Statistics 153 (Time Series) : Lecture Three

Aditya Guntuboyina

24 January 2012

1 Plan

1. In the last class, we looked at two ways of dealing with trend in time series data sets - Fitting parametric curves and smoothing.
2. Today, we will finish the story on trend by looking at Filtering (Section 2.5.2) and Differencing (Section 2.5.3).
3. A very common feature of weekly/monthly/quarterly time series data is seasonality. Seasonality is dealt with in the same way as trend (Section 2.6):
   (a) Fitting parametric seasonal functions (sines and cosines).
   (b) Smoothing.
   (c) Differencing.

2 More General Filtering for Trend Estimation

The smoothing method described in the last class (19 Jan) for estimating the trend function m_t is a special case of *linear filtering*. A linear filter converts the observed time series X_t into an estimate of the trend $\hat{m}_t$ via the linear operation:

$$\hat{m}_t = \sum_{j=-q}^{s} a_j X_{t+j}.$$

The numbers $a_{-q}, a_{-q+1}, \dots, a_{-1}, a_0, a_1, \dots, a_s$ are called the weights of the filter. The Smoothing method is clearly a special instance of filtering with $s = q$ and $a_j = 1/(2q+1)$ for $|j| \le q$ and 0 otherwise.

One can think of the filter as a (linear) *system* which takes the observed series X_t as input and produces the estimate of trend, $\hat{m}_t$ as output.

In addition to the choice $a_j = 1/(2q+1)$ for $|j| \leq q$, there are other choice of filters that people commonly use.

(1) **Binomial Weights**: Based on the following idea. When we are estimating the value of the trend m_t at t, it makes sense to give a higher weight to X_t compared to $X_{t\pm1}$ and a higher weight to $X_{t\pm1}$ compared to $X_{t\pm2}$ and so on. An example of such weights are:

$$a_j = 2^{-q}\binom{q}{q/2+j} \qquad \text{for } j = -q/2, -q/2+1, \dots, -1, 0, 1, \dots, q/2.$$

As in usual smoothing, choice of q is an issue here.

(2) **Spencer's 15 point moving average**: We have seen that simple moving average filter leaves linear functions untouched. Is it possible to design a filter which leaves higher order polynomials untouched? For example, can we come up with a filter which leaves all quadratic polynomials untouched. Yes!

For a filter with weights a_j to leave all quadratic polynomials untouched, we need the following to be satisfied for every quadratic polynomial m_t:

$$\sum_j a_j m_{t+j} = m_t \qquad \text{for all } t$$

In other words, if $m_t = \alpha t^2 + \beta t + \gamma$, we need

$$\sum_j a_j \left(\alpha(t+j)^2 + \beta(t+j) + \gamma\right) = \alpha t^2 + \beta t + \gamma \qquad \text{for all } t.$$

Simplify to get

$$\alpha t^2 + \beta t + \gamma = (\alpha t^2 + \beta t + \gamma)\sum_j a_j + (2\alpha t + \beta)\sum_j j a_j + \alpha \sum_j j^2 a_j \qquad \text{for all } t.$$

This will clearly be satisfied if

$$\sum_j a_j = 1 \qquad \sum_j j a_j = 0 \qquad \sum_j j^2 a_j = 0. \tag{1}$$

An example of such a filter is Spencer's 15 point moving average defined by

$$a_0 = \frac{74}{320}, a_1 = \frac{67}{320}, a_2 = \frac{46}{320}, a_3 = \frac{21}{320}, a_4 = \frac{3}{320}, a_5 = \frac{-5}{320}, a_6 = \frac{-6}{320}, a_7 = \frac{-3}{320}$$

and $a_j = 0$ for $j > 7$. Also the filter is symmetric in the sense that $a_{-1} = a_1, a_{-2} = a_2$ and so on. Check that this filter satisfies the condition (1).

Because this is a symmetric filter, it can be checked that it allows all cubic polynomials to pass unscathed as well.

(3) **Exponential Smoothing**: Quite a popular method of smoothing (wikipedia has a big page on this). It is also used as a forecasting technique.

