Émission de NOx lors la combustion d'un moteur (régression semi-paramétrique)

GMMA 106

Cas d'étude

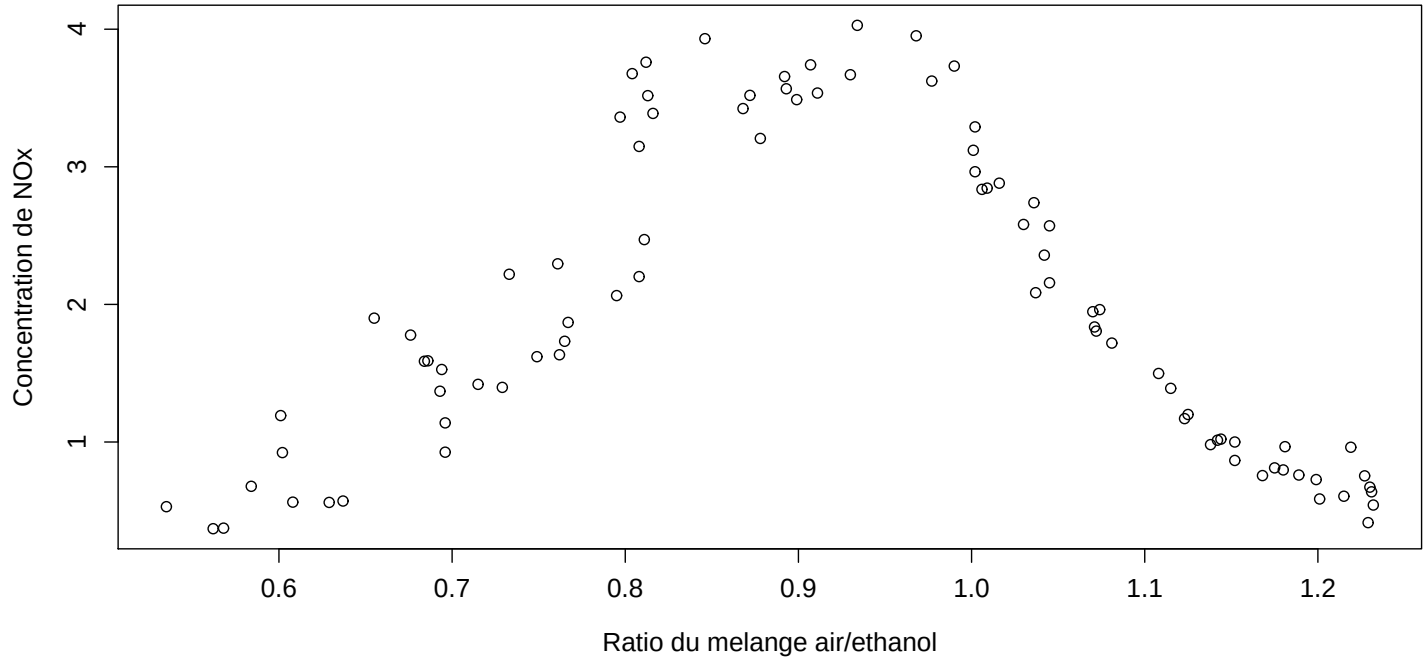

Figure 1: *Émission d'oxide nitrique (NO) et de dioxide de nitrogène (NO2) en fonction du mélange air/éthanol.*

□ Quelles remarques/constats pouvez vous faire à la vue de la Figure 1 ?

Les données

```
> library(SemiPar);data(ethanol)
> ethanol
     NOx    C     E
1  3.741 12.0 0.907
2  2.295 12.0 0.761
3  1.498 12.0 1.108
.
.
86 0.370 15.0 0.562
87 0.530 18.0 0.535
88 1.900 18.0 0.655
```

NOx Concentration de NO et NO2 lors de la combustion

C Ratio de compression du moteur

E Mesure de la richesse du mélange air/éthanol

Objectifs et éléments parcourus

- Modéliser la concentration de NOx en fonction de la qualité du mélange
- Ceci nous permettra de croiser les objets suivants :
  - Modèle linéaire, degré de liberté
  - Moindres carrés pénalisés, degré de liberté effectif
  - Décomposition de Demmler Reinsch
  - Validation croisée

1. Rappels sur la régression linéaire

Modèle linéaire

- La régression linéaire s'écrit

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

ou de manière plus compacte

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où $\mathbf{X}$ est une matrice de design et $\boldsymbol{\beta}$ un vecteur de paramètres à estimer.

