Aclaración sobre la transformación de Box y Cox

Supongamos que el problema de normalidad o homocedasticidad en el análisis de un modelo de regresión lineal pudiera corregirse mediante transformaciones de potencia en la variable de respuesta, y. Es decir, transformar y en y^λ, por ejemplo si $\lambda = 0.5$ entonces $\sqrt{y}$. Luego, el enfoque de Box y Cox consiste en estimar simultáneamente por máxima verosimilitud todos los parámtros del nuevo modelo lineal, β, σ^2 y λ.

Es claro que si $\lambda \to 0$ entonces $y^\lambda \to 1$, luego tenemos un ya que convertiría a todas las respuestas en uno y el problema perdería sentido.

Un enfoque para contrarrestar este problema es considerar la siguiente transformación

$$\frac{y^\lambda - 1}{\lambda},$$

en este caso cuando $\lambda \to 0$ tenemos que,

$$\frac{y^\lambda - 1}{\lambda} \to \ln(y)$$

Luego, se sugiere transformar las variables como

$$\begin{array}{ll} \frac{y^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \ln(y) & \text{para } \lambda = 0. \end{array}$$

Como el análisis de la varianza del modelo no se ve afectado por transformaciones lineales, se puede reescribir

$$\begin{array}{ll} y^\lambda & \text{para } \lambda \neq 0 \\ \ln(y) & \text{para } \lambda = 0. \end{array}$$