

LINEAR REGRESSION ANALYSIS

MODULE – V

Lecture - 22

Correcting Model Inadequacies Through Transformation and Weighting

Dr. Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Computational procedure

The maximum-likelihood estimate of λ corresponds to the value of λ for which residual sum of squares from the fitted model $SS_{res}(\lambda)$ is a minimum. To determine such λ , we proceed computationally as follows:

- Fit $y^{(\lambda)}$ for various values of λ . For example, start with values in $(-1, 1)$ then take the values in $(-2, 2)$ and so on. Take about 15 to 20 values of λ which are expected to be sufficient for the estimation of optimum value.
- Plot $SS_{res}(\lambda)$ versus λ .
- Find the value of λ which minimizes $SS_{res}(\lambda)$ from the graph.
- A second iteration can be performed using a finer mesh of values, if desired.

Note that the value of λ can not be selected by directly comparing the residual sum of squares from the regression of y^λ on x because for each λ , the residual sum of squares is measured on a different scale.

It is better to use simple values of λ . For example, the practical difference between $\lambda = 0.5$ and $\lambda = 0.58$ is likely to be small but $\lambda = 0.5$ is much easier to interpret.

Once λ is selected, then use

- y^λ as a study variable if $\lambda \neq 0$.
- $\ln y$ as a study variable if $\lambda = 0$.

It is entirely acceptable to use $y^{(\lambda)}$ as response for final model. This model will have a scale difference and an origin shift in comparison to model using y^λ (or $\ln y$) as the response.

An approximate confidence interval for λ

We can find an approximate confidence interval for the transformation parameter λ . This interval helps in selecting the final value of λ . For example, if $\hat{\lambda} = 0.58$ is the value of λ which is minimizing the sum of squares due to residual. But if $\lambda = 0.5$ is in the confidence interval, then one may use the square root transformation because it is easier to explain. Furthermore if $\lambda = 1$ is in the confidence interval, then it may be concluded that no transformation is necessary.

In applying the method of maximum likelihood to the regression model, we are essentially maximizing

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

or equivalently, we are minimizing $SS_{res}(\lambda)$.

An approximate $100(1-\alpha\%)$ confidence interval for λ consists of those values of λ that satisfy

$$L(\hat{\lambda}) - L(\lambda) \leq \frac{\chi^2_{\alpha}(1)}{2}$$

where $\chi^2_{\alpha}(1)$ is the upper $100\alpha\%$ point of the Chi-square distribution with one degree of freedom.

The approximate confidence interval is constructed using the following steps:

- Draw a plot of $L(\lambda)$ versus λ .
- Draw a horizontal line at height

$$L(\hat{\lambda}) - \frac{\chi_{\alpha}^2(1)}{2}$$

on the vertical scale.

- This line would cut the $L(\lambda)$ at two points.
- The location of these two points on the λ -axis defines the two end points of the approximate confidence interval.
- If sum of squares due to residuals is minimized and $SS_{res}(\lambda)$ versus λ is plotted, then the line must be plotted at the height

$$SS^* = SS_{res}(\hat{\lambda}) \exp \left[\frac{\chi_{\alpha}^2(1)}{n} \right]$$

where $\hat{\lambda}$ is the value of λ which minimizes the sum of squares due to residuals. See how:

$$\begin{aligned} L(\hat{\lambda}) - \frac{\chi_{\alpha}^2(1)}{2} &= -\frac{n}{2} \ln [SS_{res}(\hat{\lambda})] - \frac{\chi_{\alpha}^2(1)}{2} \\ &= -\frac{n}{2} \left[\ln \{SS_{res}(\hat{\lambda})\} + \frac{\chi_{\alpha}^2(1)}{n} \right] \\ &= -\frac{n}{2} \left[\ln \{SS_{res}(\hat{\lambda})\} + \ln \left\{ \exp \left(\frac{\chi_{\alpha}^2(1)}{n} \right) \right\} \right] \\ &= -\frac{n}{2} \left[\ln \left\{ SS_{res}(\hat{\lambda}) \cdot \exp \left(\frac{\chi_{\alpha}^2(1)}{n} \right) \right\} \right] \\ &= -\frac{n}{2} \ln SS^*. \end{aligned}$$

Using the expansion of exponential function as

$$\exp(t) = 1 + t + \frac{t^2}{2!} + \dots$$

$$\approx 1 + t,$$

we can approximate and replace $\exp\left[\frac{\chi^2_\alpha(1)}{n}\right]$ by $1 + \frac{\chi^2_\alpha(1)}{n}$. So in place of $\exp\left[\frac{\chi^2_\alpha(1)}{n}\right]$ in applying the confidence interval procedure, we can use the following:

$$1 + \frac{Z_{\alpha/2}^2}{\gamma} \quad \left(\text{or } 1 + \frac{Z_{\alpha/2}^2}{n} \right)$$

$$\text{or } 1 + \frac{t_{\alpha/2}^2}{\gamma} \quad \left(\text{or } 1 + \frac{t_{\alpha/2}^2}{n} \right)$$

$$\text{or } 1 + \frac{\chi_{\alpha/2}^2}{\gamma} \quad \left(\text{or } 1 + \frac{\chi_{\alpha/2}^2}{n} \right)$$

where γ is the degrees of freedom associated with sum of squares due to residuals.

