LINEAR REGRESSION ANALYSIS

MODULE – V

Lecture - 21

Correcting Model Inadequacies Through Transformation and Weighting

Dr. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Analytical methods for selecting a transformation on study variable

The Box-Cox method

Suppose the normality and/or constant variance of study variable *y* can be corrected through a power transformation on *y*. This means y is to be transformed as y^{λ} where λ is the parameter to be determined. For example, if $\lambda = 0.5$, then the transformation is square root and $\sqrt{y}$ is used as study variable in place of *y*.

Now the linear regression model has parameters β, σ^2 and λ. Box and Cox method tells how to estimate simultaneously the λ and parameters of the model using the method of maximum likelihood.

Note that as λ approaches zero, y^{λ} approaches to 1. So there is a problem at $\lambda = 0$ because this makes all the observation *y* to be unity. It is meaningless that all the observation on study variable are constant. So there is a discontinuity at $\lambda = 0$. One approach to solve this difficulty is to use $\dfrac{y^{\lambda}-1}{\lambda}$ as a study variable.

Note that as $\lambda \to 0, \dfrac{y^{\lambda}-1}{\lambda} \to \ln y$. So a possible solution is to use the transformed study variable as

$$W = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & \text{for } \lambda \neq 0 \\ \ln y & \text{for } \lambda = 0. \end{cases}$$

So family W is continuous. Still it has a drawback. As λ changes, the value of W change dramatically. So it is difficult to obtain the best value of λ. If different analyst obtain different values of λ, then it will fit different models. It may then not be appropriate to compare the models with different values of λ. So it is preferable to use an alternative form

$$y^{(\lambda)} = V = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda y_*^{\lambda-1}} & \text{for } \lambda \neq 0 \\ y_* \ln y & \text{for } \lambda = 0 \end{cases}$$

where y_* is the geometric mean of y_i 's as $y_* = (y_1 y_2 \ldots y_n)^{1/n}$ which is constant.

For calculation purpose, we can use $\ln y_* = \dfrac{1}{n}\sum_{i=1}^{n} \ln y_i$.

When V is applied to each y_i, we get $V = \left(V_1, V_2, \ldots, V_n\right)'$ as a vector of observation on transformed study variable and we use it to fit a linear model $V = X\beta + \varepsilon$ using least squares or maximum likelihood method.

The quantity $\lambda y_*^{\lambda-1}$ in the denominator is related to the n^{th} power of Jacobian of transformation. See how:

We want to convert y_i into $y_i^{(\lambda)}$ as

$$y_i^{(\lambda)} = W_i = \frac{y_i^{\lambda} - 1}{\lambda}; \quad \lambda \neq 0.$$

Let

$$y = \left(y_1, y_2, ..., y_n\right)', \quad W = (W_1, W_2, ..., W_n)'.$$

Note that if $W_1 = \dfrac{y_1^{\lambda} - 1}{\lambda}$, then

$$\frac{\partial W_1}{\partial y_1} = \frac{\lambda y_1^{\lambda-1}}{\lambda} = y_1^{\lambda-1}$$

$$\frac{\partial W_1}{\partial y_2} = 0.$$

In general,

$$\frac{\partial W_i}{\partial y_j} = \begin{cases} y_i^{\lambda-1} & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

The Jacobian of transformation is given by

$$J(y_i \rightarrow W_i) = \frac{\partial y_i}{\partial W_i} = \frac{1}{\left(\dfrac{\partial W_i}{\partial y_i}\right)} = \frac{1}{y_i^{\lambda-1}}.$$

$$J(W \to y) = \begin{vmatrix} \frac{\partial W_1}{\partial y_1} & \frac{\partial W_1}{\partial y_2} & \cdots & \frac{\partial W_1}{\partial y_n} \\ \frac{\partial W_2}{\partial y_1} & \frac{\partial W_2}{\partial y_2} & \cdots & \frac{\partial W_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial W_n}{\partial y_1} & \frac{\partial W_n}{\partial y_2} & \cdots & \frac{\partial W_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} y_1^{\lambda-1} & 0 & 0 & \cdots & 0 \\ 0 & y_2^{\lambda-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & y_n^{\lambda-1} \end{vmatrix}$$

$$= \prod_{i=1}^{n} y_i^{\lambda-1}$$

$$= \left(\prod_{i=1}^{n} y_i \right)^{\lambda-1}$$

$$J(y \to W) = \frac{1}{J(W \to Y)} = \left(1 \Big/ \prod_{i=1}^{n} y_i \right)^{\lambda-1}.$$

Since this is a Jacobian when we want to transform the whole vector *y* to whole vector *W*. If an individual y_i is to be transform into W_i, then take its geometric mean as

$$J(y_i \to W_i) = \left(1 \Big/ \left(\prod_{i=1}^{n} y_i \right)^{\frac{1}{n}} \right)^{\lambda-1}.$$

The quantity

$$J(Y \to W) = 1 \Big/ \prod_{i=1}^{n} y_i^{\lambda-1}$$

ensures that unit volume is preserved moving from the set of y_i to the set of V_i. This is a factor which scales and ensures that the residual sum of squares obtained from different values of λ can be compared.

