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 :

$$M_1: Y|X^{I_1}, X^{I_2}, Z \sim X^{I_1}A_{I_1} + X^{I_2}A_{I_2} + \varepsilon$$
 (1)

$$M_2: X^{I_2}|X^{I_1}, Z \sim X^{I_1}B_{I_1}^{I_2} + \varepsilon_{X^{I_2}}$$
 (2)

$$M_3: \qquad X^{I_1}|Z \sim GM(\alpha) \tag{3}$$

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

$$P_{M,Z}(Y,X^{I_1},X^{I_2}|\theta) = P_{M,Z}(X^{I_1}|\theta)P_{M,Z}(X^{I_2}|X^{I_1};\theta)P_{M,Z}(Y|X^{I_2},X^{I_1};\theta)$$

$$\tag{4}$$

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

$$(5)$$

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

$$P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma,\Sigma_{X^{I_2}},B^{I_2}_{I_1},A_{I_1},A_{I_2}) = P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma_{X^{I_2}},B^{I_2}_{I_1}) \text{ et}$$

$$(6)$$

$$P_{M_1,Z}(Y|X^{I_2},X^{I_1};\Sigma,\Sigma_{X^{I_2}},B_{I_1}^{I_2},A_{I_1},A_{I_2}) = P_{M_1,Z}(Y|X^{I_2},X^{I_1};\Sigma,A_{I_1},A_{I_2})$$

$$(7)$$

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 $P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\Sigma,\Sigma_{X^{I_2}},B^{I_2}_{I_1},A_{I_1},A_{I_2})$ et $P_{M_1,M_2,Z}(Y|X^{I_2},X^{I_1};\Sigma,\Sigma_{X^{I_2}},B^{I_2}_{I_1},A_{I_1},A_{I_2})$. 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 :

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

Pour alléger les notations, on va pour la suite de cette partie considérer les individus séparément (sans perte de généralité vu qu'ils sont i.i.d). On a alors Y scalaire et X vecteur de taille p. Le second membre suit une loi normale mulivariée :

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

$$(9)$$

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

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

$$(10)$$

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

$$(11)$$

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}|) + \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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{array} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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} + B_{I_{1}}^{I_{2}}A_{I_{2}} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} 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} + B_{I_{1}}^{I_{2}}A_{I_{2}} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} Y - X^{I_{1}}(A_{I_{1}} + B_{I_{1}}^{I_{2}}A_{I_{2}} \\ (X^{I_{1}} - X^{I_{2}}B_{I_{1}}^{I_{2}})^{t} + B_{I_{1}}^{I_{2}}A_{I_{2}} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} Y - X^{I_{1}}(A_{I_{1}} + B_{I_{1}}^{I_{2}}A_{I_{2}} \\ (X^{I_{1}} - X^{I_{2}}B_{I_{1}}^{I_{2}})^{t} \right)^{t} \bar{\Sigma}^{-1} \left(\begin{array}{c} Y - X^{I_{1}}(A_{I_{1}} + B_{$$

2 estimation

On veut maximiser la vraisemblance sous la contrainte $A_{I_1} + B_{I_1}^{I_2} A_{I_2} = \hat{A}_{I_1}$ où \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 $\bar{\Sigma}^{-1}$

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

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

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

$$(15)$$

$$C^{t} = A_{I_{2}}^{t} \Sigma_{X^{I_{2}}}^{t} = A_{I_{2}}^{t} \Sigma_{X^{I_{2}}} \text{ de taille } 1 \times p_{2}$$
(16)

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

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

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

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

et enfin:

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

On a donc:

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

On note $\Sigma_{X^{I_2}(-k)}$ la matrice $\Sigma_X^{I_2}$ dépourvue de la lignek 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

$$|\bar{\Sigma}| = (\sigma_Y^2 + A_{I_2}^t \Sigma_{X^{I_2}} A_{I_2}) |\Sigma_X^{I_2}| + \sum_{k=1}^{p_2} (-1)^k A_{I_2(k)} \sigma_{I_2(k)}^2 |\bar{\Sigma}(-1, -(k+1))| \text{ développement 1 colonne}$$

$$= (\sigma_Y^2 + A_{I_2}^t \Sigma_{X^{I_2}} A_{I_2}) |\Sigma_X^{I_2}| + \sum_{k=1}^{p_2} (-1)^k A_{I_2(k)} \sigma_{I_2(k)}^2 [(-1)^{k+1} A_{I_2(k)} \sigma_{I_2(k)}^2 \prod_{i \in I_2} \sigma_j^2] \text{ developpements k colonne}$$

$$(21)$$

$$= \sigma_Y^2 \prod_{j \in I_2} \sigma_j^2 + \sum_{k=1}^{p_2} A_{I_2(k)}^2 \sigma_{I_2(k)}^2 \prod_{j \in I_2} \sigma_j^2 - \sum_{k=1}^{p_2} A_{I_2(k)}^2 \sigma_{I_2(k)}^2 \prod_{j \in I_2} \sigma_j^2$$

$$= \sigma_Y^2 \prod_{j \in I_2} \sigma_j^2$$
(23)