To obtain $\hat{m}_t$ in this method, one uses only the *previous* observations $X_{t-1}, X_{t-2}, X_{t-3}, \ldots$. The weights assigned to these observations *exponentially decrease* the further one goes back in time. Specifically,

$$\hat{m}_t := \alpha X_{t-1} + \alpha(1-\alpha)X_{t-2} + \alpha(1-\alpha)^2 X_{t-3} + \cdots + \alpha(1-\alpha)^{t-2} X_1 + (1-\alpha)^{t-1} X_0.$$

Check that the weights add up to 1. α is a parameter that determines the amount of smoothing (α here is analogous to q in smoothing). If α is close to 1, there is very little smoothing and vice versa.

3 Differencing for Trend Elimination

The residuals obtained after fitting the trend function m_t in the model $X_t = m_t + W_t$ are studied to see if they are purely random or have some dependence structure that can be exploited for prediction.

Differencing is a much simpler technique which produces such de-trended residuals.

One just looks at $Y_t = X_t - X_{t-1}, t = 2, \ldots, n$. If the trend m_t in $X_t = m_t + W_t$ is linear, then this operation simply removes it because if $m_t = \alpha t + b$, then $m_t - m_{t-1} = \alpha$ so that $Y_t = \alpha + W_t - W_{t-1}$.

Suppose that the first differenced series Y_t appears purely random. What then would be a reasonable forecast for the original series: X_{n+1}? Because Y_t is purely random, we forecast Y_{n+1} by the sample mean $\bar{Y} := (Y_2 + \cdots + Y_n)/(n-1)$. But since $Y_{n+1} = X_{n+1} - X_n$, this results in the forecast $X_n + \bar{Y}$ for X_{n+1}.

Sometimes, even after differencing, one can notice a trend in the data. In that case, just difference again. It is useful to follow the notation ∇ for differencing:

$$\nabla X_t = X_t - X_{t-1} \qquad \text{for } t = 2, \ldots, n$$

and second differencing corresponds to

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2} \qquad \text{for } t = 3, \ldots, n.$$

It can be shown that quadratic trends simply disappear with the operation ∇^2. Suppose the data $\nabla^2 X_t$ appear purely random, how would you obtain a forecast for X_{n+1}?

Differencing is a quick and easy way to produce detrended residuals and is a key component in the ARIMA forecasting models (later). A problem however is that it does not result in any estimate for the trend function m_t.

4 Seasonality

In this section, we shall discuss fitting models of the form $X_t = s_t + W_t$ to the data where s_t is a periodic function of a known period d i.e., $s_{t+d} = s_t$ for all t. Such a function s models seasonality. These models are appropriate to monthly or quarterly data sets that have a seasonal pattern to them.

This model, however, will not be applicable for datasets having both trend and seasonality which is the more realistic situation. These will be focussed a little later.

Just like the trend case, there are three different approaches to dealing with seasonality: fitting parametric functions, smoothing and differencing.

4.1 Fitting a parametric seasonality function

The simplest periodic functions of period d are: $a\cos(2\pi ft/d)$ and $a\sin(2\pi ft/d)$. Here f is a positive integer that is called *frequency*. The higher f is, the more rapid the oscillations in the function are. More generally,

$$s_t = a_0 + \sum_{f=1}^{k} \left(a_j \cos(2\pi ft/d) + b_j \sin(2\pi ft/d)\right)$$

is a periodic function.

Choose a value of k (not too large) and fit this to the data.

4.2 Smoothing

Because of periodicity, the function s_t only depends on the d values $s_1, s_2, \dots, s_d$. Clearly s_1 can be estimated by the average of $X_1, X_{1+d}, X_{1+2d}, \dots$. For example, for monthly data, this corresponds to estimating the mean term for January by averaging all January observations. Thus

$$\hat{s}_i := \text{average of } X_i, X_{i+d}, X_{i+2d}, \dots$$

Note that here, we are fitting 12 parameters (one each for $s_1, \dots, s_d$) from n observations. If n is not that big, fitting 12 parameters might lead to overfitting.

4.3 Differencing

How can we obtain residuals adjusted for seasonality from the data without explicitly fitting a seasonality function? Recall that a function s is a periodic function of period d if $s_{t+d} = s_t$ for all t. The model that we have in mind here is: $X_t = s_t + W_t$.

Clearly $X_t - X_{t-d} = s_t - s_{t-d} + W_t - W_{t-d} = W_t - W_{t-d}$. Therefore, the lad-d differenced data $X_t - X_{t-d}$ do not display any seasonal trend. This method of producing deseasonalized residuals is called *Seasonal Differencing*.

5 Next Class

1. Dealing with both trend and seasonality (Section 2.6).
2. Transformations (Section 2.4)