1 Le modèle

On a un modèle génératif M complexe qui se décompose en trois morceaux $M = (M_1, M_2, M_3)$ comme suit :

$$\begin{aligned} M_1 : \quad & Y|X^{I_1}, X^{I_2}, Z \sim X^{I_1}A_{I_1} + X^{I_2}A_{I_2} + \varepsilon & (1)\\ M_2 : \quad & X^{I_2}|X^{I_1}, Z \sim X^{I_1}B_{I_1}^{I_2} + \varepsilon_{X^{I_2}} & (2)\\ M_3 : \quad & X^{I_1}|Z \sim GM(\alpha) & (3) \end{aligned}$$

On pose $\theta = (\alpha, \Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})$ Pour un Z donné on a donc

$$\begin{aligned} P_{M,Z}(Y, X^{I_1}, X^{I_2}|\theta) &= P_{M,Z}(X^{I_1}|\theta)P_{M,Z}(X^{I_2}|X^{I_1};\theta)P_{M,Z}(Y|X^{I_2}, X^{I_1};\theta) & (4)\\ &= P_{M_3,Z}(X^{I_1}|\alpha)P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})P_{M_1,M_2,Z}(Y|X^{I_2}, X^{I_1};\Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2}) & (5) \end{aligned}$$

$P_{M_3,Z}(X^{I_1}|\alpha)$ vit tout seul et est estimé par Mixmod. On sait également que

$$\begin{aligned} P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2}) &= P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma_{X^{I_2}}, B_{I_1}^{I_2}) \text{ et} & (6)\\ P_{M_1,Z}(Y|X^{I_2}, X^{I_1};\Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2}) &= P_{M_1,Z}(Y|X^{I_2}, X^{I_1};\Sigma, A_{I_1}, A_{I_2}) & (7) \end{aligned}$$

De manière générale les gens s'arrêtent là, et obtiennent le modèle complet... qui ne tient pas compte de la structure. Mais nous, nous ne sommes pas juste sous M_2 pour les sous-régression ni juste sous M_1 pour Y. Donc la dépendance persiste (c'est ce qui permet que la contrainte aie une utilité, sinon elle serait inutile).

Du coup, il faut définir proprement $P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})$ et $P_{M_1,M_2,Z}(Y|X^{I_2}, X^{I_1};\Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})$. Or il semble que ces vraisemblances ne soient pas explicites.

Vu que $P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\theta) \neq P_{M_2,Z}(X^{I_2}|X^{I_1};\theta)$ et $P_{M_1,M_2,Z}(Y|X^{I_2}, X^{I_1};\theta) \neq P_{M_1,Z}(Y|X^{I_2}, X^{I_1};\theta)$, on se limite donc à une décomposition partielle de la vraisemblance et on obtient :

$$P_{M,Z}(Y, X^{I_1}, X^{I_2}|\theta) = P_{M_3,Z}(X^{I_1}|\alpha)P_{M_1,M_2,Z}(Y, X^{I_2}|X^{I_1};\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}) \quad (8)$$

Le second membre suit une loi normale mulivariée :

$$(Y, X^{I_2}|X^{I_1}) \sim \mathcal{N}\left(\begin{pmatrix} X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}; \begin{pmatrix} \sigma_Y^2 + \sum_{i\in I_2}(\sigma_i^2 a_i^2) & \dots & a_j\sigma_j^2 & \dots \\ \vdots & \ddots & 0 & 0 \\ a_j\sigma_j^2 & 0 & \sigma_j^2 & 0 \\ \vdots & 0 & 0 & \ddots \end{pmatrix}\right) = \mathcal{N}\left(\begin{pmatrix} X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}; \begin{pmatrix} \sigma_Y^2 + \sum_{i\in I_2}(\sigma_i^2 a_i^2) & \dots a_j\sigma_j^2 \dots \\ \vdots & \\ a_j\sigma_j^2 & \Sigma_X^{I_2} \\ \vdots & \end{pmatrix}\right) \quad (9)$$