- Souvent $\boldsymbol{\beta}$ est estimé par moindre carrés, i.e, on minimise $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Exercice 1. Montrez que si $\mathbf{X}^T\mathbf{X}$ est inversible, alors la solution des moindres carrés est $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$.

Hat matrix

- Les valeurs prédites sont alors

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

- En posant

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T,$$

on a clairement

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}.$$

- La matrice $\mathbf{H}$ est appelée la hat matrix (matrice chapeau) car elle met un chapeau sur les y_i.
- On a une prédiction linéaire (en $\mathbf{y}$).

Influence et degrés de liberté

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \qquad \Longrightarrow \qquad \hat{y}_i = \sum_{j=1}^{n} H_{i,j} y_j \ (i = 1, \ldots, n)$$

- $H_{i,i}$ quantifie donc la contribution de y_i pour sa propre estimation $\hat{y}_i$

Exercice 2. Que vaut l'influence totale $\sum_{i=1}^{n} H_{i,i}$?

2. Régression semi-paramétrique

Modèle semi-paramétrique

$$Y_i = f(x_i) + \varepsilon_i, \quad f \text{ fonction de forme inconnue.}$$

- La régression semi-paramétrique consiste à décomposer f dans une base appropriée, i.e.,

$$f(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{q} b_j(x - \kappa_j)\beta_{2+j},$$

où $b_j(\cdot)$ est la j-ème fonction de base et β_j le j-ème élément du paramètre de régression $\boldsymbol{\beta}$ et κ_j sont des noeuds.

- Dans la suite on se restreindra au cas

$$b_j(x - \kappa_j) = |x - \kappa_j|^3,$$

qui correspond aux splines cubiques.

Lien avec le modèle linéaire

- Ce modèle est étroitement lié avec le modèle linéaire que vous connaissez puisqu'il peut s'écrire

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

avec

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & |x_1 - \kappa_1|^3 & \cdots & |x_1 - \kappa_q|^3 \\ 1 & x_2 & |x_2 - \kappa_1|^3 & \cdots & |x_2 - \kappa_q|^3 \\ & & & \vdots & \\ 1 & x_n & |x_n - \kappa_1|^3 & \cdots & |x_n - \kappa_q|^3 \end{bmatrix}$$

- Donc la prédiction par moindre carrés est comme précédemment

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}, \qquad \mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

Importance des noeuds

- La qualité du modèle dépend du nombre de noeuds
- Pas assez de noeuds et le modèle est peu fléxible / trop de noeuds et le modèle surajuste. Il faut donc un bon compromis !

