应用回归分析

上海财经大学　统计与管理学院

第五章加权与失拟

章节概括：

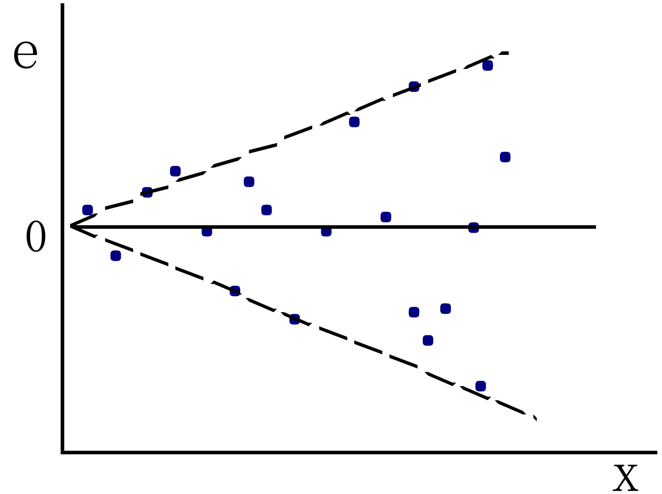
- 加权最小二乘法
- 失拟检验
- 联合置信区间

加权最小二乘法

- 多元线性模型

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p+e$$

- 方差函数



(b)

$$\mathrm{Var}(Y|X=x)=\sigma^2$$

$$\mathrm{Var}(Y|X=\mathbf{x}_i)=\mathrm{Var}(e_i)=\sigma^2/w_i$$

加权最小二乘法

- $w_1, \ldots, w_n$ 已知正数
- 多元线性模型

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \qquad \mathrm{Var}(\mathbf{e}) = \sigma^2 \mathbf{W}^{-1}$$

$$\mathbf{W} = \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \cdots & \\ & & & w_n \end{pmatrix}$$

- 加权残差平方和

$$RSS(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \sum w_i (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2$$

加权最小二乘法

- 加权最小二乘估计

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$$

- 变量转换

$$\mathbf{W}^{1/2}\mathbf{W}^{-1/2} = \mathbf{I} \qquad \mathrm{Var}(\mathbf{W}^{1/2}\mathbf{e}) = \sigma^2\mathbf{I}$$

$$\mathbf{W}^{1/2}\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{1/2}\mathbf{e}$$

$$\mathbf{Z} = \mathbf{W}^{1/2}\mathbf{Y},\ \mathbf{M} = \mathbf{W}^{1/2}\mathbf{X},\ \text{and } \mathbf{d} = \mathbf{W}^{1/2}\mathbf{e}$$

加权最小二乘法

$$\mathbf{M} = \begin{pmatrix} \sqrt{w_1} & \sqrt{w_1}x_{11} & \cdots & \sqrt{w_1}x_{1p} \\ \sqrt{w_2} & \sqrt{w_2}x_{21} & \cdots & \sqrt{w_2}x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{w_n} & \sqrt{w_n}x_{n1} & \cdots & \sqrt{w_n}x_{np} \end{pmatrix}$$

$$\mathbf{Z} = \begin{pmatrix} \sqrt{w_1}y_1 \\ \sqrt{w_2}y_2 \\ \vdots \\ \sqrt{w_n}y_n \end{pmatrix} \qquad \mathrm{Var}(\mathbf{d}) = \sigma^2\mathbf{I},$$

加权最小二乘法

$$\mathbf{Z} = \mathbf{M}\beta + \mathbf{d}$$

$$\begin{aligned}\hat{\beta} &= (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{Z} \\ &= \left((\mathbf{W}^{1/2}\mathbf{X})'(\mathbf{W}^{1/2}\mathbf{X})\right)^{-1}(\mathbf{W}^{1/2}\mathbf{X})'(\mathbf{W}^{1/2}\mathbf{Y}) \\ &= \left(\mathbf{X}'\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{X}\right)^{-1}(\mathbf{X}'\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}\mathbf{Y})\end{aligned}$$

P=1

- 一元线性回归普通最小二乘法的残差平方和为：

$$RSS(\beta_0,\beta_1)=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2$$

- 一元线性回归的加权最小二乘的残差平方和为：

$$RSS_w(\beta_0,\beta_1)=\sum_{i=1}^{n}w_i(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}w_i(y_i-\beta_0-\beta_1x_i)^2$$

$$RSS_w=\sum_{i=1}^{n}w_i(y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\cdots-\beta_px_{ip})^2$$

P=1

- 加权最小二乘估计为：

$$\begin{cases} \hat{\beta}_{0w} = \bar{y}_w - \hat{\beta}_{1w}\bar{x}_w \\ \hat{\beta}_{1w} = \dfrac{\sum_{i=1}^{n} w_i(x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum_{i=1}^{n} w_i(x_i - \bar{x}_w)^2} \end{cases}$$

其中， $\bar{x}_w = \dfrac{1}{\Sigma w_i}\sum w_i x_i$ 是自变量的加权平均；

$\bar{y}_w = \dfrac{1}{\Sigma w_i}\sum w_i y_i$ 是因变量的加权平均。

异方差

- $\mathrm{Var}(\mathbf{e}) = \mathbf{\Sigma}$

- 有参假设、 样本估计……

- $\mathrm{Var}(Y|X)$ 依赖于 $\mathrm{E}(Y|X)$

  8.3方差稳定性变换法

  (variance stabilizing transformation)

- 广义线性模型 generalized linear model

  12章 逻辑回归 (logistics regression)