On note $\bar{\Sigma}$ la matrice de variance-covariance correspondante Avec par conséquent une vraisemblance de la forme :

$$P_{M_1,M_2,Z}(Y, X^{I_2}|X^{I_1}; \Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}) = \mathcal{L}_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2}|X^{I_1}) \tag{10}$$

$$= \frac{1}{(2\pi)^{\frac{p_2+1}{2}}|\bar{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}\right) \tag{11}$$

avec $|\bar{\Sigma}|$ le déterminant de $\bar{\Sigma}$.

On passe à la log-vraisemblance :

$$L_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}; Y, X^{I_2}|X^{I_1}) = -\frac{1}{2}\left[(p_2+1)\ln(2\pi) + \ln(|\bar{\Sigma}|) + \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}\right] \tag{12}$$

2 estimation

On veut maximiser la vraisemblance sous la contrainte $A_{I_1} + B_{I_1}^{I_2}A_{I_2} = \hat{\hat{A}}_{I_1}$ où $\hat{\hat{A}}_{I_1}$ a été préalablement calculé lors de l'étape explicative. Cette contrainte n'est pas linéaire mais devient linéaire si on se limite à A ou à B. On va donc procéder par optimisation alternée, en maximisant la vraisemblance pour un B fixé puis en recommençant avec le A obtenu, et ainsi de suite.La contrainte étant d'égalité, on utilisera le théorème de Lagrange. La partie de la vraisemblance relative à X^{I_1} ne dépend pas de la contrainte ni de l'autre membre et on la maximise donc en amont par Mixmod.

On commence par expliciter la matrice $\bar{\Sigma}^{-1}$

$$\bar{\Sigma} = \begin{pmatrix} E & C^t \\ C & \Sigma_{X^{I_2}} \end{pmatrix} \tag{13}$$

$$E = \sigma_Y^2 + \sum_{i \in I_2}(\sigma_i^2 A_i^2) = \sigma_Y^2 + A_{I_2}^t \Sigma_{X^{I_2}} A_{I_2} \text{ scalaire} \tag{14}$$

$$C = \begin{pmatrix} \vdots \\ a_i\sigma_i^2 \\ \vdots \end{pmatrix} = \Sigma_{X^{I_2}} A_{I_2} \text{ de taille } p_2 \times 1 \tag{15}$$

$$C^t = A_{I_2}^t \Sigma_{X^{I_2}}^t = A_{I_2}^t \Sigma_{X^{I_2}} \text{ de taille } 1 \times p_2 \tag{16}$$

On a donc à l'aide du complément de Schur :

$$\bar{\Sigma}^{-1} = \begin{pmatrix} [E - C^t\Sigma_{X^{I_2}}^{-1}C]^{-1} & . \\ -\Sigma_{X^{I_2}}^{-1}C(E - C^t\Sigma_{X^{I_2}}^{-1}C)^{-1} & [\Sigma_{X^{I_2}}^{-1} + \Sigma_{X^{I_2}}^{-1}C(E - C^t\Sigma_{X^{I_2}}^{-1}C)^{-1}C^t\Sigma_{X^{I_2}}^{-1}] \end{pmatrix} \text{ matrice symétrique} \tag{17}$$

Le complément de Schur est récurrent, on le note $S = E - C^t \Sigma_{X^{I_2}}^{-1} C$. Après calculs on obtient $S^{-1} = \frac{1}{\sigma_Y^2}$ On obtient alors :

$$-\Sigma_{X^{I_2}}^{-1} C (E - C^t \Sigma_{X^{I_2}}^{-1} C)^{-1} = -\Sigma_{X^{I_2}}^{-1} C S^{-1} = -\Sigma_{X^{I_2}}^{-1} \Sigma_{X^{I_2}} A_{I_2} S^{-1} = -\frac{1}{\sigma_Y^2} A_{I_2} \text{ de taille } p_2 \times 1 \tag{18}$$

