

REVUE DE STATISTIQUE APPLIQUÉE

XAVIER BRY

Estimation empirique d'un modèle à variables latentes comportant des interactions

Revue de statistique appliquée, tome 52, n° 3 (2004), p. 5-35

http://www.numdam.org/item?id=RSA_2004__52_3_5_0

© Société française de statistique, 2004, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ESTIMATION EMPIRIQUE D'UN MODÈLE À VARIABLES LATENTES COMPORTANT DES INTERACTIONS

Xavier BRY

Laboratoire LISE-CEREMADE, Université de Paris IX-Dauphine
Email : bryxavier@yahoo.fr

RÉSUMÉ

On considère un modèle linéaire comportant une ou plusieurs équations reliant des variables continues latentes, chacune de celles-ci étant sous-jacente à un groupe de variables observées. Le modèle prévoit la possibilité d'interactions deux à deux entre certaines variables explicatives latentes. On propose ici une extension de l'Analyse en Composantes Thématiques qui permet d'estimer empiriquement les variables latentes du modèle en prenant en compte leurs interactions.

Mots-clés : *Analyse en Composantes Thématiques, Approche PLS, Interactions, Moindres Carrés Partiels, Régression PLS.*

ABSTRACT

We consider a linear model consisting in one or several equations binding continuous latent variables together, each variable underlying in a group of observed variables. The model assumes that some latent explanatory variables may interact pairwise. We propose an extension of Thematic Components Analysis that takes interaction effects into account through the estimation process.

Keywords : *Interaction Effects, Partial Least Squares, PLS, Path Modelling, PLS Regression, Thematic Components Analysis.*

Notations

I_n désigne couramment la matrice identité de taille n . Lorsque la dimension de la matrice est sans ambiguïté, elle sera simplement notée I .

Les minuscules carolingiennes $a, b, c, \dots, u, v, w$ désignent des vecteurs de coefficients.

Les minuscules grecques désignent des scalaires.

X, Y, Z, T désignent des tableaux matriciels décrivant n individus (en ligne) à l'aide de variables (en colonne). Ces lettres sont indifféremment utilisées pour désigner les groupes de variables en question.

Les minuscules x, y, z, t désignent indifféremment les vecteurs-colonnes de ces matrices et les variables des groupes correspondants.

M, N et P sont des métriques, *i.e.* des matrices symétriques définies positives pondérant les variables des groupes respectifs Y, Z et T .

La matrice bloc-diagonale formée des blocs diagonaux $A, B, C \dots$ sera notée $\text{Diag}(A, B, C \dots)$.

x étant un vecteur, $\text{diag}(x)$ désignera la matrice diagonale dont les éléments diagonaux sont les composantes de x .

La juxtaposition des matrices $A, B, C \dots$ sera notée $[A, B, C \dots]$.

Le sous-espace vectoriel de $\mathbf{R}^n$ engendré par un groupe de variables X sera noté $\langle X \rangle$.

Lorsqu'un groupe de variables X est décomposé en R sous-groupes, ceux-ci seront notés $X_1, \dots, X_R$.

F, G et H désignent des facteurs construits par combinaison linéaire des variables d'un groupe à préciser.

Le projecteur sur un sous-espace E parallèlement à un supplémentaire F sera noté : Π_E^F .

Le projecteur orthogonal sur un sous-espace E sera noté Π_E .

L'orthogonalité s'entend, sauf mention explicite contraire, par rapport à la métrique canonique I .

Le produit scalaire entre deux vecteurs x et y sera noté $\langle x|y \rangle$.

On notera la colinéarité de deux vecteurs x et y : $x \propto y$.

Sigles et Acronymes

AC = Analyse Canonique

ACT = Analyse en Composantes Thématiques

ARC = Analyse en Résultantes Covariantes

ARM = Analyse des Redondances Maximales

PLS = Partial Least Squares

PLS-PM = PLS Path Modelling

1. Modèles à une équation

1.1. Présentation

• Considérons un modèle à une équation formalisant le fait que R variables latentes agissent sur une variable latente (ou observée). Chaque variable latente est sous-jacente à un groupe de variables observées possédant une unité conceptuelle forte (cf. figure 1 et [Bry 2003], [Tenenhaus 1998, 1999]).

Chaque variable latente est supposée répondre à une double contrainte :

- être représentative, *i.e.* globalement liée aux variables observées de son groupe ;
- être reliée aux autres variables latentes par un schéma multivarié conforme au modèle conceptuel.

C'est la prise en compte de cette double contrainte qui permet d'élaborer une stratégie d'estimation des variables latentes. Une variable latente sera estimée par un *facteur* optimisant un critère traduisant la double contrainte ¹.

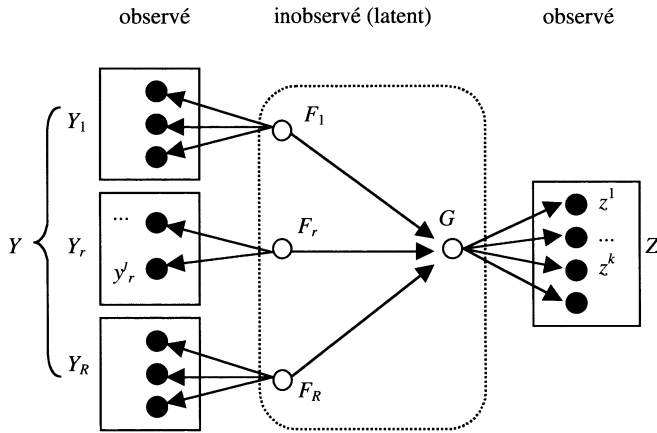


FIGURE 1

Note : un arc orienté $a \rightarrow b$ signifie : a est un déterminant de b .

• = variable observée ; ◯ = variable latente

1.2. Modèle sans interactions ; estimation par l'Analyse en Composantes Thématiques

1.2.1. Le modèle

Le modèle linéaire sans interactions correspondant à la figure 1 s'écrit :

$$G = \sum_{r=1}^R \beta_r F_r + u, \text{ où } u \text{ est un terme résiduel. Il s'agit d'estimer } G, \{F_r\}, \{\beta_r\} \text{ et } u.$$

¹ Pour alléger l'exposé, on identifiera la variable latente au facteur qui l'estime, lorsque cela ne prête pas à conséquence.

1.2.2. Préliminaires

Notion de résultante, et propriétés :

• Rappelons que la résultante d'une variable z sur un groupe X au sens d'une métrique A est notée $R_X^A z$ et est ainsi définie : $R_X^A z = XAX'z$. Elle mesure l'accord de z avec X de deux façons : sa direction donne la dimension de cet accord dans le groupe, et sa norme en mesure l'intensité (cf.[Bry 2001]).

• Si F^k est le k -ième facteur normé d'ACP de X avec la métrique A , et λ_k la valeur propre associée, on a : $R_X^A z = \sum_k \lambda_k F^k F^{k'} z = \sum_k \lambda_k \langle F^k | z \rangle F^k$. On obtient une somme des composantes de z sur la base factorielle, chaque composante étant pondérée par l'inertie captée par le facteur, *i.e.* la force structurelle de celui-ci.

• On peut plus généralement, α étant un réel ≥ 0 , définir ainsi une résultante de degré α : $R_X^{A\alpha} z = (XAX')^\alpha z$. C'est un opérateur symétrique. On a, avec les mêmes notations que précédemment $R_X^{A\alpha} z = \sum_k \lambda_k^\alpha \langle F^k | z \rangle F^k$. Cette écriture montre comment les opérateurs de résultante « attirent » la variable z vers les structures principales de X sur lesquelles elle a une composante non nulle, et ce, proportionnellement à la force de la structure et à la norme de cette composante.

• Dans le cas particulier $\alpha = 1/2$, et z étant normée, $\|R_X^{A1/2} z\|^2 = z'(XAX')z$ est l'inertie du groupe X le long de z .

• Si $A = (X'X)^{-1}$, $XAX' = \Pi_{\langle X \rangle}$. Donc, $\forall \alpha > 0$, la résultante $R_X^{A\alpha} z$ est la régression usuelle de z sur X . Elle ignore alors les structures de corrélation internes de X .

