Choice with multiple alternatives – 5.2
Specification of the deterministic part

Michel Bierlaire

Box-Cox transforms

The Box-Cox transform of a positive variable x, introduced by Box and Cox (1964), is defined as

$$x(\lambda) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log x & \text{if } \lambda = 0. \end{cases} \tag{1}$$

Note that

$$\lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} = \log x, \tag{2}$$

so that $x(\lambda)$ is continuous [Verify]. It can be embedded in the specification of a utility function:

$$V_{in} = \beta_k x_{ink}(\lambda) + \cdots , \tag{3}$$

where both β_k and λ are estimated from data. Such a specification is not linear-in-parameters. Its flexibility allows to let the data tell if the variable is involved in a linear way ($\lambda = 1$), a logarithmic way ($\lambda = 0$) or as a power law.

If the variable x may take negative values, Box and Cox (1964) propose to shift it before the transform is applied:

$$x(\lambda, \alpha) = \begin{cases} \dfrac{(x+\alpha)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x+\alpha) & \text{if } \lambda = 0, \end{cases} \tag{4}$$

where $\alpha > -x$.

There are other ways to impose the positivity of the argument of the transform. For instance, Manly (1976) suggests to use an exponential:

$$x(\lambda) = \begin{cases} \frac{e^{x\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ x & \text{if } \lambda = 0, \end{cases} \tag{5}$$

while John and Draper (1980) propose to use the absolute value:

$$x(\lambda) = \begin{cases} \operatorname{sign}(x)\frac{(|x|+1)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \operatorname{sign}(x)\log(|x|+1) & \text{if } \lambda = 0. \end{cases} \tag{6}$$

A more complex transform has been proposed by Yeo and Johnson (2000):

$$x(\lambda) = \begin{cases} \dfrac{(x+1)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, x \geq 0; \\ \\ \log(x+1) & \text{if } \lambda = 0, x \geq 0; \\ \\ \dfrac{(1-x)^{2-\lambda}-1}{\lambda-2} & \text{if } \lambda \neq 2, x < 0; \\ \\ -\log(1-x) & \text{if } \lambda = 2, x < 0. \end{cases} \tag{7}$$

Plenty of references are available in the literature. We refer the reader to Sakia (1992) for a review, and to Zarembka (1990) for a discussion in terms of model specification.

References

Box, G. E. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 211–252.

John, J. and Draper, N. (1980). An alternative family of transformations, *Applied Statistics* pp. 190–197.

Manly, B. (1976). Exponential data transformations, *The Statistician* pp. 37–42.

Sakia, R. M. (1992). The box-cox transformation technique: A review, *Journal of the Royal Statistical Society. Series D (The Statistician)* **41**(2): 169–178.
URL: *http://www.jstor.org/stable/2348250*

Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika* **87**(4): 954–959.

Zarembka, P. (1990). Transformation of variables in econometrics, *Econometrics*, Springer, pp. 261–264.