et enfin :

$$[\Sigma_{X^{I_2}}^{-1} + \Sigma_{X^{I_2}}^{-1} C (E - C^t \Sigma_{X^{I_2}}^{-1} C)^{-1} C^t \Sigma_{X^{I_2}}^{-1}] = \Sigma_{X^{I_2}}^{-1} + \frac{1}{\sigma_Y^2} A_{I_2} A_{I_2}^t \Sigma_{X^{I_2}} \Sigma_{X^{I_2}}^{-1} = \Sigma_{X^{I_2}}^{-1} + \frac{1}{\sigma_Y^2} A_{I_2} A_{I_2}^t \text{ matrice symétrique pleine} \tag{19}$$

On a donc :

$$\bar{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{\sigma_Y^2} & -\frac{1}{\sigma_Y^2} A_{I_2}^t \\ -\frac{1}{\sigma_Y^2} A_{I_2} & [\Sigma_{X^{I_2}}^{-1} + \frac{1}{\sigma_Y^2} A_{I_2} A_{I_2}^t] \end{pmatrix} \text{ matrice symétrique pleine} \tag{20}$$

On note $\Sigma_{X^{I_2}(-k)}$ la matrice $\Sigma_X^{I_2}$ dépourvue de la ligne$k$ et de la colonne k. On note $\bar{\Sigma}(-1,-k)$ la matrice $\bar{\Sigma}$ dépourvue de sa prémière colonne et de sa ligne k. On a alors

$$\begin{aligned}
|\bar{\Sigma}| &= (\sigma_Y^2 + A_{I_2}^t \Sigma_{X^{I_2}} A_{I_2}) |\Sigma_X^{I_2}| + \sum_{k=1}^{p_2} (-1)^k A_{I_2(k)} \sigma_{I_2(k)}^2 |\bar{\Sigma}(-1, -(k+1))| \text{ développement 1 colonne} && (21) \\
&= (\sigma_Y^2 + A_{I_2}^t \Sigma_{X^{I_2}} A_{I_2}) |\Sigma_X^{I_2}| + \sum_{k=1}^{p_2} (-1)^k A_{I_2(k)} \sigma_{I_2(k)}^2 [(-1)^{k+1} A_{I_2(k)} \sigma_{I_2(k)}^2 \frac{1}{\sigma_{I_2(k)}^2} \prod_{j \in I_2} \sigma_j^2] \text{ developpements k colonne} && (22) \\
&= \sigma_Y^2 \prod_{j \in I_2} \sigma_j^2 + \sum_{k=1}^{p_2} A_{I_2(k)}^2 \sigma_{I_2(k)}^2 \prod_{j \in I_2} \sigma_j^2 - \sum_{k=1}^{p_2} A_{I_2(k)}^2 \sigma_{I_2(k)}^2 \prod_{j \in I_2} \sigma_j^2 && (23) \\
&= \sigma_Y^2 \prod_{j \in I_2} \sigma_j^2 && (24)
\end{aligned}$$

On peut donc calculer plus finement la vraisemblance :

$$\begin{aligned}
L_{M,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2} | X^{I_1}) &= -\frac{1}{2} \left[(p_2+1) \ln(2\pi) + \ln(|\bar{\Sigma}|) + \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) \\ (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) \\ (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t \end{pmatrix} \right] && (25) \\
&= \ldots && (26) \\
&= -\frac{1}{2} \left[(p_2+1) \ln(2\pi) + \ln(|\bar{\Sigma}|) + (X^{I_2} - X^{I_1} B_{I_1}^{I_2}) \Sigma_{X^{I_2}}^{-1} (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t + \frac{1}{\sigma_Y^2} (Y - XA)^2 \right] && (27)
\end{aligned}$$

2.1 initialisation

On commence par choisir une valeur de B. On pourrait en prendre une arbitraire, mais on choisit de commencer par le B obtenu par estimation indépendante des sous-régressions (par maximisation de $P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma_{X^{I_2}},B_{I_1}^{I_2})$). Cela peut se faire par moindres carrés classiques appliqués de manière indépendante à chacune des sous-régressions.