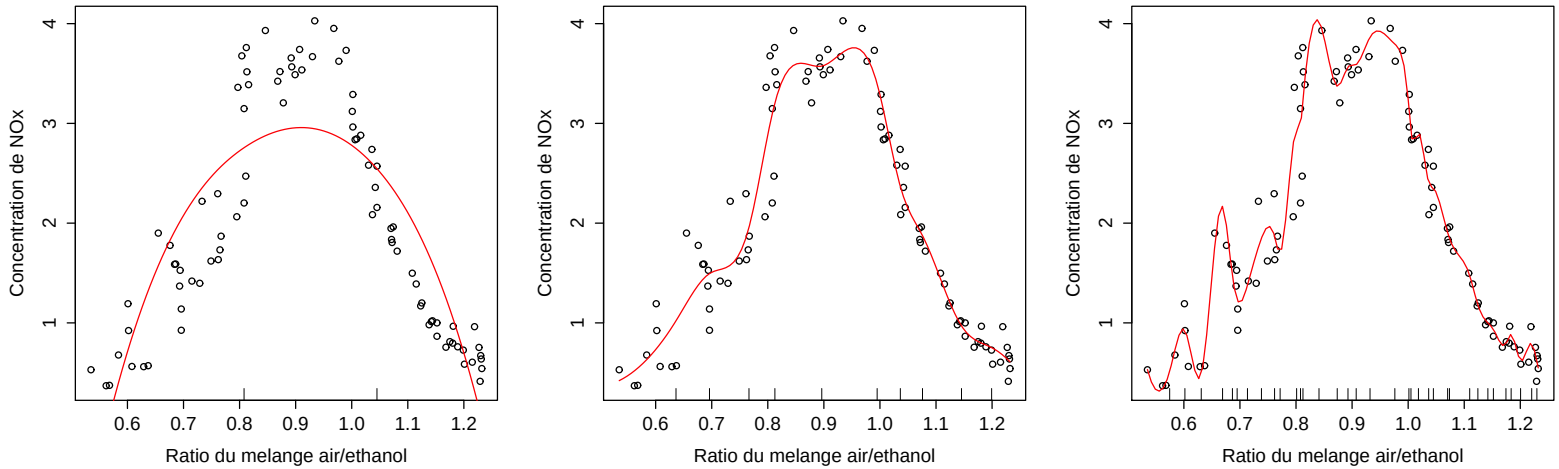

Figure 2: *Impact du nombre de noeuds sur la qualité du modèle. De gauche à droite:* $q = 2, 7, 32$. *La position des noeuds est indiquée sur l'axe des abscisses.*

Nombre de noeuds $\Rightarrow$ pénalisation

- Plutôt que de choisir le nombre et la position des noeuds $\{\kappa_j\}_{j=1,\ldots,q}$ les plus pertinents, une stratégie plus simple à mettre en oeuvre consiste à
  1. Considérer un nombre de noeuds q important
  2. Les répartir de manière pertinente (équidistants, quantiles)
  3. Ne retenir que les noeuds essentiels lors de l'ajustement.
- Pour cela nous allons donc "forcer" que $\beta_I = 0$ pour un ensemble $I \subseteq \{3, \ldots, q\}$ adéquat.

Moindres carrés pénalisés

- Plutôt que de minimiser les moindres carrés, on va donc s'intéresser au problème d'optimisation

$$\text{minimiser } \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \text{ sous la contrainte } \boldsymbol{\beta}^T\mathbf{K}\boldsymbol{\beta} \leq C,$$

  pour une constante C et une matrice $\mathbf{K}$ bien choisies.
- Si $\mathbf{K}$ est définie positive, on a $\boldsymbol{\beta}^T\mathbf{K}\boldsymbol{\beta} = \|\boldsymbol{\beta}\|_{\mathbf{K}}$ et l'on pénalise donc les $\boldsymbol{\beta}$ "trop grands" selon la norme $\|\cdot\|_{\mathbf{K}}$.
- Ce problème d'optimisation sous contrainte est équivalent à

$$\text{minimiser } \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\boldsymbol{\beta}^T\mathbf{K}\boldsymbol{\beta},$$

  pour un $\lambda > 0$ appelé le paramètre de lissage.

Exercice 3. Montrez que $\hat{\boldsymbol{\beta}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T\mathbf{y}$.

Smoothing matrix

- Les valeurs prédites sont alors

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T\mathbf{y}$$

- En posant

$$\mathbf{H}_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T,$$

on a clairement

$$\hat{\mathbf{y}} = \mathbf{H}_\lambda\mathbf{y}.$$

- La matrice $\mathbf{H}_\lambda$ est appelée la smoothing matrix (matrice lissante) et est similaire à la hat matrix vue plus haut
- On a toujours une prédiction linéaire (en $\mathbf{y}$) mais dépendant de λ.

Degrés de liberté effectif

Définition 1. Par analogie avec le modèle linéaire, on définit le degré de liberté effectif par

$$\mathrm{tr}\{\mathbf{H}_\lambda\},$$

qui peut être interprété comme le nombre réels de paramètres présents dans le modèle.

Matrice de pénalisation $\mathbf{K}$

- Plusieurs choix sont possibles pour $\mathbf{K}$
- Par exemple $\mathbf{K} = \mathrm{Id}_{q+2}$, ou encore $\mathbf{K} = \mathbf{K}_*^T \mathbf{K}_*$ avec

$$\mathbf{K}_* = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & |\kappa_1 - \kappa_1|^{3/2} & \cdots & |\kappa_1 - \kappa_q|^{3/2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & |\kappa_q - \kappa_1|^{3/2} & \cdots & |\kappa_q - \kappa_q|^{3/2} \end{bmatrix}$$

- Ce dernier choix est particulièrement intéressant pour des raisons théoriques que nous n'aborderons pas... On le considéra donc dans suite