• Lorsque $A = I$ et que les variables de X sont normées, $\forall \alpha > 0$, $R_X^{A\alpha} z$ rapproche z des facteurs d'ACP normée – *i.e.* des structures de corrélation interne – de X , dans la mesure où ils sont à la fois forts et proches de z .

• On remarquera enfin que :

1) $\forall \lambda_k > 0$, $\lambda_k^{2\alpha} \xrightarrow{\alpha \rightarrow 0} 1$. Par conséquent, $R_X^{A\alpha} z \xrightarrow{\alpha \rightarrow 0} \Pi_{\langle X \rangle} z$. Lorsque α tend vers 0, les structures internes de X sont de moins en moins prises en compte, jusqu'à ne plus l'être du tout.

2) λ_β étant la plus grande valeur propre de XAX' telle que $\langle F^\beta | z \rangle \neq 0$ (on suppose λ_β d'ordre 1), on a : $R_X^{A\alpha} z = \sum_k \lambda_k^\alpha F^k F^{k'} z \approx \lambda_\beta^\alpha F^\beta F^{\beta'} z$.

La structure la plus forte avec laquelle z ait quelque chose en commun devient alors prépondérante.

Le paramètre α peut ainsi s'interpréter comme le degré de prise en compte des structures internes de X .

Recherche d'un facteur F matérialisant la liaison entre une variable z et un groupe X

• Considérons une variable observée z dépendant d'un seul groupe de déterminants X affecté de la métrique A . Pour estimer la variable latente sous-jacente à X et liée à z , on résout le programme $\mathbf{Q1}(X, A; z) : \max_{u' Au=1} \langle X Au | z \rangle$. Il a pour solution un facteur $F = X Au$ colinéaire à $R_X^A z$ (cf. [Tenenhaus 1998] [Bry 2001]).

• Dans [Bry 2001], nous avons proposé un programme alternatif à $\mathbf{Q1}$: $\mathbf{P1}(X, A; z) : \max_{\|F\|=1} \langle R_X^A F | z \rangle$, et montré que sa solution était également colinéaire à $R_X^A z$. Ce programme peut être généralisé ainsi : $\mathbf{P1}(X, A, a; z) : \max_{\|F\|=1} \langle R_X^{Aa} F | z \rangle$. Il a pour solution $F = R_X^{Aa} z$ normée.

La résultante d'une variable sur un groupe est donc solution de programmes d'optimisation simples.

Recherche des couples de facteurs matérialisant la liaison entre deux groupes

• On considère deux groupes Y et Z , rapportés aux métriques respectives M et N . On suppose que deux variables latentes, sous-jacentes à Y et Z respectivement, sont liées entre elles. Pour les estimer, on peut alors chercher deux facteurs $F = YMu$ et $G = ZNv$ solutions du programme :

$$\mathbf{Q}(Y, M; Z, N) : \max_{\substack{u' Mu=1 \\ v' Nv=1}} \langle YMu | ZNv \rangle$$

La résolution de $\mathbf{Q}$ conduit aux deux caractérisations équivalentes suivantes des facteurs solutions de rang 1 (cf. [Bry 2001]) :

$$\text{Caractérisation (a) : } (1) R_Y^M G = \sqrt{\eta} F ; \quad (1') R_Z^N F = \sqrt{\eta} G$$

$$\text{Caractérisation (b) : } (2) R_Y^M R_Z^N F = \eta F ; \quad (2') R_Z^N R_Y^M G = \eta G$$

où η est la plus grande valeur propre des opérateurs $R_Z^N R_Y^M$ et $R_Y^M R_Z^N$.

Remarque. – Lorsque $M = (Y'Y)^{-1}$ et $N = (Z'Z)^{-1}$, les équations (1) et (1') deviennent :

$$\Pi_{<Y>} G = \sqrt{\eta} F \quad \text{et} \quad \Pi_{<Z>} F = \sqrt{\eta} G$$

Elles caractérisent les facteurs d'Analyse Canonique des deux groupes.

Notations. – Le programme $\mathbf{Q}(Y, (Y'Y)^{-1}; Z, N)$ sera noté $\mathbf{Q}(<Y>; Z, N)$ pour indiquer qu'il ne s'intéresse qu'au sous-espace engendré par Y . Le programme $\mathbf{Q}(Y, (Y'Y)^{-1}; Z, (Z'Z)^{-1})$ sera noté $\mathbf{Q}(<Y>; <Z>)$.

• Nous proposons, en alternative à **Q**, le programme suivant :

$\mathbf{P}(Y, M, \alpha; Z, N, \beta) : \max_{\|F\|=1, \|G\|=1} \langle R_Y^{M\alpha} F | R_Z^{N\beta} G \rangle$ (le cas particulier $\alpha = \beta = 1$ a été introduit dans [Bry 2001] sous le nom d'Analyse en Résultantes Covariantes).

$$\text{Résolution : } L = F' R_Y^{M\alpha} R_Z^{N\beta} G - \lambda(F' F - 1) - \mu(G' G - 1)$$

$$\frac{\partial L}{\partial F} = 0 \Leftrightarrow R_Y^{M\alpha} R_Z^{N\beta} G = 2\lambda F \quad (3); \quad \frac{\partial L}{\partial G} = 0 \Leftrightarrow R_Z^{N\beta} R_Y^{M\alpha} F = 2\mu G \quad (3')$$

$F'(3) = G'(3') \Rightarrow \lambda = \mu$; on notera $\eta = 4\lambda^2$. Les équations (3) et (3') combinées donnent alors :

$$R_Y^{M\alpha} (R_Z^{N\beta})^2 R_Y^{M\alpha} F = \eta F \quad (4); \quad R_Z^{N\beta} (R_Y^{M\alpha})^2 R_Z^{N\beta} G = \eta G \quad (4')$$

Étant vecteurs propres d'une matrice symétrique, les facteurs F (resp. G) solution des conditions du 1^{er} ordre forment un système orthogonal. Le couple (F, G) associé à la plus grande valeur propre η est la solution recherchée. Le couple associé à la seconde valeur propre par ordre décroissant est la solution du même programme sous contrainte d'orthogonalité au facteur de rang 1 pour F et pour G respectivement. Etc.

On remarquera que lorsque $M = (Y'Y)^{-1}$ et $N = (Z'Z)^{-1}$, les équations (3) et (3') deviennent :

$$\Pi_{<Z>} \Pi_{<Y>} F = \sqrt{\eta} G \quad \text{et} \quad \Pi_{<Y>} \Pi_{<Z>} G = \sqrt{\eta} F$$

Comme $F \in < Y >$ et $G \in < Z >$, $\Pi_{<Y>} F = F$ et $\Pi_{<Z>} G = G$. Les équations ci-dessus sont donc celles de l'Analyse Canonique. On notera alors plus simplement le programme : $\mathbf{P}(< Y >; < Z >)$ pour indiquer que ni α et β , ni les structures de corrélation internes de Y et Z n'y jouent de rôle.

On notera aussi qu'on aboutit au même résultat lorsque α et $\beta \rightarrow 0$ quelles que soient M et N .

1.2.3. Analyse en Composantes Thématiques (ACT)

On considère ici qu'un phénomène décrit par un groupe de variables Z a pour déterminants R groupes explicatifs $Y_1, \dots, Y_r, \dots, Y_R$ (cf. fig. 1). Dans un premier temps, l'Analyse en Composantes Thématiques extrait, dans chaque groupe, un facteur représentatif tel que l'ensemble de ces facteurs s'ajuste au modèle linéaire correspondant au graphe. Seule cette première étape est développée ci-dessous. Elle équivaut à une variante de l'algorithme de PLS-Path Modelling proposé par Lohmöller ([Lohmöller 1989] [Bry 2003]). Dans un deuxième temps, l'ACT recherche dans chaque groupe le facteur orthogonal au premier qui complète le mieux le modèle ; etc. (cf. [Bry 2003]).