2.2 étape A

pour cette étape, les paramètres $\Sigma_{X^{I_2}}$ et $B_{I_1}^{I_2}$ sont fixés. On se ramène à la maximisation de

$$\begin{aligned} \max_A f(A) &= \max_A L_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2}|X^{I_1}) && (28)\\ s.c.\ \ A_{I_1} + B_{I_1}^{I_2}A_{I_2} - \hat{\hat{A}}_{I_1} &= 0 && (29)\end{aligned}$$

En notant $\psi(A) = A_{I_1} + B_{I_1}^{I_2}A_{I_2} - \hat{\hat{A}}_{I_1} = B_{I_1}A - \hat{\hat{A}}_{I_1}$ la contrainte, le théorème de Lagrange permet de dire qu'il existe un unique λ pour lequel résoudre notre maximisation sous contrainte revient à résoudre :

$$g(A,\lambda) = \bigtriangledown f(A) + \lambda \bigtriangledown \psi(A) = 0 \tag{30}$$

pour $i \in I$:

$$\begin{aligned} \frac{\partial g(A,\lambda)}{\partial A_i} &= \frac{\partial}{\partial A_i}\left(-\frac{1}{2\sigma_Y^2}(Y-XA)^2 + \lambda(B_{I_1}A - \hat{\hat{A}}_{I_1})\right) && (31)\\ &= -\frac{1}{2\sigma_Y^2}\left(-2YX^i + 2X^iXA\right) + \lambda B_{I_1}^i && (32)\\ &= \frac{1}{\sigma_Y^2}X^i(Y-XA) + \lambda B_{I_1}^i && (33)\end{aligned}$$

On a pour $i \in I_1$: $\lambda B_{I_1}^i = \lambda_i$ et pour $i \in I_2$: $\lambda B_{I_1}^i = \sum_{j\in I_1}\lambda_j B_{j,i}$
et $\forall i \in I_1$ on a :

$$\frac{\partial g(A,\lambda)}{\partial \lambda_i} = \frac{\partial \lambda\psi(A)}{\partial \lambda_i} = (B_iA - \hat{\hat{A}}_i) = A_i + \sum_{j\in I_2} B_{i,j}A_j - \hat{\hat{A}}_i \tag{34}$$

On a donc un système à résoudre de la forme (individus iid+log-vraisemblance)

$$\begin{aligned} A_{I_1} + B_{I_1}^{I_2}A_{I_2} - \hat{\hat{A}}_{I_1} &= 0 && (35)\\ \frac{1}{\sigma_Y^2}(Y - X^{I_1}A_{I_1} - X^{I_2}A_{I_2})^tX^{I_1} + \lambda &= 0 && (36)\\ \frac{1}{\sigma_Y^2}(Y - X^{I_1}A_{I_1} - X^{I_2}A_{I_2})^tX^{I_2} + \lambda B_{I_1}^{I_2} &= 0 && (37)\end{aligned}$$

ce système s'écrit également :

$$\begin{aligned}
A_{I_1} &= \hat{\hat{A}}_{I_1} - B_{I_1}^{I_2} A_{I_2} && (38)\\
\frac{1}{\sigma_Y^2}(Y - X^{I_1}(\hat{\hat{A}}_{I_1} - B_{I_1}^{I_2} A_{I_2}) - X^{I_2} A_{I_2})^t X^{I_1} + \lambda &= 0 && (39)\\
\frac{1}{\sigma_Y^2}(Y - X^{I_1}(\hat{\hat{A}}_{I_1} - B_{I_1}^{I_2} A_{I_2}) - X^{I_2} A_{I_2})^t X^{I_2} + \lambda B_{I_1}^{I_2} &= 0 && (40)
\end{aligned}$$

