Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

OLS #15 (moindres carrés généralisés - MCG)

Régression Linéaire

- Supposons dans ce cours disposer d'un modèle linéaire
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
où $\mathbf{X}$ est une matrice satisfaisant l'hypothèse $\mathcal{H}_1$ ($\mathbf{X}$ est de rang plein).
- On supposera disposer du modèle correct (bon ensemble de covariables et correctement transformées), et que le modèle n'exhibe pas de point aberrant ni de point levier; hypothèses difficilement satisfaites en pratique!
- En revanche, on suppose
$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ et } \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega}$$
où $\boldsymbol{\Omega}$ est une matrice symétrique définie positive!
- Le modèle est dit hétéroscédastique: quelles sont les conséquences? comment y remédier?

Données Agrégées

- Supposons qu'un modèle linéaire homoscédastique standard existe entre une variable $\tilde{Y}$ et un ensemble de variables explicatives pour $k_1 + \cdots + k_n$ individus; supposons justement qu'on n'observe pas le modèle au niveau individuel mais au niveau agrégé pour les k_j individus.
  [ex. recueil du salaire dans différentes provinces du Canada! Une seule valeur moy. par province]
- On observe: $Y_i = k_i^{-1} \sum_{j=1}^{i} \tilde{Y}_j$ et le modèle s'écrit toujours: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- On note que $\mathrm{Var}(Y_i) = k_i^{-2} \sum_j \mathrm{Var}(\tilde{Y}_j) = \sigma^2 / k_i$ et ainsi

$$\mathrm{Var}\left(\mathbf{Y}\right) = \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Omega} = \sigma^2 \underbrace{\mathrm{diag}\left(\frac{1}{k_1}, \ldots, \frac{1}{k_n}\right)}_{\boldsymbol{\Omega}}$$

Groupes

- Partons d'un modèle linéaire homoscédastique $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, et supposons que les individus soient regroupés en deux groupes de taille n_1 et n_2, $n = n_1 + n_2$ (par exemple discriminés par une variable sexe);
- Il n'est pas insensé de penser que $\text{Var}(\varepsilon_i) = \sigma_1^2$ pour $i = 1, \dots, n_1$ et σ_2^2 pour $i = n_1 + 1, \dots, n_1 + n_2$.
- les paramètres σ_1^2 et σ_2^2 sont inconnus!
- Pour cet exemple, on aurait

$$\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \underbrace{\text{diag}\left(\pi_1^2, \dots, \pi_1^2, \pi_2^2, \dots, \pi_2^2\right)}_{\boldsymbol{\Omega}}$$

avec $\sigma_j^2 = \sigma^2 \pi_j^2$, $j = 1, 2$.

Corrélation Temporelle

- Supposons que le bruit soit homogène en variance. En revanche, on supposera que la série $Y_1, \ldots, Y_n$ soit une série chronologique (acquise au cours du temps). Il paraît sensé alors que
$$\text{Cov}(\varepsilon_i, \varepsilon_j) \neq 0, \quad i \neq j.$$
- L'exemple le plus simple d'une telle série est un modèle AR(1): $\varepsilon_{t+1} = \rho\varepsilon_t + \eta_t$ où $|\rho| < 1$ et η_t est un bruit blanc gaussien de variance σ^2.
- On montre que $\text{Cov}(\varepsilon_t, \varepsilon_{t+k}) = \sigma^2 \frac{\rho^k}{1-\rho^2}$ et ainsi

$$\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Omega} \text{ avec } \quad \boldsymbol{\Omega} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \vdots & \ddots & \ddots & \vdots & \\ \rho^{n-1} & \rho^{n-2} & \ldots & 1 & \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ où } \mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega}.$$

L'estimateur des MCO, s'écrit toujours $\hat{\boldsymbol{\beta}}^{\text{MCO}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ et

$$\mathbb{E}(\hat{\boldsymbol{\beta}}^{\text{MCO}}) = \boldsymbol{\beta} \quad \text{et} \quad \mathrm{Var}\left(\hat{\boldsymbol{\beta}}^{\text{MCO}}\right) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}.$$

Soit $\hat{\sigma}^2_{\text{MCO}} = \|\hat{\boldsymbol{\varepsilon}}\|/(n-p)$ alors $\mathbb{E}\hat{\sigma}^2_{\text{MCO}} = \mathrm{tr}(\mathcal{P}_{\mathbf{X}^\perp}\boldsymbol{\Omega})/(n-p) \neq \sigma^2$ (en général).