Notations. – On notera $F_r(k)$ (respectivement $G(k)$) la valeur courante, à l'itération k , du facteur du groupe Y_r (resp. Z). On notera en outre :

$$F(k) = \{F_r(k), 1 \leq r \leq R\} ; F_{-r}(k) = \{F_s(k), s \neq r\} ; \\ F_r = F_r(\infty) ; G = G(\infty).$$

Enfin, on notera désormais **R** l'un ou l'autre des programmes **Q** ou **P**, selon le choix effectué. Soit X un groupe pondéré par A . Si **R** = **Q**, la résultante utilisée $R_X^{A\alpha}$ sera la résultante de degré $\alpha = 1$. Si **R** = **P**, son degré sera celui qui a été choisi dans **P** pour le groupe X .

L'algorithme de l'ACT (restreinte aux facteurs de rang 1)

Itération 0 (initialisation)

– On détermine une valeur initiale $F_r(0)$ du premier facteur de chaque groupe Y_r , e.g. en normant la solution F de **R**($Y_r, M_r; Z, N$). Le groupe des facteurs obtenus est noté $F(0)$.

→ On détermine un facteur $G(0)$ dans le groupe Z en effectuant **R**($< F(0) >; Z, N$).

Itération $k+1$, $k \geq 0$:

– Pour chaque groupe Y_r la valeur courante $F_r(k+1)$ du 1^{er} facteur est la solution normée de **R1**($Y_r, M_r ; \Pi_{<Y_r>}^{<F_{-r}(k)>} G(k)$), i.e. est prise égale à $R_{Y_r}^{M_r\alpha} \Pi_{<Y_r>}^{<F_{-r}(k)>} G(k)$ normé. Pratiquement, on régresse $G(k)$ sur Y_r et $F_{-r}(k)$, on prend la composante sur $< Y_r >$ de la partie prédite, on en calcule la résultante de degré α sur Y_r , que l'on norme.

→ On détermine le facteur $G(k+1)$ dans le groupe Z en normant la solution G de : **R**($< F(k) >; Z, N$).

– Si les facteurs trouvés à l'itération $k+1$ sont suffisamment proches de ceux fournis par l'itération k , on s'arrête.

Cas où Z est réduit à une seule variable : l'algorithme ACT1

L'algorithme ACT ci-dessus doit être légèrement modifié lorsque le groupe Z est réduit à une seule variable z . Les étapes ci-dessus marquées d'une flèche (→) au lieu d'un tiret deviennent superflues, et le facteur G doit être remplacé par z dans les autres étapes. L'algorithme ainsi modifié sera noté **ACT1**. Une illustration intuitive de son étape courante est donnée figure 2.

N.B. : Lorsque **R** = **P**, on obtient une variante légèrement différente de celle proposée dans [Bry 2003], mais plus proche de la variante fondée sur **Q**, ce qui permet une présentation unifiée.

PROPRIÉTÉS . –

• Dans l'étape 1 de l'ACT1, lorsque l'on choisit d'ignorer les structures internes de chaque groupe Y_r en prenant $M_r = (Y_r' Y_r)^{-1}$, Le R^2 de la régression de z sur les facteurs courants ne peut que croître au cours de l'algorithme.

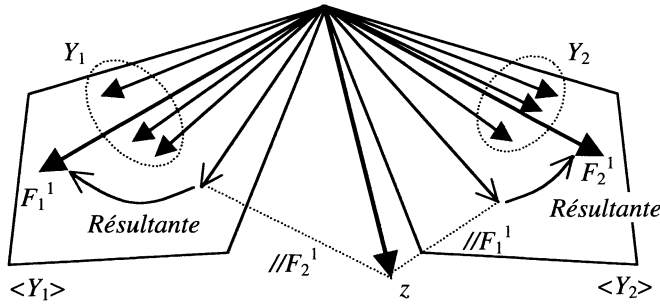


FIGURE 2
Facteurs de rang 1 de l'algorithme ACT1

Preuve. – Il est facile de voir que ce choix des matrices M_r revient à supprimer de l'algorithme tous les calculs de résultante : en effet, la résultante $R_{Y_r}^{M_r, \alpha}$ est alors le projecteur orthogonal sur $\langle Y_r \rangle$; or, ce dont on calcule la résultante sur Y_r dans les différentes étapes de l'algorithme se trouve déjà dans $\langle Y_r \rangle$. En outre, la projection, effectuée à chaque itération, de z sur $\langle Y_r \rangle$ parallèlement aux autres facteurs, est elle-même le résultat d'un programme de maximisation du R^2 sur $\langle Y_r \rangle$, ce qui garantit l'augmentation du R^2 à chaque itération de cette étape.

Note 1. – On peut montrer qu'il en va de même lors de la détermination des facteurs de rang supérieur.

Note 2. – En supprimant les calculs de résultante, on cesse de rapprocher chaque composante partielle des structures de corrélation internes de son groupe qui lui sont les plus liées. Dès que l'on réintroduit cette contrainte de rapprochement, on va contre l'optimisation du R^2 , et l'on ne peut plus garantir qu'il croisse systématiquement.

• Dans les mêmes conditions, $\hat{z}$ désignant la régression de z sur $\langle Y \rangle$, si l'on prend pour $F_r, \forall r$, la composante $\hat{z}_r$ de $\hat{z}$ sur $\langle Y_r \rangle$, on obtient un point fixe de l'algorithme.

Preuve. – $\forall r$, $\hat{z}_r$ est la projection de z sur $\langle Y_r \rangle$ parallèlement à $\langle \{\hat{z}_s, s \neq r\} \rangle$. Par conséquent, si l'on n'adjoint pas de calcul de résultante dans la détermination des facteurs, on constate que ceux-ci sont stables.

• Dans l'étape 1 de l'ACT, lorsque pour chaque groupe Y_r , on prend $M_r = (Y_r' Y_r)^{-1}$ et pour $Z, N = (Z' Z)^{-1}$, le R^2 de la régression de G sur les facteurs F courants ne peut que croître au cours de l'algorithme.

Preuve. – Il est aisé de voir que les étapes supplémentaires de ACT par rapport à ACT1, lesquelles déterminent le facteur G courant, sont dans ce cas des étapes d'Analyse Canonique (programme $R(\langle F \rangle; \langle Z \rangle)$). Dans ce cas, le facteur G courant optimise le $\cos^2$ de son angle avec le sous-espace $\langle F \rangle$, i.e. le R^2 de sa régression sur F .

1.3. Modèle avec interactions

Nous n'avons envisagé jusqu'ici que des effets propres (constants) des variables explicatives. Relâchons à présent cette contrainte en permettant à l'effet d'une variable explicative de dépendre linéairement d'une autre.

1.3.1. Les problèmes posés par l'estimation de variables interactives

Les variables explicatives intervenant dans l'interaction peuvent être observées ou latentes. Les stratégies d'estimation du modèle dépendent du statut de l'une et de l'autre. On considérera provisoirement la variable dépendante z comme observée, pour plus de simplicité.

1.3.1.1. Interaction entre variable observée et variable latente

- Supposons que z dépende d'une variable x observée, d'un groupe T de variables observées et d'une variable latente F sous-jacente à un groupe observé Y . Si l'on pense que l'effet de x sur z peut dépendre de la valeur courante de F – et réciproquement – on envisagera le modèle avec interactions suivant :

$$z = \alpha + \beta x + \varepsilon F + \delta Fx + Tc + u$$

- En l'absence d'effet d'interaction δ , l'ACT calcule le facteur F en deux temps : dans un premier temps, on maximise le R^2 en régressant z sur $\{x, T, Y\}$ et en prenant pour valeur transitoire de F la composante sur $\langle Y \rangle$ de la partie expliquée $\hat{z}$. Dans un deuxième temps, la valeur finale de F est rapprochée des structures fortes de Y par un calcul de résultante.

En présence d'interaction, la maximisation du R^2 doit faire intervenir les deux termes F et Fx . Elle ne peut donc plus être effectuée par une simple projection sur un sous-espace parallèlement à un autre.