On en déduit :

$$\begin{aligned}
\lambda &= -\frac{1}{\sigma_Y^2}(Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2})^t X^{I_1} && (41)\\
\text{d'où} \quad 0 &= (Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2})^t X^{I_2} - \sigma_Y^2 \frac{1}{\sigma_Y^2}(Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2})^t X^{I_1} && (42)\\
0 &= (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (Y - X^{I_1} \hat{\hat{A}}_{I_1} - (X^{I_2} - X^{I_1} B_{I_1}^{I_2}) A_{I_2}) && (43)\\
\text{d'où} \quad (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (Y - X^{I_1} \hat{\hat{A}}_{I_1}) &= (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (X^{I_2} - X^{I_1} B_{I_1}^{I_2}) A_{I_2} && (44)\\
\text{et enfin} \quad \hat{A}_{I_2} &= \left((X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (X^{I_2} - X^{I_1} B_{I_1}^{I_2})\right)^{-1} (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (Y - X^{I_1} \hat{\hat{A}}_{I_1}) && (45)\\
\text{puis} \ \hat{A}_{I_1} &= \hat{\hat{A}}_{I_1} - B_{I_1}^{I_2} A_{I_2} && (46)
\end{aligned}$$

On remarque que $(X^{I_2} - X^{I_1} B_{I_1}^{I_2}) = \varepsilon_{X^{I_2}}$ résidu des sous-régressions et que $Y - X^{I_1} \hat{\hat{A}}_{I_1} = \tilde{\varepsilon}$ résidu du modèle explicatif. On se contente donc en fait d'estimer les résidus du modèle explicatif.

$$Y - X^{I_1} \hat{\hat{A}}_{I_1} = \varepsilon_{X^{I_2}} A_{I_2} + \varepsilon_Y \tag{47}$$

De ce fait, on peut utiliser tout estimateur et méthode de sélection propre à la régression linéaire en l'appliquant à ces données modifiées.

2.3 étape B

Pour cette étape, les paramètres Σ, A_{I_1} et A_{I_2} sont fixés (ici aussi, la connaissance de A permet d'obtenir le Σ associé). Comme pour l'étape A, on va résoudre :

$$\bigtriangledown f(B) + \lambda \bigtriangledown \psi(B) = 0 \tag{48}$$

où $f(B) = L_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2}|X^{I_1})$

$$g(B, \lambda) = \bigtriangledown f(B) + \lambda \bigtriangledown \psi(B) = 0 \tag{49}$$

On note $I_1^j = \{i \in I_1 | Z_{i,j} \neq 0\}$. pour $i \in I_1, j \in I_2$ et $Z_{i,j} \neq 0$ (i.e. : $i \in I_1^j$) :

$$\begin{aligned}
\frac{\partial g(B,\lambda)}{\partial B_{i,j}} &= \frac{\partial}{\partial B_{i,j}}\left(-\frac{1}{2}[(X^{I_2} - X^{I_1}B_{I_1}^{I_2})\Sigma_{X^{I_2}}^{-1}(X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t]\right) + \frac{\partial}{\partial B_{i,j}}\left(\lambda(B_{I_1}A - \hat{\hat{A}})\right) && (50)\\
&= -\frac{1}{2}\frac{\partial}{\partial B_{i,j}}\left(\sum_{l\in I_2}\frac{1}{\sigma_l^2}(X^l - \sum_{k\in I_1} X^k B_{k,l})^2\right) + \lambda_i A_j && (51)\\
&= \frac{X^i}{\sigma_j^2}(X^j - \sum_{k\in I_1} X^k B_{k,j}) + \lambda_i A_j && (52)\\
&= \frac{X^i}{\sigma_j^2}(X^j - \sum_{k\in I_1^j} X^k B_{k,j}) + \lambda_i A_j && (53)\\
\frac{\partial g(B,\lambda)}{\partial B_{I_1^j,j}} &= \frac{1}{\sigma_j^2}(X^j - X^{I_1^j}B_{I_1^j}^j)^t X^{I_1^j} + A_j\lambda_{I_1^j} && (54)
\end{aligned}$$