- $\hat{\boldsymbol{\beta}}^{\text{MCO}}$ reste un estimateur sans biais, donc la lecture des coefficients estimés reste pertinente même en présence d'un bruit coloré.
- Ceci est confirmé: on pourrait montrer (en rajoutant quelques hypothèses) que $\hat{\boldsymbol{\beta}}^{\text{MCO}}$ est un estimateur consistant de $\boldsymbol{\beta}$.
- En revanche, σ^2 n'est pas correctement estimée, et par conséquence $\text{Var}(\hat{\boldsymbol{\beta}}^{\text{MCO}})$ n'est pas correctement estimée!
- Ceci signifie que la lecture des erreurs standard n'est pas pertinente; les IC , RC , tests d'hypothèses ne tenant pas compte du bruit coloré sont donc (a priori) sans intérêt et dangereux.
- certaines variables peuvent paraître significatives alors qu'elles ne le sont pas (et inversement): il faut remédier à cette situation!

- On supposera dans un premier temps que $\mathbf{\Omega}$ est connue, et notons $\mathbf{\Omega}^{-1/2}$ sa racine carré inverse (qui existe car $\mathbf{\Omega} > 0$).
- On construit le modèle

$$\underbrace{\mathbf{\Omega}^{-1/2}\mathbf{Y}}_{:=\mathbf{Y}'} = \underbrace{\mathbf{\Omega}^{-1/2}\mathbf{X}}_{:=\mathbf{X}'}\boldsymbol{\beta} + \underbrace{\mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}}_{:=\boldsymbol{\varepsilon}'}.$$

- Il est clair que le modèle $\mathbf{Y}' = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$ est un modèle linéaire homoscédastique car
  - $\mathbf{X}'$ est de plein rang!
  - $\mathrm{Var}(\boldsymbol{\varepsilon}') = \mathrm{Var}(\mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}) = \mathbf{\Omega}^{-1/2}\mathrm{Var}(\boldsymbol{\varepsilon})\mathbf{\Omega}^{-1/2} = \sigma^2\mathbb{I}_n$.

On appelle estimateur des moindres carrés généralisés l'estimateur $\hat{\boldsymbol{\beta}}^{\mathsf{MCG}}$ égal à l'estimateur des MCO du modèle $\mathbf{Y}' = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$. Ainsi

$$\hat{\boldsymbol{\beta}}^{\mathsf{MCG}} = (\mathbf{X}'^{\top}\mathbf{X}')^{-1}\mathbf{X}'^{\top}\mathbf{Y}' = \left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{Y}.$$

- Conséquence du théorème de Gauss-Markov: $\hat{\boldsymbol{\beta}}^{\text{MCG}}$ est l'estimateur linéaire sans biais de variance minimale du modèle $\mathbf{Y}' = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$; $\hat{\boldsymbol{\beta}}^{\text{MCO}}$ étant un autre estimateur linéaire sans biais de $\boldsymbol{\beta}$
$$\mathrm{Var}(\hat{\boldsymbol{\beta}}^{\text{MCG}}) \leq \mathrm{Var}(\hat{\boldsymbol{\beta}}^{\text{MCO}}).$$

- Conséquence du théorème de Gauss-Markov: $\hat{\boldsymbol{\beta}}^{\text{MCG}}$ est l'estimateur linéaire sans biais de variance minimale du modèle $\mathbf{Y}' = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$; $\hat{\boldsymbol{\beta}}^{\text{MCO}}$ étant un autre estimateur linéaire sans biais de $\boldsymbol{\beta}$
$$\mathrm{Var}(\hat{\boldsymbol{\beta}}^{\text{MCG}}) \leq \mathrm{Var}(\hat{\boldsymbol{\beta}}^{\text{MCO}}).$$

Supposons $\boldsymbol{\Omega} = \mathrm{diag}(\pi_i^2, i = 1, \ldots, n)$, on peut montrer que la procédure des MCG est une procédure de MC pondérés, i.e.

$$\hat{\boldsymbol{\beta}}^{\text{MCG}} = \mathrm{argmin} f(\boldsymbol{\beta}), \quad f(\boldsymbol{\beta}) = \|\boldsymbol{\Omega}^{-1/2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\|^2 = \sum_{i=1}^n w_i (Y_i - \mathbf{x}_i' \boldsymbol{\beta})^2$$

avec $w_i = 1/\pi_i^2$.