1.3.1.2. Interaction entre deux variables latentes

- Si deux variables latentes F et G interagissent l'une sur l'autre dans le modèle de z , on obtient un modèle du type :

$$z = \alpha + \beta F + \delta G + \varepsilon FG + Tc + u$$

- La situation est un peu plus complexe que dans le cas d'une interaction entre une variable latente et une observée, car la maximisation du R^2 dépend de trois termes où interviennent les variables latentes. Mais il est aisé de concevoir un algorithme itératif qui tour à tour considère F (resp. G) comme fixé pour optimiser G (resp. F).

1.3.2. Estimation du modèle d'une variable observée

1.3.2.1. Interaction entre une variable latente et une ou plusieurs observées

On considère une variable latente F sous-jacente au groupe explicatif Y . On suppose que cette variable interagit avec plusieurs variables explicatives observées $x^1, \dots, x^J$. Chacune d'elles forme un groupe à elle seule : on ne cherche pas à résumer $\{x^1, \dots, x^J\}$ par une variable latente. On note $T_1, \dots, T_r, \dots, T_R$ les groupes des autres variables explicatives, et T leur réunion. Les variables latentes des groupes $T_1, \dots, T_r, \dots, T_R$ sont respectivement notées $H_1, \dots, H_r, \dots, H_R$ et on note $H = \{H_1, \dots, H_R\}$. On note $X = \{x^1, \dots, x^J\}$. Enfin, $b = \{\beta_j\}_{j=1 \text{ à } J}$, $d = \{\delta_j\}_{j=1 \text{ à } J}$ et $c = \{\gamma_r\}_{r=1 \text{ à } R}$ sont des vecteurs de coefficients.

Notation. – Soit une variable x . On notera $Y \otimes x$ (ou $x \otimes Y$) le groupe des variables obtenues en faisant le produit de x et de celles de Y , soit : $Y \otimes x = \{xy^j | y^j \in Y\}$. Matriciellement, on a : $Y \otimes x = \text{diag}(x)Y$.

Le modèle de z s'écrit : $z = \alpha + \sum_{j=1}^J \beta_j x^j + \varepsilon F + \sum_{j=1}^J \delta_j x^j F + \sum_{r=1}^R \gamma_r H^r + u$, soit :
 $z = \alpha + Xb + \varepsilon F + (F \otimes X)d + Hc + u$, où u est un résidu décorrélié aux explicatives. Il est utile de noter que ce modèle s'écrit aussi : $z = \alpha + Xb + (\varepsilon + Xd)F + Hc + u$.

a) Estimation sans calcul de résultante

Oublions provisoirement tout calcul de résultante destiné à rapprocher les facteurs des structures fortes. Il s'agit alors simplement de déterminer $F \in \langle Y \rangle$ tel que la régression de z sur $\{F, X, F \otimes X, T\}$ ait le R^2 maximal. On pourrait certes utiliser une méthode numérique telle que l'algorithme de Newton-Raphson. Mais pour rester dans l'esprit de PLS et de l'ACT, nous proposons un algorithme géométrique plus simple, qui garantit seulement qu'à chaque itération, le R^2 augmente :

Initialisation

On régresse z sur $\{X, Y, x^1 \otimes Y, \dots, x^J \otimes Y, T\}$.

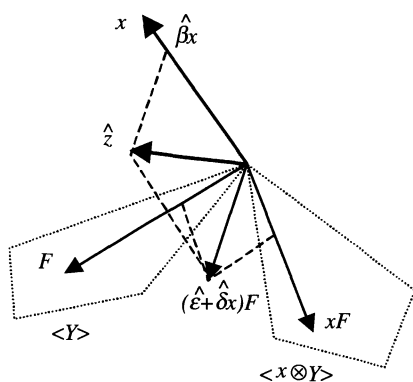
On prend la composante $\hat{z}_Y$ de $\hat{z}$ sur $\langle Y \rangle$ pour valeur initiale de F . Cette valeur initiale met l'accent sur l'effet « propre » de F , et non sur ses interactions. Mais les modèles avec interactions font naturellement intervenir des effets propres, ceux-ci apparaissant à la moindre translation sur une variable (e.g. son centrage).

Étape courante 1

On régresse z sur $\{X, F, F \otimes X, T\}$. On obtient ainsi : $\hat{z} = \hat{\alpha} + X\hat{b} + \hat{\varepsilon}F + (F \otimes X)\hat{d} + T\hat{c}$, que l'on écrit aussi : $\hat{z} = \hat{\alpha} + X\hat{b} + (\hat{\varepsilon} + X\hat{d})F + T\hat{c}$.

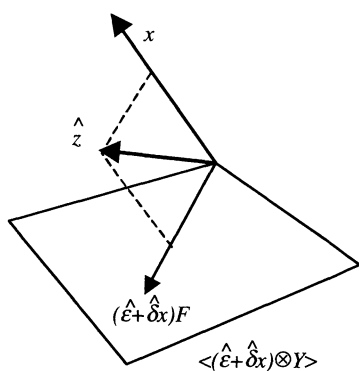
Étape courante 2

On forme le groupe $(\hat{\varepsilon} + X\hat{d}) \otimes Y$. Puis, on régresse z sur $\{(\hat{\varepsilon} + X\hat{d}) \otimes Y, X, T\}$. On prend alors la composante de $\hat{z}$ sur $\langle (\hat{\varepsilon} + X\hat{d}) \otimes Y \rangle$, et on la divise par $(\hat{\varepsilon} + X\hat{d})$. On obtient ainsi une nouvelle valeur courante de F : $\text{diag}(\hat{\varepsilon} + X\hat{d})^{-1} \Pi_{\langle (\hat{\varepsilon} + X\hat{d}) \otimes Y \rangle}^{\langle X, T \rangle} z$.



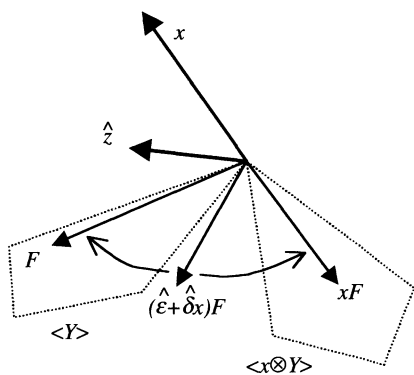
Étape 1

z est régressée sur x et les variables courantes F et xF . Ceci fournit les valeurs courantes de $\hat{\beta}$, $\hat{\varepsilon}$ et $\hat{\delta}$.



Étape 2 (1^{ère} phase)

Le sous-espace $\langle \hat{\varepsilon} + \hat{\delta}x \rangle \otimes Y$ est formé. z est régressée sur x et ce sous-espace. Ceci fournit une nouvelle composante $(\hat{\varepsilon} + \hat{\delta}x)F$.



Étape 2 (2^{ème} phase)

On en tire la nouvelle valeur courante de F , donc aussi celle de xF .

FIGURE 3
Étapes courantes de l'algorithme de base
(sans résultantes ni autres variables explicatives)

Si elle diffère assez peu de la précédente, on arrête. Sinon, on reprend l'étape courante 1.

L'algorithme est illustré figure 3 (avec une seule variable x et en l'absence de groupe T pour simplifier les représentations; b et d sont des scalaires, notés respectivement β et δ).

PROPRIÉTÉS . –

- Le R^2 augmente :

Preuve. – Les coefficients de l'étape courante 1 maximisent le R^2 sur le sous-espace $\langle F, X, F \otimes X, T \rangle$. Ils le font donc en particulier sur le sous-espace $\langle (\hat{\varepsilon} + X\hat{d})F, X, T \rangle$, auquel appartient la solution $\hat{z}$. Or, à l'étape courante 2, ce sous-espace est élargi à $\langle (\hat{\varepsilon} + X\hat{d}) \otimes Y, X, T \rangle$. La valeur de F issue de cette nouvelle maximisation ne peut donc qu'augmenter le R^2 .

- Si la régression $\hat{z}$ de z sur $\{X, Y, x^1 \otimes Y, \dots, x^J \otimes Y, T\}$ peut s'écrire de façon unique sous la forme : $\hat{z} = \hat{\alpha} + X\hat{b} + \hat{\varepsilon}F + (F \otimes X)\hat{d} + T\hat{c}$, alors il s'agit d'un point fixe de l'algorithme.

