IX Regression models

IX.1 Regression with errors in the variables

Idea: Before $Y_i = \alpha + \beta x_i + \varepsilon_i$ with fixed x_i

Now X_i is going to be a random variable.

Def. EIV model

$$\begin{cases} Y_i = \underbrace{\alpha + \beta \xi_i}_{=:\eta_i} + \varepsilon_i & \text{with } \varepsilon_i \sim n(0, \sigma_\varepsilon^2) \\ X_i = \xi_i + \delta_i & \text{with } \delta_i \sim n(0, \sigma_\delta^2) \end{cases}$$

ξ_i and η_i are called the latent variables.

Data fitting: Since the x-variable has some errors

the least squares (based on vertical distance) is replaced by

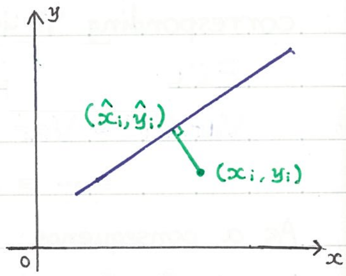
the total least squares $= \sum_{i=1}^{n} \left((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right)$

(with the line given $a+bx$) $= \frac{1}{1+b^2} \sum_{i=1}^{n} \left((y_i - (a + b x_i) \right)^2$

By minimization over a and b, one gets

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{-(S_{xx} - S_{yy}) + \sqrt{(S_{xx} - S_{yy})^2 + 4 S_{xy}^2}}{2 S_{xy}}$$



Recall: $Y_i \sim n(\alpha + \beta \xi_i, \sigma_\varepsilon^2)$;

$X_i \sim n(\xi_i, \sigma_\delta^2)$

This leads to the likelihood function for the sample $\{(X_i, Y_i)\}_{i=1}^{n}$

$$L(\alpha, \beta, \xi_1, \cdots, \xi_n, \sigma_\varepsilon^2, \sigma_\delta^2 \mid \underline{x}, \underline{y}) =$$

$$= \frac{1}{(2\pi)^n} \frac{1}{(\sigma_\varepsilon^2 \sigma_\delta^2)^{n/2}} \exp\left(-\sum \frac{(x_i - \xi_i)^2}{2\sigma_\delta^2}\right) \exp\left(-\sum_{i=1}^{n} \frac{(y_i - (\alpha + \beta \xi_i))^2}{2\sigma_\varepsilon^2}\right)$$