Impact du paramètre de lissage λ

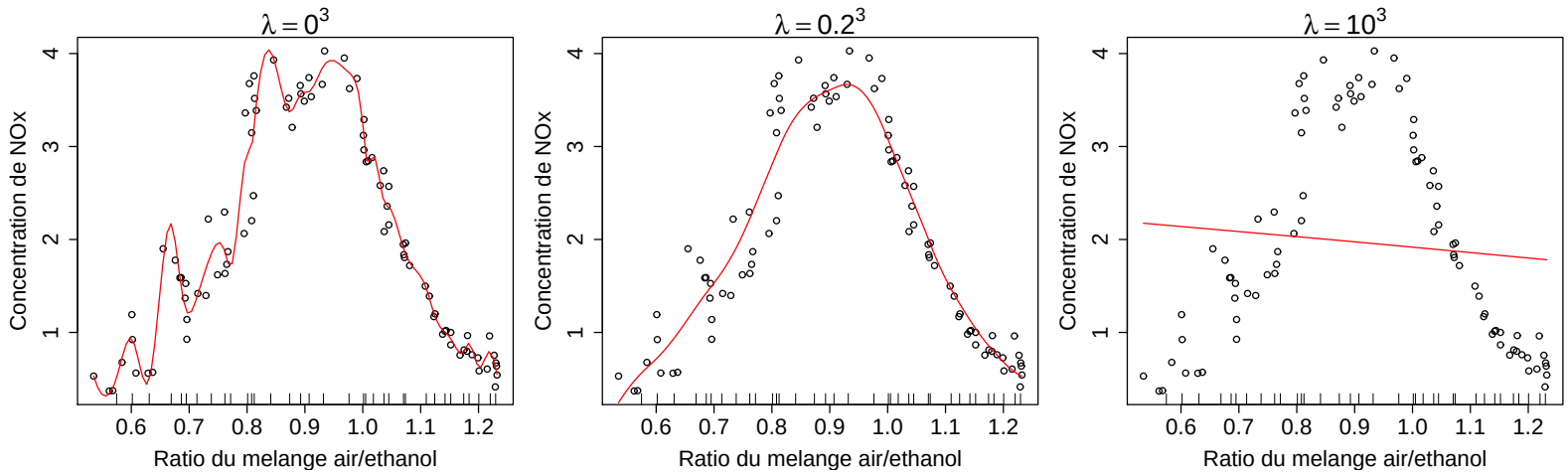

Figure 3: *Effet du paramètre de lissage sur le modèle avec de gauche à droite* $\lambda = 0, 0.2^3, 10^3$.

- Si $\lambda = 0$ alors le problème est non contraint : surajustement
- Si $\lambda \to \infty$ alors modèle linéaire simple : peu flexible
- Il faut donc choisir le λ "optimal"

Qu'est ce qu'un bon modèle selon vous ?

Qu'est ce qu'un bon modèle selon vous ?

- Prédit bien des observations futures y_+, i.e., $\{y_+ - \hat{f}(x_+; \lambda)\}^2$ petit
- Peu variable, i.e., les prédictions ne sont pas largement influencées par un nombre réduit d'observations
- Une manière de quantifier ces deux points est connue sous le nom de validation croisée

$$\mathrm{CV}(\lambda) = \sum_{i=1}^{n} \left\{ y_i - \hat{f}_{-i}(x_i; \lambda) \right\}^2,$$

où $\hat{f}_{-i}$ corresponds à l'estimation semi-paramétrique de y_i estimée sans l'aide de la i-ème observation (x_i, y_i).

- On cherchera donc λ_* minimisant $\mathrm{CV}(\lambda)$

Une formule bien utile !!!

- On a

$$\mathsf{CV}(\lambda) = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - \mathbf{H}_{\lambda,ii}} \right)^2,$$

où $\mathbf{H}_{\lambda,ii}$ est le i-ème élément diagonal de la matrice $\mathbf{H}_\lambda$ et $\hat{y}_i$ est la prédiction de la i-ème observation du modèle ajusté avec toutes les observations.

- Cette formule est bien plus pratique car elle évite d'ajuster n modèles semi-paramétriques.
- Cela dit il faut calculer $\mathbf{H}_\lambda$ (donc inverser une matrice) pour chaque valeur de λ...