Preuve. – À l'étape 1, la régression de $\hat{z}$ sur $\langle X, F, F \otimes X, T \rangle$, sous-espace qui contient $\hat{z}$, fournit dans ce cas sous l'hypothèse d'unicité de l'optimum les coefficients $\hat{b}, \hat{\varepsilon}, \hat{d}, \hat{c}$ de l'écriture ci-dessus. À l'étape 2, la régression de z sur $\langle X, (\hat{\varepsilon} + X\hat{d}) \otimes Y, T \rangle$, sous-espace qui contient également $\hat{z}$, donne le facteur F de l'écriture ci-dessus. Les estimations sont donc stables.

- L'algorithme part de la régression sans interaction. Si, dans le vrai modèle, X n'interagit avec aucune des variables y^j , F est directement trouvé lors de l'initialisation, le coefficient $\hat{d}$ de l'étape 1 est nul, et l'étape courante 2 ne change pas la valeur de F . Le R^2 est d'ailleurs maximum dès l'initialisation.

- Le cas particulier d'une « interaction » de Y avec la constante ($X = x = 1$) donne la régression multiple sur $\{Y, x, T\}$

La propriété est évidente. En effet, $x = 1 \Rightarrow x \otimes Y = Y$ et $\forall a, b : \langle (a + bx) \otimes Y \rangle = \langle Y \rangle$. Par conséquent, l'étape courante ne fait que répéter l'initialisation, qui fournit $F = \hat{z}_Y$.

b) Estimation avec calcul de résultante

On introduit à présent dans l'algorithme précédent les calculs de résultantes qui rapprochent les facteurs de structures de corrélation fortes de leur groupe. Le groupe Y est pondéré par la métrique M , et chaque groupe T_r est pondéré par une métrique P_r .

Algorithme ACT1-ILO (ACT1 avec Interaction Latente-Observée)**Initialisation**

On réalise l'ACT1 de z sur les groupes explicatifs $\{x^1\}, \dots, \{x^J\}, Y, T_1, \dots, T_R$, ce qui fournit des valeurs initiales possibles de F et $H_1, \dots, H_R$. On note $H = \{H_1, \dots, H_R\}$.

Étape courante 1 : recalcul de F **Itération courante 1**

On régresse z sur $\{X, F, F \otimes X, H\}$. On obtient ainsi : $\hat{z} = \hat{\alpha} + X\hat{b} + \hat{\varepsilon}F + (F \otimes X)\hat{d} + H\hat{c}$. Soit encore : $\hat{z} = \hat{\alpha} + X\hat{b} + (\hat{\varepsilon} + X\hat{d})F + H\hat{c}$.

Itération courante 2

On forme le groupe $(\hat{\varepsilon} + X\hat{d}) \otimes Y$. Puis, on régresse z sur $\{(\hat{\varepsilon} + X\hat{d}) \otimes Y, X, H\}$. On prend alors la composante de $\hat{z}$ sur $\langle (\hat{\varepsilon} + X\hat{d}) \otimes Y \rangle$, et on la divise par $(\hat{\varepsilon} + X\hat{d})$. On calcule la résultante sur Y de la variable obtenue. On obtient ainsi une nouvelle valeur courante de F :

$$F = R_Y^M \alpha \text{diag}(\hat{\varepsilon} + X\hat{d})^{-1} \Pi_{(\hat{\varepsilon} + X\hat{d}) \otimes Y}^{<X, H>} z \text{ normé.}$$

Si elle diffère assez peu de la précédente, on passe à l'étape courante 2. Sinon, on reprend l'itération courante 1.

Étape courante 2 : recalcul de H

On réalise l'ACT1 de z sur les groupes explicatifs $\{x^1\}, \dots, \{x^J\}, \{F\}, \{x^1 F\}, \dots, \{x^J F\}, T_1, \dots, T_R$, ce qui donne les nouvelles valeurs de H . Si les valeurs des facteurs diffèrent assez peu des précédentes, on arrête. Sinon, on reprend l'étape courante 1.

PROPRIÉTÉS . –

- Lorsqu'il n'existe aucune interaction entre X et Y , l'algorithme trouve la solution lors de l'initialisation, puisque dans l'itération courante, $\hat{d} = 0$, donc $\hat{\varepsilon} + X\hat{d} = \text{constante}$ et $\langle (\hat{\varepsilon} + X\hat{d}) \otimes Y \rangle = \langle Y \rangle$. Les itérations successives reviennent à poursuivre l'ACT1 sans interaction entamée à l'initialisation.

- Le cas particulier d'une « interaction » de Y avec la constante ($X = x = 1$) donne l'ACT1 de z sur $\{Y, x, T\}$.

En effet : l'initialisation fournit pour F la valeur calculée par l'ACT1 ; l'étape courante 1 régresse z sur $\{F, x, H\}$ et calcule la résultante de la composante sur $\langle Y \rangle$, ce qui est exactement le recalcul de F par l'ACT1 ; enfin, l'étape courante 2 recalcule H selon l'algorithme ACT1.

1.3.2.2. Interaction entre plusieurs variables latentes

On considère ici Q variables latentes interactives $F_1, \dots, F_Q$ respectivement sous-jacentes aux groupes explicatifs $Y_1, \dots, Y_Q$ (respectivement pondérés par les métriques $M_1, \dots, M_Q$). Par ailleurs, on considère comme précédemment R variables latentes non interactives $H_1, \dots, H_R$ respectivement sous-jacentes aux groupes explicatifs $T_1, \dots, T_R$. On suppose donc que le modèle de z s'écrit : $z = \alpha + \sum_q \beta_q F_q + \sum_{q,l/q < 1} \delta_{ql} F_q F_l + Hc + u$, où u est un résidu décorrélié avec les explicatives. Lorsque F_q et F_l n'interagissent pas *a priori*, $\delta_{ql} = 0$.

Algorithme ACT1-ILL (ACT1 avec Interaction Latente-Latente)

Initialisation

On réalise l'ACT1 de z sur les groupes explicatifs $Y_1, \dots, Y_Q$ et $T_1, \dots, T_R$, ce qui fournit les valeurs initiales de $F_1, \dots, F_Q$.

Étape courante 1 : recalcul des F_Q

Pour $q = 1$ à Q : Itération courante q (recalcul de F_q) :

Pour recalculer F_q , on utilise l'étape 1 de l'algorithme ACT1-ILO en considérant toutes les autres variables $F_r (r \neq q)$ comme observées.

Lorsque les valeurs des facteurs sont jugées suffisamment stables, on arrête. Sinon, on reprend l'étape courante 1.

Étape courante 2 : recalcul de H

C'est exactement l'étape courante 2 de l'algorithme ACT1-ILO.

PROPRIÉTÉS . —

- Si l'on omet les calculs de résultante, il est clair qu'à chaque étape de l'algorithme, le R^2 augmente, puisqu'il le fait au cours de ACT1-ILO.

- S'il n'y a pas d'interaction entre les Y_q dans le vrai modèle de z , l'algorithme ACT1-ILL revient à faire l'ACT1 de z sur les groupes explicatifs.

En effet, la première étape de l'algorithme ACT1-ILL est composée d'itérations exécutant ACT1-ILO. Or, en l'absence d'interactions dans le vrai modèle de z , l'algorithme ACT1-ILO revient à l'ACT1, et cette itération de l'ACT1 utilisant alternativement les F_q n'est autre que l'ACT1 utilisant les groupes Y_q .

1.3.3. Estimation du modèle d'une variable latente

On considère à présent un groupe Z de variables dépendantes, dans lequel on calculera un facteur G estimant la variable latente sous-jacente. Par ailleurs, on a Q groupes de variables explicatives observées $Y_1, \dots, Y_Q$, certains d'entre eux pouvant

éventuellement être réduits à une seule variable. On postule l'existence d'interactions entre ces groupes. Enfin, on considère d'autres groupes explicatifs $T_1, \dots, T_R$ dans lesquels on cherchera *a priori* des variables latentes $H_1, \dots, H_R$ non-interactives.

