Time Series Concepts

Kurt M. Anstreicher

A *Time Series* is a sequence of observations of a single variable, $X_1, X_2, \ldots X_T$, where X_t is the observation in time period t. Commonly used time periods include hours, days, weeks, months, quarters and years. There is special terminology associated with time series, and also certain operations that are only applied to time series data.

1 Components of a Time Series

- **Trend.** The trend of a series refers to its overall behavior over time, ignoring up and down variation. The most common trends are level (no trend), linear, and exponential (growth or decay).
- **Seasonal.** The seasonal component of a series represents up and down variation with a known periodicity, usually corresponding to obvious, regular fluctuations. A typical example would be regular variation in the monthly or quarterly sales for a particular product corresponding to different times of the year.
- **Cyclical.** The cyclical component of a series represents up and down variation with a period that is not associated with obvious factors, and that may change over time. Time series of macroeconomic data often contain a nontrivial cyclical component.
- **Random:** The random component of a series represents irregular variation that is typically not associated with a known source.

2 Operations on a Time Series

2.1 Moving Average

A k-period simple, or trailing moving average of an original series $\{X_t\}_{t=1}^T$ is defined as

$$M_t = \frac{X_t + X_{t-1} + \ldots + X_{t-(k-1)}}{k}.$$

So the value of M_t is simply the average of X_t and the $(k-1)$ previous values in the series. Note that the first $k-1$ values of the moving average series are "lost," and the first observation of M_t is for $t = k$. The main effects of a moving average are to:

1. Remove k-period seasonality;
2. Damp, or reduce, the cyclical and random components.

It is also possible to define a *centered moving average*, where past and future values are used. If k is an odd number, $k = 2p + 1$ then a k-period centered moving average is defined as

$$C_t = \frac{X_t + (X_{t-1} + \ldots + X_{t-p}) + (X_{t+1} + \ldots + X_{t+p})}{2p+1}.$$

So each value C_t is the average of X_t and the past p and future p values of the series. Note that for the series C_t the first p and last p observations are lost. If a series has non-level trend, then the values of a centered moving average are more representative of the original series than those of a simple moving average, since a simple moving average will lag (be behind) the original series. However a centered moving average cannot be used for forecasting, since future observations are already required to define the last p values of the series C_t. It is worth noting that the centered moving average and simple moving average are related by

$$C_t = M_{t+p} \ .$$

When k is even (for example $k = 12$, as would be natural for monthly data), the definition of a centered moving average is slightly more complicated because there is no one time period exactly in the middle of k consecutive values of the series. The simplest definition of the centered moving average for $k = 2p$ is in terms of the simple moving average,

$$C_t = \frac{M_{t+(p-1)} + M_{t+p}}{2} \ .$$

2.2 Lag

The one-period lag of a series is simply the previous observation,

$$\mathrm{Lag} X_t = X_{t-1}.$$

Note that the first observation of $\mathrm{Lag}X_t$ is for $t = 2$. The lag of a series is often used in regression models. It is also possible to consider the k-period lag,

$$\mathrm{Lag}_k X_t = X_{t-k}.$$

The k-period lag is often more appropriate than the one-period lag if a series has seasonality with period k.

2.3 Difference

The one-period difference of a series is the change from the previous time period,

$$\mathrm{Diff} X_t = X_t - X_{t-1}.$$

If a series has a linear trend then the difference of the series will be level (no trend). In addition the difference will damp the cyclical and random components. For a series with k-period seasonality it may be more appropriate to consider the k-period difference,

$$\mathrm{Diff}_k X_t = X_t - X_{t-k}.$$

3 Autocorrelation

The term *autocorrelation* refers to correlation between a series and its lag. More generally, one can consider the period-k autocorrelation as the correlation between a series and its k-period lag. Autocorrelation of residuals is a common problem in regressions involving time series data. In Excel, autocorrelation can be easily computed using the `correl()` cell function. For example, if a series is in cells A2:A51, then the autocorrelation can be obtained using the syntax

$$= \texttt{correl}(\texttt{A2} : \texttt{A50}, \texttt{A3} : \texttt{A51})\ .$$

In the context of residuals in a time series regression a reasonable goal is to have the autocorrelation be below .3 in absolute magnitude (positive autocorrelation is much more common than negative autocorrelation). The significance of autocorrelation is usually evaluated using a measure known as the *Durbin-Watson statistic*. The exact definition is not important, but it is useful to note that if r is the autocorrelation of the residuals, and DW is the Durbin-Watson statistic, then

$$\text{DW} \approx 2 - 2r.$$

Thus DW ranges between 0 and 4, DW=2 corresponds to no autocorrelation, and values of DW below 2 correspond to positive autocorrelation.