Linear regression

Lets assume that some output or dependent time series, say, x_t, for $t = 1,...,n$ is being influenced by a collection of possible inputs or independent series, say, $z_{t_1}, z_{t_2}, \ldots, z_{t_q}$, where we first regard the inputs as fixed and known.

We express this relation through the linear regression model

$$x_t = \beta_0 + \beta_1 z_{t_1} + \beta_2 z_{t_2} + \cdots \beta_q z_{t_q} + w_t$$

Expressed in vectors form

$$x_t = \boldsymbol{\beta}^T \boldsymbol{z}_t + w_t$$

where $\beta = [\beta_0, \beta_1, \ldots, \beta_q]^T$ and $\boldsymbol{z}_t = [1, z_{t_1}, z_{t_2}, \ldots, z_{t_q}]$, and sequentially,

$$x_t - \boldsymbol{\beta}^T \boldsymbol{z}_t = w_t$$

Error measure

To calculate the error in our model, we usually use the squared error defined as:

$$Q(\boldsymbol{\beta}) = \sum_{t=1}^{N} w_t^2 = \sum_{t=1}^{N} (x_t - \boldsymbol{\beta}^T \boldsymbol{z}_t)^2$$

And we want to get

$$\operatorname{argmin}_{\boldsymbol{\beta}} Q(\boldsymbol{\beta})$$

Why?

Optimization step

To minimize, we usually compute the gradient and set to zero

$$\nabla_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) = \mathbf{0}$$

If we evaluate it, we will find that the solution must satisfy

$$\sum_{t=1}^{N} (x_t - \boldsymbol{\beta}^T \boldsymbol{z}_t)\boldsymbol{z}_t = \mathbf{0}$$

If $\sum_{t=1}^{N} \boldsymbol{z}_t^T \boldsymbol{z}_t$ is not singular (has inverse), the estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{N} \boldsymbol{z}_t^T \boldsymbol{z}_t \right)^{-1} \sum_{t=1}^{N} x_t \boldsymbol{z}_t$$

What is another appropiated method for optimize a regression parameters?

In general, it is necessary for time series data to be stationary so that averaging lagged products over time, as in the previous section, will be a sensible thing to do.

Hence, to achieve any meaningful statistical analysis of time series data, it will be crucial that, if nothing else, the mean and the autocovariance functions satisfy the conditions of stationarity (for at least some reasonable stretch of time). Often, this is not the case, and we will mention some methods in this section for playing down the effects of nonstationarity so the stationary properties of the series may be studied.

Trend stationarity

We may write a trend stationary model as

$$x_t = \mu_t + y_t$$

where x_t are the observations, μ_t denotes the trend, and y_t is a stationary process.

Quite often, strong trend will obscure the behavior of the stationary process, y_t. Hence, there is some advantage to removing the trend as a first step in an exploratory analysis of such time series.

What could be an appropiate model for detrend? Why?

Detrending using random walk with drift

We might model trend as a stochastic component using the random walk with drift model

$$\mu_t = \delta + \mu_{t-1} + w_t$$

If the appropriate model is $x_t = \mu_t + y_t$, then *differencing* the data, x_t, yields a stationary process;

$$x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = \delta + w_t + y_t - y_{t-1}$$

where $y_t - y_{t-1}$ is a stationary process.

Advantages and disadvantages of differencing

One advantage of differencing over detrending to remove trend is that no parameters are estimated in the differencing operation.

One disadvantage, however, is that differencing does not yield an estimate of the stationary process y_t.

If an estimate of y_t is essential, then detrending may be more appropriate. If the goal is to coerce the data to stationarity, then differencing may be more appropriate. Differencing is also a viable tool if the trend is fixed.

Because differencing plays a central role in time series analysis, it receives its own notation. The first difference is denoted as

$$\nabla x_t = x_t - x_{t-1}$$

The first difference eliminates a linear trend. A second difference can eliminate a quadratic trend, and so on.

Backshift operator

Definición 1. *We define the backshift operator (or lag operator $\boldsymbol{L}$) by*

$$\boldsymbol{B}\,x_t = x_{t-1}$$

and extend it to powers $\boldsymbol{B}^2 x_t = \boldsymbol{B}(\boldsymbol{B}\,x_t) = \boldsymbol{B}\,x_{t-1} = x_{t-2}$*, and so on. Thus,*

$$\boldsymbol{B}^k x_t = x_{t-k}$$

Forward-shift operator

The idea of an inverse operator can also be given if we require $\boldsymbol{B}^{-1}\boldsymbol{B} = 1$, so that

$$x_t = \boldsymbol{B}^{-1}\boldsymbol{B}\, x_t = \boldsymbol{B}^{-1}x_{t-1}$$

In addition, it is clear that we may rewrite

$$\nabla x_t = x_t - x_{t-1} = (1 - \boldsymbol{B})x_t$$

and we may extend the notion further. For example, the second difference becomes

$$\nabla^2 x_t = (1 - \boldsymbol{B})^2 x_t = (1 - 2\boldsymbol{B} + \boldsymbol{B}^2)x_t = x_t - 2x_{t-1} + x_{t-2}$$

Higher order differences

Definición 2. *Differences of order d are defined as*

$$\nabla^d = (1 - \boldsymbol{B})^d$$

where we may expand the operator $(1 - \boldsymbol{B})^d$ *algebraically to evaluate for higher integer values of* d*. When* $d = 1$*, we drop it from the notation.*

The first difference is an example of a linear filter applied to eliminate a trend. Other filters, formed by averaging values near x_t, can produce adjusted series that eliminate other kinds of unwanted fluctuations.

Smoothing in the time series context

This method is useful in discovering certain traits in a time series, such as long-term trend and seasonal components. In particular, if x_t represents the observations, then

$$m_t = \sum_{j=-k}^{k} a_j x_{t-j}$$

where $a_j = a_{-j} \geq 0$, and $\sum_{j=-k}^{k} a_j = 1$ is a symmetric moving average of the data.