方差稳定变换

- 方差稳定变换

(1) 如果σ_i^2与$E(y_i)$存在一定的比例关系，使用$y' = \sqrt{y}$；

(2) 如果σ_i与$E(y_i)$存在一定的比例关系，使用$y' = \log(y)$；

(3) 如果$\sqrt{\sigma_i}$与$E(y_i)$存在一定的比例关系，使用$y' = \frac{1}{y}$

失拟检验

- 方差已知
- 假定

$$e_i \sim \mathrm{NID}(0, \sigma^2 / w_i)$$

$$w_i \text{ and } \sigma^2 \qquad \text{已知}$$

- 统计量

$$X^2 = \frac{RSS}{\sigma^2} = \frac{(n-p')\hat{\sigma}^2}{\sigma^2}$$

服从卡方分布，自由度 $n-p'$

失拟检验

- 方差未知

$$\begin{aligned} \text{NH:} \quad & \mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{e} \\ \text{AH:} \quad & \mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{e} \end{aligned}$$

- 统计量

$$F = \frac{(RSS_{NH} - RSS_{AH})/(\mathrm{df}_{NH} - \mathrm{df}_{AH})}{RSS_{AH}/\mathrm{df}_{AH}}$$

当NH成立时服从 $F(\mathrm{df}_{NH} - \mathrm{df}_{AH}, \mathrm{df}_{AH})$

检验功效

- 当AH成立时

$$\frac{RSS_{NH}-RSS_{AH}}{\sigma^2}\sim\chi^2(\mathrm{df}_{NH}-\mathrm{df}_{AH},\kappa)$$

$$\kappa=\frac{\boldsymbol{\beta}_2'\mathbf{X}_2'(\mathbf{I}-\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2\boldsymbol{\beta}_2}{q\sigma^2}$$

联合置信区域

- $\boldsymbol{\beta}$ 的 $(1-\alpha)\times 100\%$ 置信区域为

$$\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{p'\hat{\sigma}^2} \le F(\alpha; p', n-p')$$

$$\frac{(\boldsymbol{\beta}_1-\hat{\boldsymbol{\beta}}_1)'\mathbf{S}^{-1}(\boldsymbol{\beta}_1-\hat{\boldsymbol{\beta}}_1)}{q\hat{\sigma}^2} \le F(\alpha; q, n-p')$$

联合置信区域

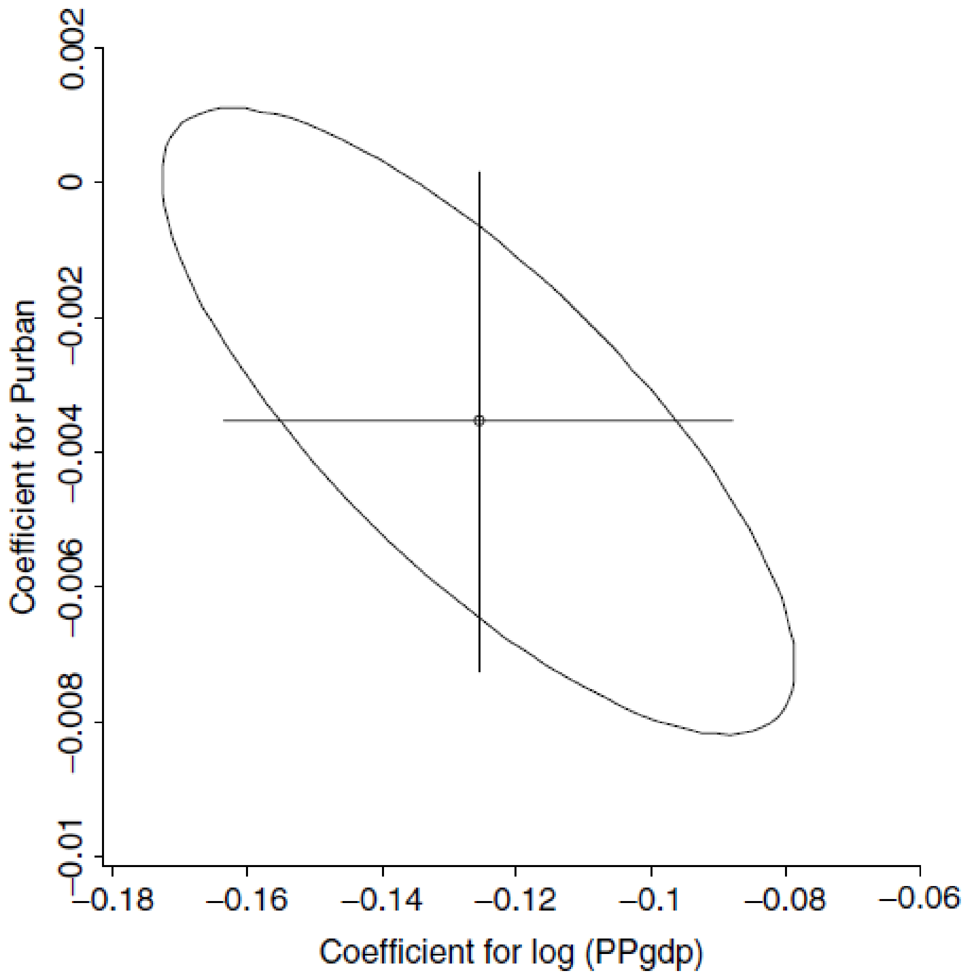


FIG. 5.3 95% confidence region for the UN data.

Thank You !