Décomposition de Demmler Reinsch

- A partir des décompositions

$$\mathbf{X^T X} = \mathbf{R}^T\mathbf{R}, \qquad \mathbf{R}^{-T}\mathbf{K}\mathbf{R}^{-1} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T,$$

où $\mathbf{R}$ est une matrice inversible et $\mathbf{U}$ orthogonale, il n'est pas trop dur de montrer (algèbre linéaire simple) que

$$\mathbf{H}_\lambda = \mathbf{A}(1+\lambda\mathbf{\Lambda})^{-1}\mathbf{A}^T,$$

avec $\mathbf{A} = \mathbf{X}\mathbf{R}^{-1}\mathbf{U}$.

- Ainsi il suffira juste d'inverser $(1+\lambda\mathbf{\Lambda})$ pour chaque λ—ce qui est facile non ?

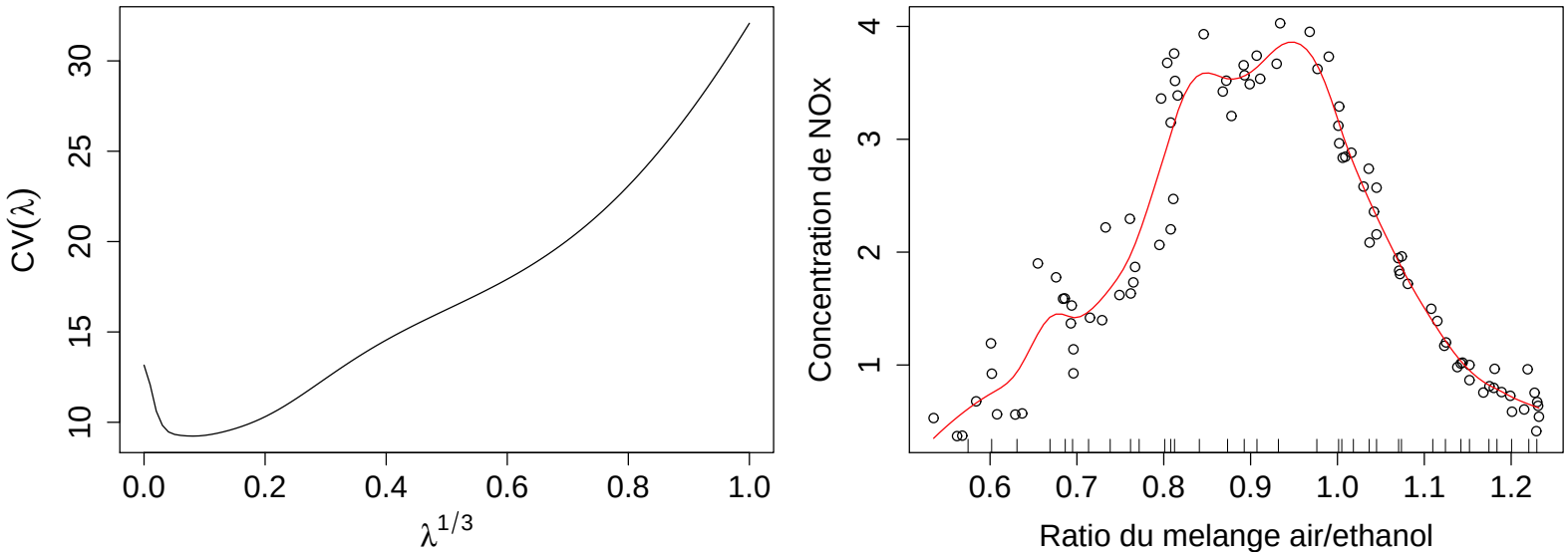

Figure 4: *Évolution de CV(λ) en fonction de λ pour les données d'émission de NOx et prédiction associé au meilleur λ.*

- Le "λ optimal" est obtenu à 0.08^3
- Le degré de liberté effectif est de 14.11 au lieu des $2 + 35$ paramètres présents...

3. Ce que nous ne verrons pas

Pour les personnes intéressées

Ceci était juste une petite initiation aux modèles semi-paramétriques. En particulier nous n'avons pas vu les éléments suivants

- GCV (generalized cross validation)
- Bandes de variabilités
- Différentes bases (thin plate, B-splines, . . .)