Algorithme ACT-I (ACT avec Interactions)

Initialisation

On réalise l'ACT de Z sur les groupes explicatifs $Y_1, \dots, Y_Q, T_1, \dots, T_r, \dots, T_R$, ce qui fournit les valeurs initiales de $G, F_1, \dots, F_Q$ et $H_1, \dots, H_R$. On note $H = \{H_1, \dots, H_R\}$.

Étape courante 1

On utilise l'algorithme ACT1-ILL de G sur $Y_1, \dots, Y_Q$ et H pour calculer les nouvelles valeurs de $F_1, \dots, F_Q$.

Étape courante 2

On résout $\mathbf{R}(< F_1, \dots, F_Q, \{F_q F_l / \gamma_{ql} \neq 0\}, H_1 \dots H_R >; Z, N)$ pour calculer la nouvelle valeur de G .

Lorsque les valeurs des facteurs sont jugées suffisamment stables, on arrête. Sinon, on reprend l'étape courante 1.

1.3.4. Recherche de dimensions supplémentaires dans les groupes

Lorsque l'on ne se contente pas d'une seule variable latente dans certains groupes, il est loisible, comme dans l'ACT, de réitérer les algorithmes précédents en substituant à chacun de ces groupes ses résidus de régression sur les facteurs qu'on y a précédemment trouvés. En réitérant l'algorithme sur ces résidus, on n'oubliera pas de considérer comme un groupe en soi chacun des facteurs précédemment calculés (cf. [Bry 2003]) ainsi que chacun de leurs produits d'interactions.

Notons que si l'on a postulé des interactions possibles entre variables latentes de deux groupes Y_1 et Y_2 , dans lesquels on a déjà déterminé les facteurs de rang 1 (F_1^1 et F_2^1 respectivement), et que l'on cherche dans chacun de ces groupes un facteur orthogonal au précédent (notés F_1^2 et F_2^2 respectivement), seule l'interaction entre F_1^2 et F_2^2 sera considérée, et non celles entre F_2^1 et F_1^2 ni celle entre F_1^1 et F_2^2 .

1.3.5. Applications comparées de l'ACT1 et de l'ACT1-ILL

1.3.5.1. Données simulées

Nous allons d'abord utiliser des données simulées de sorte à illustrer clairement la puissance d'ACT1-ILL relativement à ACT1 lorsque certaines variables explicatives interagissent. Toutes les méthodes utilisent ici le programme $\mathbf{R} = \mathbf{Q}$.

- Pour 100 observations, nous avons généré 20 variables aléatoires indépendantes extraites de la loi uniforme $\mathcal{U}[-1, 1]$, scindées en deux groupes « thématiques » A

et B , et notées $a1$ à $a10$ pour le groupe A , et $b1$ à $b10$ pour le groupe B . Puis, nous avons calculé la variable c suivante :

$$c = 0.2(a1 + b1)/\sqrt{2/3} + 0.8(a2/10 + b2/10 + a2 \cdot b2)/\sqrt{53/450}$$

Note. – Les valeurs des variables ont été arrondies à la deuxième décimale.

- Toutes les variables sont naturellement centrées, y compris les produits $ai \cdot bj$, puisque $E(ai \cdot bj) = E(ai) \cdot E(bj) = 0$.

- La variable c s'écrit $c = 0.2c1 + 0.8c2$, où $c1 = (a1 + b1)/\sqrt{2/3}$ et $c2 = (a2/10 + b2/10 + a2 \cdot b2)/\sqrt{53/450}$ sont deux variables indépendantes de variance unitaire :

$$\forall i : V(ai) = V(bi) = (1 + 1)^2/12 = 1/3 ;$$

$$\forall i, j : V(ai \cdot bj) = E(ai \cdot bj)^2 = E(ai^2) \cdot E(bj^2) = V(ai) \cdot V(bi) = 1/9$$

$$\begin{aligned} \forall i, j : \text{Cov}(ai, bj) &= 0 ; \text{Cov}(ai, ai \cdot bj) = E(ai \cdot ai \cdot bj) = E(ai^2 \cdot bj) \\ &= E(ai^2) \cdot E(bj) = 0 \end{aligned}$$

De même : $\text{Cov}(bj, ai \cdot bj) = 0$.

Il s'ensuit que $V(a1 + b1) = V(a1) + V(b1) = 2/3$ et que $V(a2/10 + b2/10 + a2 \cdot b2) = V(a2)/100 + V(b2)/100 + V(a2 \cdot b2) = 53/450$. On a donc $V(c1) = V(c2) = 1$.

- On note que le coefficient de $c1$ dans c est quatre fois plus petit que celui de $c2$. La variable $c1$ ne contient pas d'interaction entre $a1$ et $b1$, tandis que dans $c2$, l'interaction de $a2$ et $b2$ est prépondérante en termes de variance.

Une méthode qui ne voit que les effets propres devrait donc dépister $a1$ et $b1$ comme premières composantes thématiques de c , tandis que la prise en compte des interactions possibles devrait réorienter les composantes thématiques de rang 1 vers $a2$ et $b2$ respectivement.

- On remarquera que chacun des groupes A et B étant composé de variables de même variance et décorréliées, il est dépourvu de structure et la recherche de composantes principales préalablement à une éventuelle régression n'est d'aucune utilité.

Enfin, le nombre des interactions possibles entre variables respectives des deux groupes s'élève à 100 (c'est le nombre des produits $ai \cdot bj$). En y adjoignant les effets propres des variables des deux groupes, on obtient 120 coefficients. Comme il n'y a que 100 observations, il est impossible de procéder à la régression de c sur $\{ai, bj, ai \cdot bj\}_{ij}$ pour estimer directement le modèle. La situation n'est donc guère confortable.

ACT sans interactions

On a procédé à l'ACT1 de c sur les deux groupes thématiques $A = \{ai\}_i$ et $B = \{bj\}_j$. On s'intéresse aux deux premiers facteurs obtenus dans chaque groupe, et tout particulièrement au facteur de rang 1. Le facteur k du groupe A (resp. B) est noté AFk (resp. BFk). Les plans factoriels (1,2) des groupes A et B sont donnés par les figures 4 et 5.

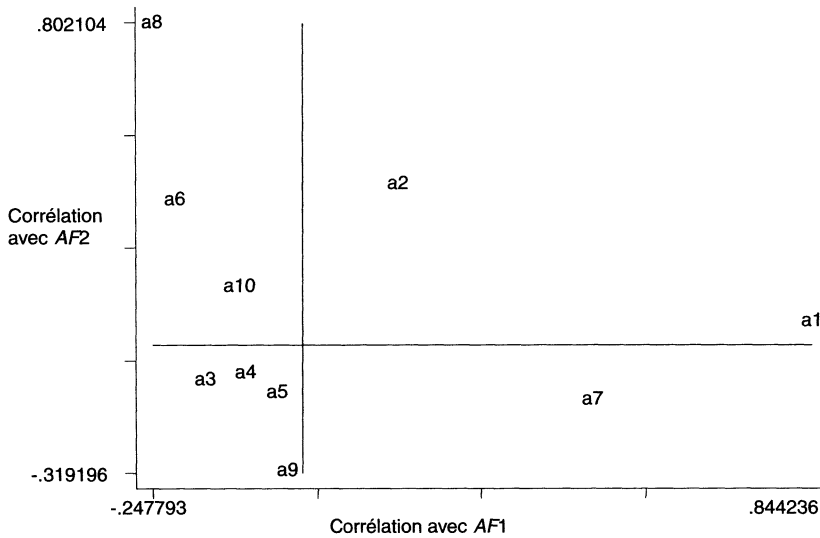


FIGURE 4
ACT, Plan (1,2) du groupe A

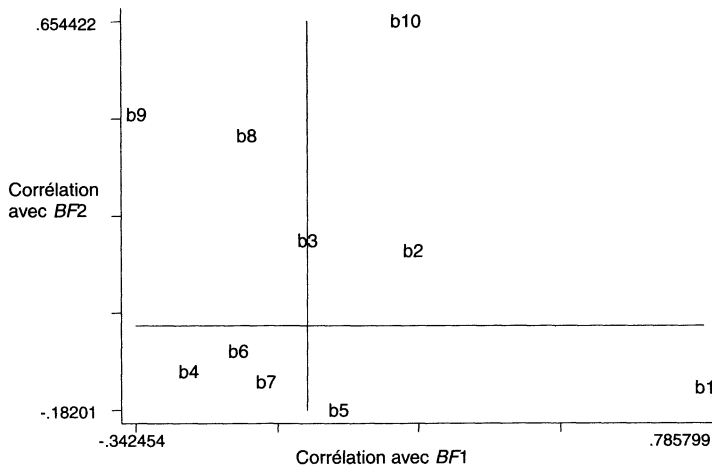


FIGURE 5
ACT, Plan (1,2) du groupe B

Les corrélations faites par les facteurs de rangs 1 et 2 avec les variables de leurs groupes respectifs sont les suivantes :

