

## Empiriske Metoder Vinter 2003

### OPGAVE 2

I en regressionsmodel er man interesseret i at bestemme en krum kurve af formen

$$y_i = \alpha x_i^\lambda + e_i ,$$

hvor restleddene  $e_i$  er uafhængige normalfordelte med middelværdi nul. Desuden er man interesseret i heteroskedasticitet, der svarer til, at variationen er proportional med niveauet. Det betyder, at standardafvigelsen på  $y_i$  er proportional med  $x_i^\lambda$ , således at der gælder

$$\text{var}(y_i) = \text{var}(e_i) = \sigma^2 x_i^{2\lambda} .$$

Antag nu at der foreligger  $n$  observationer af  $y_i$  og  $x_i$ . I denne model er den simultane tæthedsfunktion givet ved

$$\begin{aligned} & \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 x_i^{2\lambda}}} \exp\left( -\frac{1}{2} \frac{(y_i - \alpha x_i^\lambda)^2}{\sigma^2 x_i^{2\lambda}} \right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 x_i^{2\lambda}}} \exp\left( -\frac{y_i^2 x_i^{-2\lambda}}{2\sigma^2} - \frac{\alpha^2}{2\sigma^2} + \frac{\alpha y_i x_i^{-\lambda}}{\sigma^2} \right) . \end{aligned}$$

a)

Opskriv loglikelihoodfunktionen og vis at scorevektoren er

$$\mathbf{q}_\alpha = -\frac{n\alpha}{\sigma^2} + \frac{1}{\sigma^2} \sum y_i x_i^{-\lambda}$$

$$\mathbf{q}_\lambda = -\sum \log(x_i) + \frac{1}{\sigma^2} \sum y_i^2 x_i^{-2\lambda} \log(x_i) - \frac{\alpha}{\sigma^2} \sum y_i x_i^{-\lambda} \log(x_i)$$

$$\mathbf{q}_\sigma = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum \frac{(y_i - \alpha x_i^\lambda)^2}{x_i^{2\lambda}}$$

Husk at

$$\frac{d}{d\gamma} x^\gamma = \log(x) x^\gamma .$$

I en konkret estimation benyttes data, hvor  $y$  betegner  $\text{co}_2$  udledningen pr. indbygger og  $x$  betegner urbaniseringsgraden for  $n = 155$  lande, hvorfra data er vist på figuren. Først estimeres en lineær model med restriktionen  $\lambda = 1$  til

$$\hat{\alpha} = 0.0653 \quad \hat{\sigma}^2 = 0.0000455 \quad \text{og altså} \quad \tilde{\lambda} = 1 .$$

Den dertil svarende rette linie er indlagt på figuren. Loglikelihoodfunktionens værdi er for disse parameterverdier - 388.7.

Maksimum likelihood estimaterne, når alle parametre estimeres frit, er

$$\hat{\alpha} = 0.00578 \quad \hat{\lambda} = 1.580 \quad \hat{\sigma}^2 = 0.0000285 ,$$

idet maksimum af loglikelihoodfunktionen er - 361.8. Den dertil svarende krumme kurve er også indlagt på figuren.

Desuden estimeres variansen (atter med parametrene i rækkefølgen  $\alpha, \lambda, \sigma^2$ ) ved  $-\mathbf{Q}^{-1}$  til

$$-\mathbf{Q}^{-1} = \begin{bmatrix} 0.00000103 & -0.0000366 & 0.00000000724 \\ -0.0000366 & 0.00158 & -0.000000313 \\ 0.00000000724 & -0.000000313 & 0.000000000710 \end{bmatrix} .$$

**b)**

Test hypotesen  $\lambda = 1$  ved et likelihoodratio test og ved et Wald test.

**c)**

Opstil Wald testet for den sammensatte hypotese  $\alpha = 0.006$  og  $\lambda = 1.5$  og  $\sigma^2 = 0.000025$ , men I skal ikke regne noget ud ved matrixinversion og matrixmultiplikation - kun opstille regnestykket.

**d)**

I Danmark er urbaniseringsgraden  $x = 85$  (dvs. at 85% af befolkningen bor i byer). Beregn talværdien af forudsigelsen af  $\text{co}_2$  udslippet pr. indbygger ud fra den estimerede model og beregn variansen på denne forudsigelse, dvs. beregn  $\text{var}(\hat{y}) = \text{var}(\hat{\alpha} \times 85^{\hat{\lambda}})$  - også som et tal.

