classical approach: moving averages

When the analytic description of the phenomenon is not possible, for example because the dynamic exhibited by the phenomenon is extremely irregular,
the moving averages are a good answer. Intuitively, moving averages tries to capture the dynamic of the phenomenon without, with this, necessarily define a functional form that holds over the entire time span as a variation law.

6.1 Main characteristics of moving averages

The moving averages logic can be simply synthesized as follows: to estimate a component using moving averages, it should be applied to the observed series $x_t$ a transformation $m$ such that the desired component is preserved and other fluctuations are smoothed. In this view, moving averages are a weighted sum of observed values of $x_t$ around the generic instant $t$.

$$x^*_t = \sum_{i=-m_1}^{m_2} \theta_i x_{t+i}$$

where $m_1 + m_2 + 1$ is the order of the moving average. If $\theta_i = -\theta_i$ and all $\theta_i = \frac{1}{m_1+m_2+1}$, then the moving average is called simple and if $m_2 = m_2$ it is called centered.

A time series $x_t$ is invariant to the moving average $M$, i.e. if $Mx_t = x_t$, in this case the moving average is said to preserve $x_t$. In general, formal conditions to guarantee that a moving average $M$ preserves a degree $p$ polynomial are the following:

$$\sum_{i=-m_1}^{m_2} \theta_i = 1; \quad \sum_{i=-m_1}^{m_2} i^{r} \theta_i = 0 \quad for \ r = 1, 2, ..., p$$

6.2 Using moving averages

Using moving averages it is possible to:

1. estimate the trend
2. remove seasonality from series
3. reduce the random component

Having in mind the three objectives enumerated above, the following are basics for using moving averages:
1. There are moving averages that preserve the polynomial trend up to a certain order; hence, when estimating the trend component, be careful of using moving averages that preserve it, i.e. choose opportunely the coefficients $\theta$.

2. If a moving average of order $s$ is applied to an observed series characterized by a wave of period $s$, this wave is eliminated; hence, when removing the seasonal component, choose accurately the order of the moving average.

3. The variability due to the error component $\epsilon$ can be reduced if a moving average with coefficient such that $\sum_{i=-m}^{m} \theta_i^2 < 1$ are used.

To estimate the trend component in an observed time series $x_t$ using the moving averages (in moving averages language we could say that we want to emphasize the trend component), one must take into account the issues presented above.

So, if $x_t$ is a monthly time series, one needs an average over 12 subsequent values to produce the series $x^*_t$ that is not influenced by the seasonal component (hence leaving only the trend component!). Thus, a moving average of 13 terms is a good option to estimate trend (trend-cycle in case there is one) and a moving average of 5 terms is good for quarterly data.

The moving averages are used also to estimate the seasonal component with the further aim, in general, of removing seasonality from the series in order to produce what is called a deseasonalized time series. The latter is extremely informative, especially where it is of interest a comparison of the level taken by the phenomenon in subsequent months (for example to evaluate the real dynamic) without this judgment being spuriously affected by the seasonality.

Estimating and removing seasonality are a rather interesting issue and a deeper discussion is of worth. Here we do not focus on this (for brevity reasons), but other details can be found in time series manuals (e.g. Chatfield, 1989).

### 6.3 Moving averages as bandpass filters

Moving averages can be seen as filters, more precisely they are bandpass filters. The idea is to fit a polynomial curve not on the whole series, but only to different parts. Put it differently, when a trend is estimated and other
fluctuations are smoothed, this is like filtering out the desired component and exclude the non-desired components.
Whenever a symmetric filter is applied, there is likely to be an end-effects problem. In some situations this is not necessarily a problem, in some other it is and it is then important to get smoothed values up to \( t = N \). To do this an option (only a possible one...) for the analyst is to use asymmetric filters which only involve present and past values of \( x_t \). For example a popular technique is the exponential smoothing that can be seen as a particular asymmetric filter:

\[
x^*_t = \sum_{j=0}^{\infty} \alpha(1 - \alpha)^j x_{t-j}
\]

where \( \alpha \) is a constant such that \( 0 < \alpha < 1 \). The weights applied to observations are \( \alpha(1 - \alpha)^j \) and decrease geometrically with \( j \).

How do we choose an appropriate filter? The answer to this question requires a considerable knowledge about the phenomenon and the properties of filters. For example, to get smoothed values we want to remove the high frequency variation and preserve the low frequency variation, in this case we want what is called lowpass filter. On the contrary, if we are interested in removing long term variations and focus on high frequency variations, we will use what is called highpass filter.