4. Les mains dans le cambouis

De nouvelles données

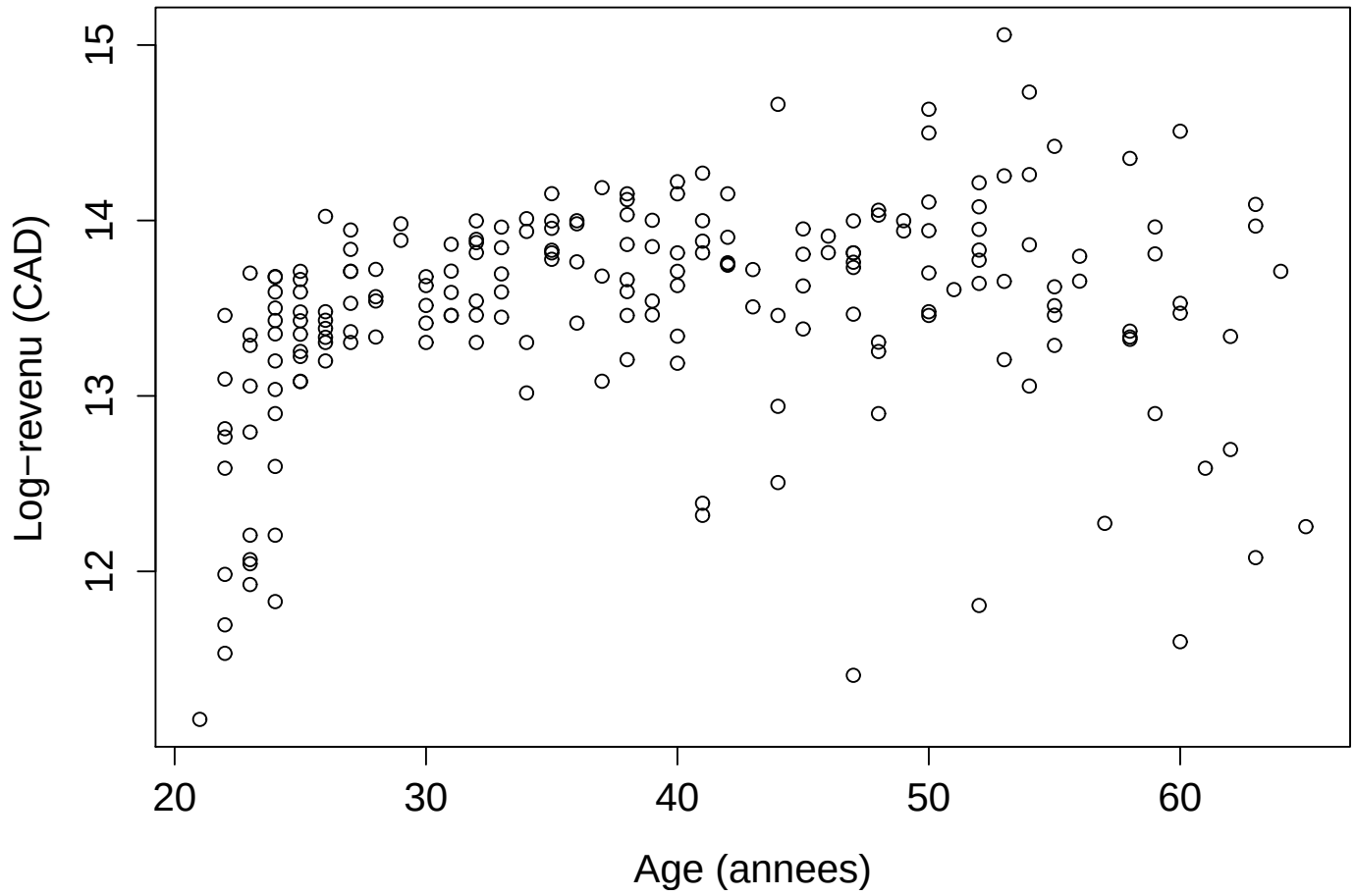

Figure 5: *Evolution du logarithme du revenu annuel ($ CAD) en fonction de l'âge. Données* `age.income` *du package* `SemiPar`.

Ce que j'aimerais que vous fassiez

- ☐ Faire une fonction `R` qui lisse nos données précédentes—le paramètre de lissage étant fixé par l'utilisateur.
- ☐ Faire une fonction `R` trouvant le paramètre de lissage optimal selon le critère CV
- ☐ Représenter la modèle semi-paramétrique ainsi obtenu
- ☐ Représenter la dérivée du modèle estimé
- ☐ Faire vos interprétations