Groupe A	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
AF1	0.844	0.159	-0.160	-0.093	-0.042	-0.210	0.485	-0.248	-0.017	-0.106
AF2	0.067	0.400	-0.088	-0.070	-0.119	0.360	-0.137	0.802	-0.319	0.144

Groupe B	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
BF1	0.786	0.209	-0.002	-0.239	0.060	-0.142	-0.087	-0.125	-0.342	0.190
BF2	-0.134	0.161	0.182	-0.100	-0.182	-0.056	-0.122	0.408	0.454	0.654

L'ACT, ne « voyant » pas l'interaction entre a_2 et b_2 , a dépisté dans les deux groupes les variables respectives dont les effets propres expliquent le mieux la variance de c , à savoir a_1 et b_1 . Ce faisant, l'ACT n'explique qu'une part modeste de variance ($R^2 = 0.291$ pour la régression de c sur AF1 et BF1; $R^2 = 0.197$ pour celle de c sur a_1 et b_1).

Les facteurs de rang 2 AF2 et BF2 sont illustrés par les variables a_8 et b_{10} respectivement, mais s'avèrent ne jouer aucun rôle significatif dans la prédiction de c (la régression de c sur AF1, AF2, BF1 et BF2 donne un $R^2 = 0.313$, et les probabilités critiques* associées à AF2 et BF2 sont de 0.321 et 0.260 respectivement).

L'ACT sans interactions a donc manqué le phénomène représenté par la variable c_2 , lequel explique l'essentiel des variations de c .

ACT avec interactions

On a utilisé l'algorithme ACT1-ILL avec 5 itérations dans chaque boucle. Là encore, on s'intéresse aux facteurs de rang 1 et 2 de chaque groupe. Le facteur k du groupe A (resp. B) est noté $A\Phi k$ (resp. $B\Phi k$). On obtient les plans des figures 6 et 7.

Les corrélations des facteurs avec les variables de leurs groupes respectifs sont :

Groupe A	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
$A\Phi 1$	0.082	0.948	-0.113	0.225	-0.222	-0.302	0.093	0.173	0.029	-0.121
$A\Phi 2$	0.913	-0.152	0.003	-0.254	0.066	0.010	0.402	-0.205	0.008	-0.121

Groupe B	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
$B\Phi 1$	-0.004	0.930	0.136	0.176	-0.081	0.219	0.001	-0.277	-0.096	0.249
$B\Phi 2$	0.933	-0.075	0.041	0.091	-0.254	-0.097	-0.259	-0.246	0.197	-0.254

* Le calcul des facteurs de l'ACT impliquant la variable endogène, la probabilité critique n'est pas utilisable pour l'inférence. On ne la fait figurer ici qu'à titre descriptif, comme indicateur de l'importance de la liaison marginale entre le facteur et l'endogène.

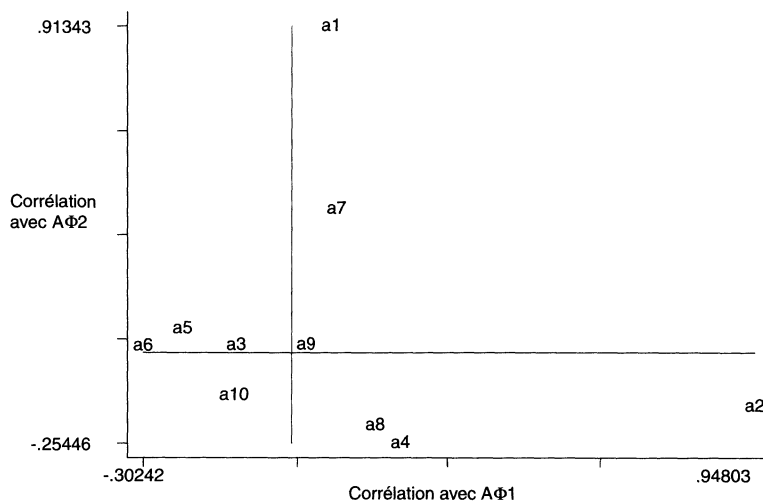


FIGURE 6
ACT-ILL, Plan (1,2) du groupe A

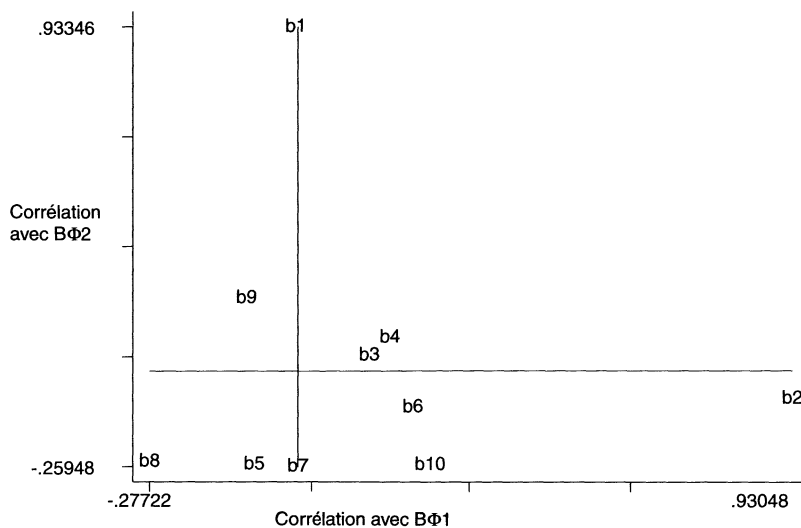


FIGURE 7
ACT-ILL, Plan (1,2) du groupe B

L'ACT1-ILL, dans un premier temps, a très bien dépisté les variables interactives des deux groupes qui expliquent l'essentiel de la variance de c : a_2 et b_2 respectivement. Dans un deuxième temps, elle a trouvé les variables dont les effets propres améliorent le mieux la prédiction de c , i.e. a_1 et b_1 .

Comme le modèle prévoit les interactions entre les deux groupes, il est nécessaire de calculer le produit des facteurs de rang 1 comme celui des facteurs de rang 2. La régression de c sur tous les facteurs et tous les produits ainsi calculés permet

de juger de l'existence et de l'importance de chaque effet. Cette régression donne ici un $R^2 = 0.926$, mais l'interaction entre facteurs de rang 2 apparaît ne pas jouer de rôle dans la prédiction de c , ce qui permet d'ôter du modèle le produit de facteurs correspondant. On obtient alors un modèle de c en fonction de $A\Phi 1$, $A\Phi 2$, $B\Phi 1$, $B\Phi 2$ et $A\Phi 1 \times B\Phi 1$. Son R^2 est de 0.921 et chaque facteur et terme d'interaction y joue un rôle significatif.

Dans notre exemple, le remplacement de chaque facteur par la variable de son groupe qui lui est la plus corrélée permet de retrouver le vrai modèle.

