

Chapter 6

Transformation and Weighting to Correct Model Inadequacies

Graphical methods help detect violations of basic assumptions in regression analysis. Now we consider methods and procedures for building models that use data transformation when assumptions are violated.

Variance stabilizing transformations

In regression analysis, it is assumed that the variance of disturbances is constant, i.e., $Var(\varepsilon_i) = \sigma^2$, $i = 1, 2, \dots, n$. Suppose this assumption is violated. A common reason for such isolation is that the study variable follows a probability distribution whose variance is a function of the mean.

Some commonly used variance-stabilizing transformations in the order of their strength are as follows:

Relation of σ^2 to $E(y)$	Transformation
$\sigma^2 \propto \text{constant}$	$y^* = y$ (no transformation)
$\sigma^2 \propto E(y)$	$y^* = \sqrt{y}$ (Poisson data)
$\sigma^2 \propto E(y)[1 - E(y)]$	$y^* = \sin^{-1}(\sqrt{y})$ (Binomial proportion $0 \leq y_i \leq 1$)
$\sigma^2 \propto [E(y)]^2$	$y^* = \ln(y)$
$\sigma^2 \propto [E(y)]^3$	$y^* = 1/\sqrt{y}$
$\sigma^2 \propto [E(y)]^4$	$y^* = \frac{1}{y}$

After making a suitable transformation, use y^* as a study variable in the respective case.

The strength of a transformation depends on the degree of curvature present in the curve between the study and the explanatory variable. The transformation mentioned here ranges from relatively mild to relatively strong. The square root transformation is relatively mild, and the reciprocal transformation is relatively strong. The square root transformation is relatively mild, and the reciprocal transformation is relatively strong.

In general, a mild transformation is applied when the minimum and maximum values do not range much (e.g. $y_{\max} / y_{\min} < 2, 3$) and such a transformation has little effect on the curvature. On the other hand, when the minimum and maximum vary greatly, a strong transformation is needed to have a substantial impact on the analysis.

Analytical methods for selecting a transformation on the study variable

The Box-Cox method

Suppose the normality and/or constant variance of the 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 the square root and $\sqrt{y}$ is used as a study variable in place of y .

Now the linear regression model has parameters β, σ^2 and λ . The Box-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 observations y to be one. It is meaningless for all observations on the study variable to be constant. So there is a discontinuity at $\lambda = 0$. One approach to solve this difficulty is to use $\frac{y^\lambda - 1}{\lambda}$ as a study variable.

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

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

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

When V is applied to each y_i , we get $V = (V_1, V_2, \dots, V_n)'$ as a vector of observations on the transformed study variable, and we use it to fit a linear model

$$V = X\beta + \varepsilon$$

using the least squares or maximum likelihood method.

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 the method of maximum likelihood to the likelihood function for $y^{(\lambda)}$, the log-likelihood function $\ln L[y^{(\lambda)}]$ gives

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

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

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 of the 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 a response to the final model. This model will have a scale difference and an origin shift in comparison to the 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 the residual. But if $\lambda = 0.5$ is within 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

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

where $\chi_{\alpha}^2(1)$ is the upper $\alpha\%$ 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 the height

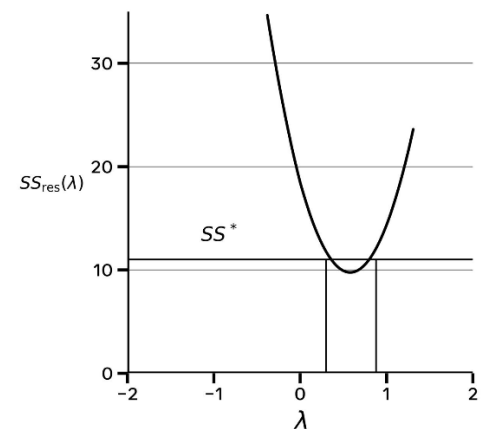
$$L(\hat{\lambda}) - \frac{\chi_{\alpha}^2(1)}{2}$$

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 endpoints of the approximate confidence interval.
- If the 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.



Using the expansion of the exponential function as

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

$$\approx 1 + t,$$

we can approximate and replace $\exp\left[\frac{\chi_{\alpha}^2(1)}{n}\right]$ by $1 + \frac{\chi_{\alpha}^2(1)}{n}$. So in place of $\exp\left[\frac{\chi_{\alpha}^2(1)}{n}\right]$ 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 the sum of squares due to residuals.

Transformation of explanatory variables: Box and Tidwell procedure

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

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

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