On peut donc calculer plus finement la log-vraisemblance :

$$L_{M,Z}(\Sigma, A_{I_{1}}, A_{I_{2}}, \Sigma_{X^{I_{2}}}, B_{I_{1}}^{I_{2}}); Y, X^{I_{2}}|X^{I_{1}}) = -\frac{1}{2} \left[(p_{2}+1)\ln(2\pi) + \ln(|\bar{\Sigma}|) + \left(\frac{Y - X^{I_{1}}(A_{I_{1}} + B_{I_{1}}^{I_{2}}A_{I_{2}})}{(X^{I_{2}} - X^{I_{1}}B_{I_{1}}^{I_{2}})^{t}} \right)^{t} \bar{\Sigma}^{-1} \left(\frac{Y - X^{I_{1}}(A_{I_{1}} + B_{I_{1}}^{I_{2}}A_{I_{2}})}{(X^{I_{2}} - X^{I_{1}}B_{I_{1}}^{I_{2}})^{t}} \right) \right] (25)$$

$$= \dots \qquad (26)$$

$$= -\frac{1}{2} \left[(p_{2}+1)\ln(2\pi) + \ln(|\bar{\Sigma}|) + (X^{I_{2}} - X^{I_{1}}B_{I_{1}}^{I_{2}})\Sigma_{X^{I_{2}}}^{-1} (X^{I_{2}} - X^{I_{1}}B_{I_{1}}^{I_{2}})^{t} + \frac{1}{\sigma_{Y}^{2}} (Y - XA)^{2} \right] \qquad (27)$$

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 $P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma_{X^{I_2}},B^{I_2}_{I_1})$). 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

$$\max_{A} f(A) = \max_{A} L_{M_1, M_2, Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}; Y, X^{I_2} | X^{I_1})$$
(28)

$$s.c. A_{I_1} + B_{I_1}^{I_2} A_{I_2} - \hat{\hat{A}}_{I_1} = 0 (29)$$

En notant $\psi(A) = A_{I_1} + B_{I_1}^{I_2} A_{I_2} - \hat{A}_{I_1} = B_{I_1} A - \hat{A}_{I_1}$ 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 :

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

pour $i \in I$:

$$\frac{\partial g(A,\lambda)}{\partial A_i} = \frac{\partial}{\partial A_i} \left(-\frac{1}{2\sigma_Y^2} (Y - XA)^t (Y - XA) + \lambda (B_{I_1} A - \hat{\tilde{A}}_{I_1}) \right)$$
(31)

$$= -\frac{1}{2\sigma_V^2} \left(-2(X^i)^t Y + 2(X^i)^t X A \right) + \lambda B_{I_1}^i$$
 (32)

$$= \frac{1}{\sigma_Y^2} (X^i)^t (Y - XA) + \lambda B_{I_1}^i \tag{33}$$

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 :

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

On revient maintenant à la notation matricielle Y vecteur de taille n et X matrice de taille $n \times p$. On a donc un système à résoudre de la forme (individus

iid+log-vraisemblance)

$$A_{I_1} + B_{I_1}^{I_2} A_{I_2} - \hat{A}_{I_1} = 0 (35)$$

$$\frac{1}{\sigma_V^2} (X^{I_1})^t (Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2}) + \lambda^t = 0$$
(36)

$$\frac{1}{\sigma_V^2} (X^{I_2})^t (Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2}) + (B_{I_1}^{I_2})^t \lambda^t = 0$$
(37)

ce système s'écrit également :

$$A_{I_1} = \hat{A}_{I_1} - B_{I_1}^{I_2} A_{I_2} \tag{38}$$

$$\frac{1}{\sigma_V^2} (X^{I_1})^t (Y - X^{I_1} (\hat{A}_{I_1} - B_{I_1}^{I_2} A_{I_2}) - X^{I_2} A_{I_2}) + \lambda^t = 0$$
(39)

$$\frac{1}{\sigma_V^2} (X^{I_2})^t (Y - X^{I_1} (\hat{A}_{I_1} - B_{I_1}^{I_2} A_{I_2}) - X^{I_2} A_{I_2}) + (B_{I_1}^{I_2})^t \lambda^t = 0$$
(40)

