Approach 5: Variance-Stabilizing Transformation

For certain distributions of Y, the variance is a known function of the mean, and consequently a known function $h(Y)$ has an approximately constant variance.

For example:

Distribution	Mean-variance	Transformation $h(Y)$
Poisson	$\sigma^2 = \mu$	$\sqrt{Y}$
Binomial	$\sigma^2 \propto \mu(1-\mu)$	$\sin^{-1}\sqrt{Y}$
lognormal	$\sigma^2 \propto \mu^2$	$\log Y$

Empirically, transformation often was found also to make the dependence linear.

So for example counted data Y, where the Poisson distribution might be relevant, might be modeled as

$$\mathrm{E}\left(\sqrt{Y_j}\,\middle|\,\mathbf{x}_j\right) = \mathbf{x}_j^T\boldsymbol{\beta},$$
$$\mathrm{var}\left(\sqrt{Y_j}\,\middle|\,\mathbf{x}_j\right) = \sigma^2.$$

Remarks

- We may not be so lucky!
- Even if we are, $\boldsymbol{\beta}$ is hard to interpret.

Approach 6: Box-Cox Transformation

Let the data determine the transformation in a parametric family $h(Y, \lambda)$.

Box and Cox used the power family:

$$h(Y, \lambda) = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log Y & \lambda = 0. \end{cases}$$

- Assume linearity and normality: $h(Y_j, \lambda) \sim N\left(\mathbf{x}_j^T \boldsymbol{\beta}, \sigma^2\right)$.
- Estimate $\lambda, \boldsymbol{\beta},$ and σ^2 by ML.

Remarks

Same issues as before.

One transformation is assumed to achieve

- linearity,
- normality,
- and constant variance.

β is hard to interpret.

Approach 7: Transform Both Sides (TBS)

Suppose we have scientific reasons for a model $f(\mathbf{x}, \boldsymbol{\beta})$.

Assume that, conditionally on $\mathbf{x}_j$,

$$h(Y_j, \lambda) = h\{f(\mathbf{x}_j, \boldsymbol{\beta}), \lambda\} + e_j$$

The deviations $e_1, e_2, \ldots, e_n$ satisfy the standard assumptions:

- constant variance σ^2;
- independence and normality.

Estimate $\boldsymbol{\beta}$, λ, and σ^2 by ML.

- Still requires a strong assumption.
- Still difficult to interpret.

But: if h is smooth and invertible and σ is small relative to $f(\mathbf{x}, \beta)$, we can use a linear approximation:

$$\begin{aligned} Y_j &= h^{-1}\left[h\left\{f\left(\mathbf{x}_j, \beta\right), \lambda\right\} + e_j, \lambda\right] \\ &\approx f\left(\mathbf{x}_j, \beta\right) + \frac{1}{h_y\left\{f\left(\mathbf{x}_j, \beta\right), \lambda\right\}} \times e_j \end{aligned}$$

where

$$h_y(y, \lambda) = \frac{\partial h(y, \lambda)}{\partial y}.$$

So to this order of approximation,

$$\mathrm{E}\left(Y_j | \mathbf{x}_j \right) \approx f\left(\mathbf{x}_j, \boldsymbol{\beta} \right)$$

and

$$\begin{aligned} \mathrm{var}\left(Y_j | \mathbf{x}_j \right) &\approx \frac{\sigma^2}{h_y\left\{ f\left(\mathbf{x}_j, \boldsymbol{\beta} \right), \lambda \right\}^2} \\ &\approx \frac{\sigma^2}{h_y\left\{ \mathrm{E}\left(Y_j \right), \lambda \right\}^2}. \end{aligned}$$

For the Box-Cox family, $h_y(y, \lambda) = y^{\lambda - 1}$, so

$$\mathrm{var}\left(Y_j | \mathbf{x}_j \right) \approx \sigma^2 f\left(\mathbf{x}_j, \boldsymbol{\beta} \right)^{2(1-\lambda)} \approx \sigma^2 \mathrm{E}\left(Y_j | \mathbf{x}_j \right)^{2(1-\lambda)}.$$

So:

normality and constant variance on the transformed scale

are (approximately) equivalent to

normality and non-constant variance on the original scale,

with the variance given by a certain function of the mean.

Henceforth, we do not consider transformations. Instead, we model variances, and sometimes distributions, explicitly.

Approach 8: ML for an Assumed Model

E.g. Y is a count, which we could model as Poisson-distributed, with a model for the mean:

$$E(Y|\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}).$$

This implies a model for the variance,

$$\text{var}(Y|\mathbf{x}) = E(Y|\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta})$$

but we do not need it in likelihood-based inference.

Approach 9: Generalized Least Squares (GLS)

Assume

$$\mathrm{E}(Y|\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}),$$
$$\mathrm{var}(Y|\mathbf{x}) = \sigma^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x})^2.$$

Here:

- $\boldsymbol{\beta}$ is, as before, the vector of parameters in the mean function;
- $\boldsymbol{\theta}$ is a possible additional parameter in the structure of the variance function;
- σ^2 is a possible over-dispersion parameter.

Ad hoc estimation scheme:

1. Get initial estimate of $\boldsymbol{\beta}$ using OLS;
2. Get initial estimate of $\boldsymbol{\theta}$, if needed, and construct initial *estimated weights*

$$\hat{w}_j = \frac{1}{g\left(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}, \hat{\boldsymbol{\theta}}, \mathbf{x}\right)^2}$$

3. Re-estimate $\boldsymbol{\beta}$ using WLS, treating $\hat{w}_j$ as fixed: solve the estimating equation

$$\sum_{j=1}^{n} \hat{w}_j \left\{Y_j - f\left(\mathbf{x}_j, \boldsymbol{\beta}\right)\right\} f_{\boldsymbol{\beta}}\left(\mathbf{x}_j, \boldsymbol{\beta}\right) = \mathbf{0}.$$

Remarks

In step iii, equivalently, $\hat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ minimizes

$$S_{\hat{w}}(\boldsymbol{\beta}) = \sum_{j=1}^{n} \hat{w}_j \left\{ Y_j - f\left(\mathbf{x}_j, \boldsymbol{\beta}\right)\right\}^2 .$$

As noted before, this is a computational alternative to solving the estimating equation, but is not generally preferred.

The estimating-equation approach may be extended to cases where no such objective function exists.

Could repeat step ii with $\hat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ instead of $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, then repeat step iii.

We could also iterate, perhaps to convergence.

Approach 10: ML

Same moment structure as in Approach 9:

$$\mathrm{E}(Y|\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}),$$
$$\mathrm{var}(Y|\mathbf{x}) = \sigma^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x})^2.$$

Add an assumption about the distribution of $Y|\mathbf{x}$:

$$Y = f(\mathbf{x}, \boldsymbol{\beta}) + \sigma g(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x}) Z$$

where Z has mean 0 and variance 1, such as $N(0, 1)$, and use likelihood methods.

When $g(\cdot)$ involves $\boldsymbol{\beta}$, the ML estimating equation for $\boldsymbol{\beta}$ is *not* the same as in GLS.

Comments on Approaches 5–10

Approaches 5–7 require (lucky) transformation, and the parameters can be hard for interpretation.

Approaches 8–10 are the most interesting.

Depending on the form of the functions $f(\cdot)$ and $g(\cdot)$, these approaches will sometimes yield the same estimating equations.