

Stat 328 Handout on Time Series Analysis (Vardeman's Version of Chapter 13)

Time series data are (cleverly enough) data collected over time. For our present purposes we will consider ones collected at regular time intervals (weekly, monthly, quarterly, yearly, etc.). We'll use the notation

$$y_t = \text{observed value at period } t$$

The objectives of time series analysis are 1) the understanding of trends and seasonal effects in a record and (usually) 2) short term forecasting based on a record "to date." Typically one tries to measure on a "y" scale such that "local variation" in y_t is approximately constant and the general trend is simple (e.g. straight-line). This sometimes means that it is necessary to transform initial observations by taking a logarithm or square root, etc.

Overall Trend Assessment

The first order of business in a time series analysis is usually evaluation of any overall "trend" in the data. The most flexible and "model free" ways of doing this are through the use of various "data smoothing" algorithms.

A very simple method of smoothing is through the use of "moving averages." For a "span" of k periods, an **ordinary moving average** at time t is

$$\tilde{y}_t = \frac{y_t + y_{t-1} + y_{t-2} + \cdots + y_{t-k+1}}{k}$$

(This is like the text's formula on page 13-28, but there the authors are already thinking of using \tilde{y}_t to predict/forecast y_{t+1} .) So called **exponentially weighted moving averages** are computed by successively using (for some weight $0 < w < 1$)

$$\tilde{y}_t = wy_t + (1-w)\tilde{y}_{t-1}$$

(This is like the text's formula on page 13-31, but there the authors are already thinking of using \tilde{y}_t to predict/forecast y_{t+1} .)

Other more flexible smoothing methods have been developed and implemented in standard statistical software. JMP gives you access to "spline" smoothers. These have a so called "stiffness parameter" that controls how much they are allowed to "wiggle" in order to follow the local jaggedness of a time series record. The larger the stiffness, the less wiggling that is allowed.

A second way of smoothing a time series record is through the fitting (using regression analysis) of some function of t to the record. For example, one might fit a linear (in time) trend

$$\hat{y}_t = b_0 + b_1t$$

using least squares regression with the "predictor" variable t . Other common possibilities (easily fit using JMP) include low order polynomials in t and exponential functions of t (that is, functions like $\hat{y}_t = c_0 10^{c_1 t}$). One possible advantage of using such functions to represent trends, is that they can provide forecasts extrapolating an apparent current

trend. (In contrast, the smoothing ideas above can only forecast that the next value of a time series is the current smoothed value.)

Accounting for Seasonality

Many economic time series have regular seasonal patterns that are ignored or smoothed out when one assesses an overall trend. It is essential to account for/model these if one is going to do practical time series analysis and forecasting. There are two standard ways of doing so, that treat seasonal effects as respectively “additive” and “multiplicative.”

Additive treatment of seasonal effects can be done using dummy variables in a regression analysis, as discussed on pages 13-10 through 13-12 of the text. If there are I periods in a cycle, one makes up $I - 1$ “season” dummy variables and either includes them in a regression of y_t on the function(s) of t used to smooth the data, or uses them alone on a set of residuals

$$e_t = y_t - \hat{y}_t \text{ or } e_t = y_t - \tilde{y}_t$$

One’s forecast at a given period in the future then becomes a forecast for the smooth part of the data pattern **plus** the estimated seasonal effect for the period,

$$\text{forecast} = \widehat{\text{smooth}} + \widehat{\text{seasonal}}$$

Multiplicative treatment of seasonal effects is done by first dividing each observed y_t by a corresponding smoothed or fitted value. This gives an estimate of a multiplicative seasonal effect

$$\frac{y_t}{\tilde{y}_t} \text{ or } \frac{y_t}{\hat{y}_t}$$

One can then average these estimates over all periods that come from the same part of a cycle (for example, for all Januaries in a time series data set of monthly data with $I = 12$ months in a cycle) to get a multiplicative factor for that season. One’s forecast at a given period in the future then becomes a forecast for the smooth part of the data pattern **times** the estimated seasonal effect for the period,

$$\text{forecast} = \widehat{\text{smooth}} \times \widehat{\text{seasonal}}$$

This is discussed on pages 13-12 through 13-16 of the text.

When a government agency announces the latest value of an economic or social time series (for example, housing starts, retail sales, unemployment rate, etc.) it often announces both the raw and a “seasonally adjusted” version of the value. That is, both

$$y_t \text{ and } \frac{y_t}{\text{seasonal factor}}$$

are announced (where the “*seasonal factor*” corresponding to y_t presumably comes from analysis of historical data).

Exploiting Serial Correlation

“Trend” and Seasonal Effects” can describe much of what is seen in business time series. But sometimes, even more careful modeling/analysis can take advantage of additional structure in the data, related to “serial-” or “auto-” correlations.

That is, when time series data have trends and seasonal effects, successive observations will empirically be correlated (in the sense that the size of y_t will tend to be related to the size of y_{t-1}). But this can also happen even when no trends or seasonal effects are obvious. And (even after “detrrending” and “removing seasonal effects”) residuals

$$e_t = y_t - \left(\widehat{smooth}_t + \widehat{seasonal}_t \right) \text{ or } e_t = y_t - \left(\frac{\widehat{smooth}_t}{\widehat{seasonal}_t} \right)$$

for many time series have empirical structure that can be exploited to improve forecasts.

Define the ***k*th order lag** of a time series y_t to be y_{t-k} . (The ***k*th order lag** of a residual series e_t is, of course, e_{t-k} .) The ***k*th order (sample) autocorrelation** for the y_t series is then the sample correlation between the values y_t and y_{t-k} . (And the corresponding terminology holds for the residual series.) Sample autocorrelations can be computed and plots of y_t vs y_{t-k} (or e_t vs e_{t-k}) made using the “multivariate” function in JMP (after using the “lag” function in the JMP calculator to create lagged versions of the series).

Where autocorrelations are important, one way of exploiting them is through so called “autoregressions.” As the name suggests, this is regression of a series (original or residual) onto itself (onto lags for itself). That is, if empirical investigation suggests that there are important patterns in residuals over time, one might fit an equation like

$$\hat{e}_t = b_0 + b_1 e_{t-1} \tag{*}$$

to the residuals in hand, and use b_0 and b_1 to predict next period’s residual from the current one. This can be added to a prediction based on smooth plus seasonal effects, to produce a more sophisticated prediction. Of course, once one thinks in these terms, the possibility of using not just 1st order lags, but higher order lags as well comes into play. That is, one could go beyond (*) to

$$\hat{e}_t = b_0 + b_1 e_{t-1} + b_2 e_{t-2}$$

(or even further).