Modèles linéaires - Statistical Analysis System (SAS)
Chapitre 2 : Hypothèses

Ordinary Least Squares Estimator (OLS)

Emmanuelle Gautherat[a]

(a) Crieg-Regards, Université de Reims Champagne Ardenne

Second semestre - 4 ECTS

Outline

Introduction

Méthode OLS fondée sur la corrélation, la moyenne empirique et utilise pleinement l'ensemble des hypothèses
–> très sensible à l'écart à ces hypothèses.
Importance de la détection des écarts aux hypothèses.
–> très sensibles aux valeurs extrêmes ou atypiques.
Importance de la détection des valeurs atypiques : plusieurs sens donné à 'atypique'.

Ici : cadre du modèle OLS standard avec constante.
Par défaut, l'ensemble des hypothèses est supposé vérifié.

Objets génériques : Matrice H

On note dorénavant $H = (h_{ij})_{i,j=1,\dots,n}$ (prononcer "matrice hat") la matrice de projection $\Pi_{(x)}$.

$$\widehat{Y} = \Pi_{(x)} Y = HY.$$

En notant $\tilde{x}$ la matrice dont les colonnes sont les variables $x^{(k)}$ centrées :

$$\tilde{x} = \begin{pmatrix} x_1^{(1)} - \overline{x^{(1)}} & \dots & x_1^{(K)} - \overline{x^{(K)}} \\ \vdots & & \vdots \\ x_n^{(1)} - \overline{x^{(1)}} & \dots & x_n^{(K)} - \overline{x^{(K)}} \end{pmatrix},$$

on obtient

$$\begin{aligned} \widehat{Y}_i &= (HY)_i = \sum_{j=1}^n h_{ij} Y_j = \sum_{j=1}^n (\frac{1}{n} + \tilde{x}(\tilde{x}^t \tilde{x})^{-1} \tilde{x}^t)_{ij} Y_j \\ h_{ii} &= \frac{1}{n} + \sum_{k=1}^K \left(\frac{(v^k)^t (x_i - \overline{x})}{\sqrt{\lambda_k}} \right)^2 \end{aligned}$$

avec λ_k val. propre associée au vect. propre normé v^k de $\tilde{x}^t \tilde{x}$.
R : Comme $H = \Pi_{(x)}$, on obtient la diagonale de la matrice H avec

```
diag(x%*% solve( t(x) %*% x )%*% t(x))
```

Ou encore

`hat(x)`.

Estimateur du niveau de bruit $\widehat{\sigma}_{(-i)}$

On rappelle que $\widehat{u}_i$, le résidu pour l'observation i, est défini par

$$\begin{aligned}\widehat{u} = \Pi_{x^\perp} Y = (Id_n - H)Y \quad &\sim \quad \mathcal{N}_n(0\,;\sigma^2(Id - H)) \\ \widehat{u}_i \quad &\sim \quad \mathcal{N}_1(0\,;\sigma^2(1 - h_{ii}).\end{aligned}$$

Estimateur du niveau de bruit sans l'observation i : $\widehat{\sigma}^2_{(-i)}$.

- $$\widehat{\sigma}^2 = \frac{\|\widehat{u}\|_2^2}{n-K-1} = \frac{\widehat{u}^t\widehat{u}}{n-K-1}.$$
- Soit $\widehat{\sigma}^2_{(-i)}$ un estimateur du niveau de bruit calculé sans utiliser l'observation i. C'est-à-dire :

$$\begin{aligned}\widehat{\sigma}^2_{(-i)} \quad &= \quad \frac{\widehat{u}^t_{(-i)}\widehat{u}_{(-i)}}{n-K-2} \\ \widehat{\sigma}^2_{(-i)} \quad &\sim \quad \chi^2((n-1)-(K+1)) = \chi^2(n-K-2)\end{aligned}$$

.

ModLin Chap 1

E. Gautherat

Objets génériques

Résidus standardisés $\widehat{e}_i$, résidus studentisés $\widehat{t}_i$

Rappel : Pour tout i, $\frac{\widehat{u}_i}{\sqrt{\sigma^2(1-h_{ii})}} \sim \mathcal{N}_1(0;1)$.

Résidus standardisés pour l'observation i : $\widehat{e}_i$

On définit $\widehat{e}_i$ par :

$$\widehat{e}_i = \frac{\widehat{u}_i}{\sqrt{\widehat{\sigma^2}(1-h_{ii})}} \sim T(n-K-1).$$

Résidus studentisés pour l'observation i : $\widehat{t}_i$

On définit $\widehat{t}_i$ par :

$$\widehat{t}_i = \frac{\widehat{u}_i}{\sqrt{\widehat{\sigma}^2_{(-i)}(1-h_{ii})}} \sim T(n-K-2).$$

Avec R :

Pour reg<-lm(.....)

$\widehat{u}$ ——-> residuals(reg)

$\widehat{e}$ ——-> rstandard(reg)

$\widehat{t}$ ——-> rstudent(reg)

$\widehat{\sigma}^2_{(-i)}$ ——-> Influ2<-lm.influence(reg) ; Influ2$sigma

Prédiction sans l'individu "i" : $\widehat{Y}_{(-i)}$ et $\widehat{\beta}_{(-i)}$

On désigne par $\widehat{Y}_{(-i)}$ la prédiction de Y lorsque l'individu i a été omis du calcul de $\widehat{\beta}$.

Soit $x_{(-i)}$, la matrice x à laquelle on a retiré la ligne i : $x_{(-i)} \in \mathcal{M}_{n-1,K+1}(\mathbb{R})$.
Soit $Y_{(-i)}$le vecteur des endogènes auquel on a retiré l'observation i : $Y_{(-i)} \in \mathcal{M}_{n-1,1}(\mathbb{R})$

On remarque que $Y_{(-i)}$ a bien n valeurs : la valeur de Y pour i est prédite à partir des autres individus.
On définit :

$$\begin{aligned} \widehat{\beta}_{(-i)} &= (x^t_{(-i)} x_{(-i)})^{-1} x^t_{(-i)} Y_{(-i)} \\ \widehat{Y}_{(-i)} &= x\widehat{\beta}_{(-i)}. \end{aligned}$$

On a bien

$$\widehat{\beta}_{(-i)} \in \mathcal{M}_{K+1,1}(\mathbb{R}).$$