On en déduit :

$$\lambda^{t} = -\frac{1}{\sigma_{Y}^{2}} (X^{I_{1}})^{t} (Y - X^{I_{1}} A_{I_{1}} - X^{I_{2}} A_{I_{2}}) \tag{41}$$

$$\text{d'où} \quad 0 = \sigma_Y^2 \frac{1}{\sigma_Y^2} (X^{I_2})^t (Y - X^{I_1} (\hat{\hat{A}}_{I_1} - B_{I_1}^{I_2} A_{I_2}) - X^{I_2} A_{I_2}) - \sigma_Y^2 \frac{1}{\sigma_Y^2} (B_{I_1}^{I_2})^t (X^{I_1})^t (Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2})$$
(42)

$$0 = (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (Y - X^{I_1} \hat{A}_{I_1} - (X^{I_2} - X^{I_1} B_{I_1}^{I_2}) A_{I_2})$$

$$(43)$$

$$d'où \quad (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t(Y - X^{I_1}\hat{A}_{I_1}) = (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t(X^{I_2} - X^{I_1}B_{I_1}^{I_2})A_{I_2}$$

$$(44)$$

et enfin
$$\hat{A}_{I_2} = \left((X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (X^{I_2} - X^{I_1} B_{I_1}^{I_2}) \right)^{-1} (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t (Y - X^{I_1} \hat{A}_{I_1})$$
 (45)

$$puis \hat{A}_{I_1} = \hat{A}_{I_1} - B_{I_1}^{I_2} A_{I_2}$$
(46)

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

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

De ce fait, on peut utiliser tout estimateur et méthode de sélection propre à la régression linéaire en l'appliquant à ces données modifiées. Ce modèle est appelé modèle prédictif. L'idée est maintenant d'optimiser B sachant A pour ensuite venir recalculer ce modèle, et ainsi de suite jusqu'à convergence vers le maximum de vraisemblance de la loi jointe du modèle (M).

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 :

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

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

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

On note $I_1^j = \{i \in I_1 | Z_{i,j} \neq 0\}$. Alors $\forall j \in I_2, \forall i \in I_1^j$:

$$\frac{\partial g(B,\lambda)}{\partial B_{i,j}} = \frac{\partial}{\partial B_{i,j}} \sum_{k=1}^{n} \left(-\frac{1}{2} \left[(X_k^{I_2} - X_k^{I_1} B_{I_1}^{I_2}) \Sigma_{X^{I_2}}^{-1} (X_k^{I_2} - X_k^{I_1} B_{I_1}^{I_2})^t \right] \right) + \frac{\partial}{\partial B_{i,j}} \left(\lambda (B_{I_1} A - \hat{A}) \right)$$
(50)

$$= -\frac{1}{2} \frac{\partial}{\partial B_{i,j}} \left(\sum_{l \in I_2} \frac{1}{\sigma_l^2} (X^l - \sum_{k \in I_1} X^k B_{k,l})^t (X^l - \sum_{k \in I_1} X^k B_{k,l}) \right) + \lambda_i A_j$$
 (51)

$$= \frac{1}{\sigma_j^2} (X^j - \sum_{k \in I_1} X^k B_{k,j})^t X^i + \lambda_i A_j$$
 (52)

$$= \frac{1}{\sigma_j^2} (X^j - \sum_{k \in I_j^j} X^k B_{k,j})^t X^i + \lambda_i A_j$$
 (53)

$$\frac{\partial g(B,\lambda)}{\partial B_{I_1^j}^j} = \frac{1}{\sigma_j^2} (X^{I_1^j})^t (X^j - X^{I_1^j} B_{I_1^j}^j) + A_j \lambda_{I_1^j}^t$$
(54)

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

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

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

$$\frac{1}{\sigma_j^2} (X^{I_1^j})^t (X^j - X^{I_1^j} B_{I_1^j}^j) + A_j \lambda_{I_1^j}^t = 0$$
 (56)

$$A_{I_1} + \sum_{j \in I_2} A_j B_{I_1}^j - \hat{A}_{I_1} = 0 (57)$$

On a donc

$$\frac{1}{\sigma_j^2} (X^{I_1^j})^t X^j - \frac{1}{\sigma_j^2} (X^{I_1^j})^t X^{I_1^j} B_{I_1^j}^j + A_j \lambda_{I_1^j}^t = 0$$
(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)

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).