avec $\lambda_{I_1^j}$ qui désigne le vecteur des λ_i associés aux élements de I_1^j. et pour tout $i \in I_1$:

$$\frac{\partial g(B,\lambda)}{\partial \lambda_i} = \frac{\partial \lambda\psi(B)}{\partial \lambda_i} = (B_i A - \hat{\hat{A}}_i) = A_i + \sum_{j\in I_2} B_{i,j}A_j - \hat{\hat{A}}_i = 0 \qquad (55)$$

Et donc on obtient le système $\forall j \in I_2$:

$$\begin{aligned}
\frac{1}{\sigma_j^2}(X^j - X^{I_1^j}B_{I_1^j}^j)^t X^{I_1^j} + A_j\lambda_{I_1^j} &= 0 && (56)\\
A_{I_1^j} + \sum_{j\in I_2} A_j B_{I_1^j}^j - \hat{\hat{A}}_{I_1^j} &= 0 && (57)
\end{aligned}$$

On a donc

$$\begin{aligned}
\frac{1}{\sigma_j^2}(X^j)^t X^{I_1^j} - \frac{1}{\sigma_j^2}(B_{I_1^j}^j)^t (X^{I_1^j})^t X^{I_1^j} + A_j\lambda_{I_1^j} &= 0 && (58)\\
(X^{I_1^j})^t X^{I_1^j} B_{I_1^j}^j &= \sigma_j^2 \frac{1}{\sigma_j^2}(X^{I_1^j})^t X^j + \sigma_j^2 A_j (\lambda_{I_1^j})^t && (59)\\
B_{I_1^j}^j &= \left((X^{I_1^j})^t X^{I_1^j}\right)^{-1}\left[(X^{I_1^j})^t X^j + \sigma_j^2 A_j(\lambda_{I_1^j})^t\right] && (60)
\end{aligned}$$

On constate ici que si A_j ou σ_j est nul, alors on revient à l'estimateur des moindre carrés classiques pour la sous-régression concernée. On peut donc continuer en supposant que A_j et σ_j sont strictement non nuls (puisque sinon, on connaît les $B_{i,j}$ correspondant et donc on n'a plus besoin de les chercher).

$$B^j_{I_1^j} = \left((X^{I_1^j})^t X^{I_1^j}\right)^{-1}\left[(X^{I_1^j})^t X^j + \sigma_j^2 A_j (\lambda_{I_1^j})^t\right] \tag{61}$$

$$B^j_{I_1^j} = \left((X^{I_1^j})^t X^{I_1^j}\right)^{-1}(X^{I_1^j})^t X^j + \sigma_j^2 A_j \left((X^{I_1^j})^t X^{I_1^j}\right)^{-1}(\lambda_{I_1^j})^t \tag{62}$$

On peut donc écrire la contrainte sous la forme

$$A_{I_1^j} + \sum_{j \in I_2}\left((X^{I_1^j})^t X^{I_1^j}\right)^{-1}\left[(X^{I_1^j})^t X^j + \sigma_j^2 A_j (\lambda_{I_1^j})^t\right] A_j = \hat{\hat{A}}_{I_1^j} \tag{63}$$

d'où

$$\sum_{j \in I_2}\left((X^{I_1^j})^t X^{I_1^j}\right)^{-1}\left[(X^{I_1^j})^t X^j + \sigma_j^2 A_j (\lambda_{I_1^j})^t\right] A_j = \hat{\hat{A}}_{I_1} - A_{I_1} \tag{64}$$

$$\tag{65}$$