These expressions are based on the fact that

$$\chi^2(1) = Z^2 \approx t_\gamma^2 \quad \text{if } \gamma \text{ is small.}$$

It is debatable to use either γ or n but practically the difference is very little between the confidence interval results.

Box-Cox transformation was originally introduced to reduce the nonnormality in the data. It also helps in reducing the nonlinearity. The approach is to find out the transformations which attempt to reduce the residuals associated with outliers and also reduce the problem of non constant error variance if there was no acute nonlinearity to begin with.

Transformation on explanatory variables: Box and Tidwell procedure

Suppose the relationship between y and one or more of the explanatory variables is nonlinear. Other usual assumptions normally and independently distributed study variable with constant variance are at least approximately satisfied.

We want to select an appropriate transformation on the explanatory variable so that the relationship between y and transformed explanatory variable is as simple as possible.

Box and Tidwell procedure describes a general analytical procedure for determining the form of transformation on x

Suppose that the study variable y is related to the power of explanatory variables. Box and Tidwell procedures for explanatory variables chooses the variables as

$$z_{ij} = \begin{cases} \frac{x_{ij}^{\alpha_j} - 1}{\alpha_j} & \text{when } \alpha_j \neq 0, i = 1, 2, \dots, n; j = 1, 2, \dots, k \\ \ln x_{ij} & \text{when } \alpha_j = 0. \end{cases}$$

We need to estimate α_j 's. Since the dependent variable is not being transformed, we need not worry about the changes of scale and minimize

$$\sum_{i=1}^n [y_i - \beta_0 - \beta_1 z_{i1} - \dots - \beta_k z_{ik}]^2$$

by using nonlinear least squares techniques.

We consider this for simple linear regression model instead of nonlinear regression model.

Assume y is related to $\xi = x^\alpha$ as

$$E(y) = f(\xi, \beta_0, \beta_1) = \beta_0 + \beta_1 \xi$$

where $\xi = \begin{cases} x^\alpha & \text{if } \alpha \neq 0 \\ \ln x & \text{if } \alpha = 0 \end{cases}$

where β_0, β_1 and α are the unknown parameters.

Suppose α_0 is the initial guess of constant α .

Usually, first guess is $\alpha_0 = 1$ so that $\xi = x$ or no transformation is applied in the first iteration.

Expand about the initial guess in a Taylor series and ignoring terms of order higher than one gives

$$\begin{aligned} E(y) &= f(\xi_0, \beta_0, \beta_1) + (\alpha - \alpha_0) \left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}} \\ &= \beta_0 + \beta_1 x + (\alpha - 1) \left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}}. \end{aligned}$$

Suppose the term $\left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}}$ is known, then it can be treated just like as an additional explanatory variable. Then

the parameters β_0, β_1 and α can be estimated by least squares method.

The estimate of α can be considered as an improved estimate of the transformation parameter.

This term can be written as

$$\left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}} = \left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\xi} \right\}_{\xi=\xi_0} \left\{ \frac{d\xi}{d\alpha} \right\}_{\alpha=\alpha_0}.$$

Since the form of transformation is known, i.e., $\xi = x^\alpha$,

so

$$\frac{d\xi}{d\alpha} = x \ln x.$$

Furthermore

$$\left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\xi} \right\}_{\xi=\xi_0} = \frac{d(\beta_0 + \beta_1 x)}{dx} = \beta_1.$$

So β_1 can be estimated by fitting the model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ by least squares method.

Then an “adjustment” to initial guess $\alpha_0 = 1$ is computed by defining a second regression variable as

$$\omega = x \ln x$$

estimating the parameter in

$$\begin{aligned} E(y) &= \beta_0^* + \beta_1^* x + (\alpha - 1)\beta_1 \omega \\ &= \beta_0^* + \beta_1^* x + \gamma \omega \end{aligned}$$

by least squares.

This gives

$$\hat{y} = \hat{\beta}_0^* + \hat{\beta}_1^* x + \hat{\gamma} \omega$$

$$\hat{\gamma} = (\alpha - 1) \hat{\beta}_1$$

or

$$\alpha_1 = \frac{\hat{\gamma}}{\hat{\beta}_1} + 1$$

as the revised estimate of α .

Note that $\hat{\beta}_1$ is obtained from $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ and $\hat{\gamma}$ is obtained from $\hat{y} = \hat{\beta}_0^* + \hat{\beta}_1^* x + \hat{\gamma} \omega$.

Generally, $\hat{\beta}_1$ and $\hat{\beta}_1^*$ will differ.

This procedure may be repeated using a new regression $x^* = x^{\alpha_1}$ in the calculation.

This procedure generally converges rapidly. Usually, the first stage result α_1 is a satisfactory estimate of α . The round-off error is a potential problem. If enough decimal places are not taken care, then the successive values of α may oscillate badly. If the standard deviation of error (σ) is large or the range of explanatory variable is very small relative to its mean then the estimator may face convergence problems. This situation implies that the data do not support the need for any transformation.