To find the appropriate family, consider

$$y^{(\lambda)} = V = X\beta + \varepsilon$$

where

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda y_*^{\lambda-1}}, \ \varepsilon \sim N(0, \sigma^2 I).$$

Applying method of maximum likelihood for likelihood function for $y^{(\lambda)}$,

$$L\left[y^{(\lambda)}\right] = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n} \varepsilon_i^2}{2\sigma^2}\right]$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{\varepsilon'\varepsilon}{2\sigma^2}\right]$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{(y^{(\lambda)} - X\beta)'(y^{(\lambda)} - X\beta)}{2\sigma^2}\right]$$

$$\ln L\left[y^{(\lambda)}\right] = -\frac{n}{2}\ln\sigma^2 - \left[\frac{(y^{(\lambda)} - X\beta)'(y^{(\lambda)} - X\beta)}{2\sigma^2}\right]$$ (ignoring constant).

Solving

$$\frac{\partial \ln L\left[y^{(\lambda)}\right]}{\partial \beta} = 0$$

$$\frac{\partial \ln L\left[y^{(\lambda)}\right]}{\partial \sigma^2} = 0$$

gives the maximum likelihood estimators

$$\hat{\beta}(\lambda) = (X'X)^{-1}X'y^{(\lambda)}$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{n} y^{(\lambda)}{}'\left[I - X(X'X)^{-1}X'\right]y^{(\lambda)} = \frac{y^{(\lambda)}{}'\bar{H}y^{(\lambda)}}{n}$$

for a given value of λ .

Substituting these estimates in the log likelihood function $\ln L\left[y^{(\lambda)}\right]$ gives

$$L(\lambda) = -\frac{n}{2}\ln\hat{\sigma}^2 = -\frac{n}{2}\ln\left[SS_{res}(\lambda)\right]$$

where $SS_{res}(\lambda)$ is the sum of squares due to residuals which is a function of λ . Now maximize $L(\lambda)$ with respect to λ. It is difficult to obtain any closed form of the estimator of λ. So we maximize it numerically.

The function $-\frac{n}{2}\ln\left[SS_{res}(\lambda)\right]$ is called as the **Box-Cox objective function.**

Let $\lambda_{\max}$ be the value of λ which maximizes the Box-Cox objective function. Then under fairly general conditions, for any other λ

$$n \ln\left[SS_{res}(\lambda)\right] - n \ln\left[SS_{res}(\lambda_{\max})\right]$$

has approximately $\chi^2(1)$ distribution. This result is based on the large sample behaviour of the likelihood ratio statistic. This is explained as follows:

The likelihood ratio test statistic in our case is

$$\eta_n \equiv \eta = \frac{\underset{\Omega_o}{Max}\, L}{\underset{\Omega}{Max}\, L}$$

$$= \frac{\underset{\Omega_o}{Max}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}}}{\underset{\Omega}{Max}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}}}$$

$$= \frac{\left(\frac{1}{\hat{\sigma}^2(\lambda)}\right)^{\frac{n}{2}}}{\left(\frac{1}{\hat{\sigma}^2(\lambda_{\max})}\right)^{\frac{n}{2}}}$$

$$= \left(\frac{1/SS_{res}(\lambda)}{1/SS_{res}(\lambda_{\max})}\right)^{\frac{n}{2}}$$

$$\ln\eta = \frac{n}{2}\ln\left(\frac{SS_{res}(\lambda_{\max})}{SS_{res}(\lambda)}\right)$$

$$-\ln\eta = \frac{n}{2}\ln\left(\frac{SS_{res}(\lambda)}{SS_{res}(\lambda_{\max})}\right)$$

$$= \frac{n}{2}\ln\left[SS_{res}(\lambda)\right] - \frac{n}{2}\ln\left[SS_{res}(\lambda_{\max})\right]$$

$$= -L(\lambda) + L(\lambda_{\max})$$

where

$$L(\lambda) = -\frac{n}{2}\ln\left[SS_{res}(\lambda)\right]$$

$$L(\lambda_{\max}) = -\frac{n}{2}\ln\left[SS_{res}(\lambda_{\max})\right].$$

Since under certain regularity conditions, $-2\ln\eta_n$ converges in distribution to $\chi^2(1)$ when the null hypothesis is true, so

$$-2\ln\eta \sim \chi^2(1)$$

or $-\ln\eta \sim \dfrac{\chi^2(1)}{2}$

or $L(\lambda_{\max}) - L(\lambda) \sim \dfrac{\chi^2(1)}{2}.$