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

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

$$(62)$$

Soit J la matrice identité de taille p. On introduit une matrice de passage $J_{I_1}^{I_1^j}$ matrice identité de taille p_1 restreinte en nombre de colonnes. On peut donc écrire la contrainte sous la forme

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

$$(63)$$

d'où

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

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

$$(65)$$

$$\sum_{j \in I_2} J_{I_1}^{I_1^j} \left((X^{I_1^j})^t X^{I_1^j} \right)^{-1} \sigma_j^2 A_j^2 (\lambda_{I_1^j})^t = (\hat{A}_{I_1} - A_{I_1}) - \sum_{j \in I_2} J_{I_1}^{I_1^j} \left((X^{I_1^j})^t X^{I_1^j} \right)^{-1} (X^{I_1^j})^t X^j A_j$$
 (66)

En notation moins matricielle on a le système

$$\forall j \in I_2, \forall i \in I_1^j : \frac{1}{\sigma_j^2} (X^j - \sum_{k \in I_1^j} X^k B_{k,j})^t X^i + \lambda_i A_j = 0$$
(67)

$$\forall i \in I_1: \ A_i + \sum_{j \in I_2} A_j B_{i,j} - \hat{\hat{A}}_i = 0 \tag{68}$$

et donc

$$\forall j \in I_2, \forall i \in I_1^j : \frac{1}{\sigma_j^2} (\sum_{k \in I_1^j} X^k B_{k,j})^t X^i = \frac{1}{\sigma_j^2} (X^j)^t X^i + \lambda_i A_j$$
(69)

$$(B_{I_1^j}^j)^t (X^{I_1^j})^t X^i = (X^j)^t X^i + \sigma_j^2 \lambda_i A_j$$
(70)

En notation plus large on a le système :

$$A_{I_1} + \sum_{j \in I_2} A_j B_{I_1}^j - \hat{A}_{I_1} = 0$$

$$B \cdot Z = 0$$
(71)

$$B \cdot Z = 0 \tag{72}$$

$$\frac{1}{\sigma_j^2} (X^j - X^{I_1} B_{I_1}^j)^t X^{I_1} + A_j \lambda = 0 (73)$$

(74)

d'où

$$\frac{1}{\sigma_j^2} (X^j)^t X^{I_1} - \frac{1}{\sigma_j^2} (B_{I_1}^j)^t (X^{I_1})^t X^{I_1} + A_j \lambda_{I_1} = 0$$
(75)

$$(X^{I_1})^t X^{I_1} B^j_{I_1} = \sigma_j^2 \frac{1}{\sigma_j^2} (X^{I_1})^t X^j + \sigma_j^2 A_j \lambda^t$$
 (76)

$$B_{I_1}^j = \left((X^{I_1})^t X^{I_1} \right)^{-1} \left[(X^{I_1})^t X^j + \sigma_j^2 A_j \lambda^t \right]$$
 (77)

donc la contrainte s'écrit

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

$$(78)$$

$$\sum_{j \in I_2} \left((X^{I_1})^t X^{I_1} \right)^{-1} \left[(X^{I_1})^t X^j + \sigma_j^2 A_j \lambda^t \right] A_j = (\hat{A}_{I_1} - A_{I_1})$$
(79)

$$\sum_{j \in I_2} (X^{I_1})^t X^j A_j + \sum_{j \in I_2} \sigma_j^2 A_j^2 \lambda^t = (X^{I_1})^t X^{I_1} (\hat{\hat{A}}_{I_1} - A_{I_1})$$
(80)

$$(X^{I_1})^t X^{I_2} A_{I_2} + (A^t_{I_2} \Sigma_{X^{I_2}} A_{I_2}) \lambda^t = (X^{I_1})^t X^{I_1} (\hat{A}_{I_1} - A_{I_1})$$
(81)

$$\lambda^{t} = \frac{1}{A_{I_{2}}^{t} \sum_{X^{I_{2}}} A_{I_{2}}} \left[(X^{I_{1}})^{t} X^{I_{1}} (\hat{\tilde{A}}_{I_{1}} - A_{I_{1}}) - (X^{I_{1}})^{t} X^{I_{2}} A_{I_{2}} \right]$$
(82)

$$\lambda_{I_1^j}^t = \frac{1}{A_{I_2}^t \sum_{X^{I_2}} A_{I_2}} (X^{I_1^j})^t \left[X^{I_1} (\hat{A}_{I_1} - A_{I_1}) - X^{I_2} A_{I_2} \right]$$
 (83)

$$\lambda_i = \frac{1}{A_{I_2}^t \sum_{X^{I_2}} A_{I_2}} (X^i)^t \left[X^{I_1} (\hat{\hat{A}}_{I_1} - A_{I_1}) - X^{I_2} A_{I_2} \right]$$
 (84)