
# Distributional Modelling in R

## Transformation Models

Thomas Kneib, Nikolaus Umlauf

https://nikum.org/dmr.html

# Motivation

- In an ideal world, all regression models would be of the form

$$Y = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \sigma^2),$$

which implies

$$Y \sim \mathrm{N}(\boldsymbol{x}'\boldsymbol{\beta}, \sigma^2)$$

or equivalently

$$\frac{Y - \boldsymbol{x}\boldsymbol{\beta}}{\sigma} \sim \mathrm{N}(0, 1)$$

# Transformations

- Transformations are a common approach to make the normality assumption hold at least approximately, e.g.
  - the log-transformation
$$\log(Y) \sim \mathrm{N}(\boldsymbol{x}'\boldsymbol{\beta}, \sigma^2)$$
  - or the Box-Cox-transformation
$$\left(\frac{Y^\lambda - 1}{\lambda}\right) \sim \mathrm{N}(\boldsymbol{x}'\boldsymbol{\beta}, \sigma^2).$$

# Conditional Transformation Models

- Conditional transformation models take this one step further and assume

$$h(Y|\boldsymbol{x}) \stackrel{\mathcal{D}}{=} Z \sim \mathrm{N}(0,1)$$

  where $h(\cdot|\boldsymbol{x})$ is a covariate-dependent transformation function that is strictly increasing in $y$.
- Can be reversed to the generative model

$$Y \stackrel{\mathcal{D}}{=} h^{-1}(Z), \quad Z \sim \mathrm{N}(0,1).$$

# Conditional Transformation Models

- Implies the the density

$$f(y_i|\boldsymbol{x}_i) = f_{\mathrm{ref}}(h_{\boldsymbol{x}_i}(y_i)) \left| \frac{\partial h_{\boldsymbol{x}_i}(y_i)}{\partial y_i} \right|$$

  and cumulative distribution function

$$\mathbb{P}(y_i \leq c|\boldsymbol{x}_i) = F_{y_i}(c|\boldsymbol{x}_i) = F_{\mathrm{ref}}(h_{\boldsymbol{x}_i}(c)) = \mathbb{P}(z_i \leq h_{\boldsymbol{x}_i}(c)).$$

  where $f_{\mathrm{ref}}$ and $F_{\mathrm{ref}}$ are the density and the CDF of the reference distribution (usually the standard normal).

# Heteroscedastic Linear Model

- The transformation model

$$h(y|\boldsymbol{x}) = \boldsymbol{x}'\tilde{\boldsymbol{\beta}} + y\boldsymbol{x}'\boldsymbol{\alpha}, \quad \text{with } \boldsymbol{x}'\boldsymbol{\alpha} > 0$$

  implies

$$Y \sim \mathrm{N}\left(-\frac{\boldsymbol{x}'\tilde{\boldsymbol{\beta}}}{\boldsymbol{x}'\boldsymbol{\alpha}}, \frac{1}{(\boldsymbol{x}'\boldsymbol{\alpha})^2}\right).$$

# Likelihood-Based Inference

- Conditional transformation models can be estimated in a likelihood-based framework where the likelihood arises from

$$f_{Y|\boldsymbol{x}}(y) = f_{\mathrm{ref}}(h(y|\boldsymbol{x})) \left| \frac{\partial}{\partial y} h(y|\boldsymbol{x}) \right| .$$

  (the density implied by the transformation model).
- Main difficulty: Ensuring monotonicity of the transformation function.
- Can, for example, be implemented via Bernstein polynomial approximations for all functions envolved in the model specification.