1.3.5.2. Échantillon d'immobilier locatif

Présentation

On considère ici un échantillon composé de 41 biens immobiliers mis en location à Dakar (cf. annexe). Pour chacun des biens, on a relevé le montant du loyer mensuel (cette seule variable constitue le groupe X1, qu'il s'agit d'expliquer), ainsi que trois groupes de caractéristiques explicatives :

- un groupe de caractéristiques de taille du bien : $X2 = \{\text{surface du terrain, surface habitable, surface habitable rapportée au nombre de pièces résidentielles, nombre total de pièces, nombre de pièces résidentielles, nombre de salles de bain, nombre de chambres, nombre de salons, nombre de WC, nombre de cuisines}\}$.
- un groupe de caractéristiques qualitatives du bien : $X3 = \{\text{villa, standing, état, jardin, cour ; piscine, garage, installations de haute technologie}\}$.
- un groupe de caractéristiques de situation (caractéristiques qualitatives du quartier) : $X4 = \{\text{distance au centre-ville, présence d'une zone commerciale, situation par rapport au bord de mer, présence d'hôtellerie, proximité d'un gros axe routier, standing du quartier ; quartier d'affaires}\}$.

Le modèle qui nous a paru le plus naturel est le suivant : la qualité du bien immobilier et celle du quartier déterminent la cherté du bien par unité de taille (la taille étant une variable latente mesurée indirectement de multiples manières). Cette cherté et la taille du bien déterminent le loyer qui en est demandé (cf. figure 8). Cette dernière détermination a toutes les chances d'être multiplicative sous l'hypothèse d'un rendement locatif variant peu.

On adoptera les notations suivantes :

Loyer = L (observée); Qualité quartier = Q (latente); Qualité bâtiment = B (latente); Taille = T (latente); Cherté = C (latente). Du modèle conceptuel ci-dessus, on tire les équations :

$$(a) \quad L = l_0 + C T ; \quad (b) \quad C = c_0 + c_Q Q + c_B B$$

On notera que dans ce modèle, on n'a mentionné aucune variable observable mesurant d'une façon ou d'une autre la cherté par unité de taille. Nous ne pouvons donc pas appliquer l'algorithme ACTIS directement à ce modèle conceptuel. Il nous faut transformer ce modèle de sorte que la cherté n'y apparaisse plus nommément.

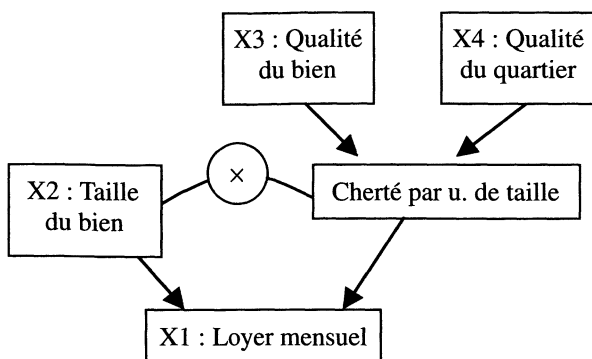


FIGURE 8
Modèle conceptuel du loyer

Des équations (a) et (b), on dérive :

$$L = l_0 + (c_0 + c_Q Q + c_B B)T \Leftrightarrow (c) \quad L = l_0 + c_0 T + c_Q T_Q + c_B T B$$

Nous avons finalement un modèle à une équation (c) avec deux interactions, correspondant au schéma conceptuel de la figure 9.

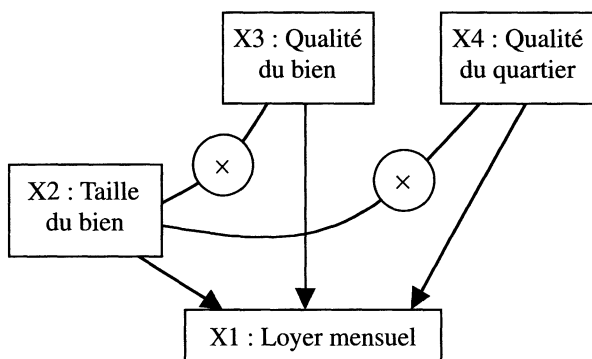


FIGURE 9
Une autre vision du modèle

Estimation

Il y a deux interactions impliquant la variable *Taille*. Nous avons utilisé le programme **R** = **Q** et successivement procédé aux estimations suivantes :

- Estimation sans calcul de résultantes, d'abord sans interactions, puis avec.
- Estimation avec calcul de résultantes, d'abord sans interactions, puis avec.

Lorsque l'estimation des facteurs a été faite sans postuler d'interactions, nous avons quand même calculé les produits des facteurs obtenus pour évaluer le modèle complet (avec interactions) fondé sur ces facteurs.

Les résultats des différentes estimations sont comparés dans les tableaux 1 à 4. En l'absence de résultantes, des résultats stables ont été atteints en dix itérations, tandis qu'avec résultantes, cinq ont suffi.

Qualité d'ajustement des modèles

Le modèle le mieux ajusté est, sans surprise, le modèle estimé avec interactions et sans résultantes (le R^2 est alors le critère à optimiser). La remarquable qualité de l'ajustement est dans ce cas attribuable à plusieurs choses :

- l'adéquation entre le modèle (qui contient des interactions) et la technique d'estimation (qui en tient compte);
- l'absence de contraintes sur les facteurs les rapprochant des structures fortes;
- la faiblesse relative des variations de loyer à taille et autres caractéristiques données, par rapport aux variations globales du loyer dans l'échantillon (lequel va de la chambre modeste à la villa de luxe).
- le nombre relativement élevé de variables explicatives.

TABLEAU 1
 R^2 des modèles

Modèle	T, B, Q, TB, TQ	T, B, Q, TB	T, B, Q, TQ	T, B, Q
Estimation				
Sans résultantes				
Sans interactions	0.9345	0.9215	0.9335	0.9168
Avec interactions	0.9934	0.8853	0.9773	0.8300
Avec résultantes				
Sans interactions	0.9170	0.8445	0.8569	0.7828
Avec interactions	0.9249	0.8390	0.8585	0.7570

On notera qu'en l'absence de résultantes, le fait de tenir compte des interactions lors de la détermination des facteurs permet un gain d'ajustement (passant de 0.934 à 0.993) bien supérieur à ce qu'il est en présence de résultantes (0.917 à 0.925), ce qui est normal au sens où l'introduction de résultantes restreint la liberté de l'estimation.

On remarquera enfin la perte de qualité d'ajustement lorsque les termes d'interaction TB et TQ sont omis dans le modèle. L'interaction entre taille et quartier apparaît plus importante que celle entre taille et qualité du bien.

Interprétation des facteurs

L'interprétation d'un facteur peut être faite de deux façons complémentaires :

- À partir des corrélations entre le facteur et les variables de son groupe : on répertorie les variables fortement corrélées positivement et négativement avec le facteur. On s'intéresse alors à la liaison globale entre le facteur et chaque variable.

– À partir des coefficients des variables dont la combinaison linéaire a donné le facteur. On s'intéresse alors à la liaison partielle entre le facteur et chaque variable, les autres variables étant fixées². Nous avons ici considéré les variables standardisées, afin de pouvoir comparer leurs coefficients en valeur absolue.

TABLEAU 2

Facteur taille

	Surf. terrain	Surf. habita.	Surf./ p. résid.	Nb pièces	Nb p. résid	Nb s. bains	Nb chambres	Nb s.a m.	Nb WC	Nb cuisines
Corrélations										
<i>Sans résultantes</i>										
Sans interactions	0.9148	0.7657	0.6581	0.7103	0.6269	0.6977	0.5838	0.6196	0.8234	0.2637
Avec interactions	0.9217	0.9317	0.8143	0.8750	0.8130	0.8446	0.7584	0.8005	0.8883	0.3303
<i>Avec résultantes</i>										
Sans interactions	0.8451	0.9837	0.7689	0.9797	0.9497	0.9111	0.9130	0.8698	0.8953	0.3631
Avec interactions	0.8421	0.9825	0.7678	0.9801	0.9502	0.9103	0.9118	0.8741	0.8934	0.3755
Coefficients										
<i>Sans résultantes</i>										
Sans interactions	.8360	.6336	-.4996	-.0828	-.3363	-.0778	-.4635	.0304	.7779	.0535
Avec interactions	.3515	.9798	-.2281	-.0724	-.2432	-.0741	-.3632	.0898	.3656	-.0554
<i>Avec résultantes</i>										
Sans interactions	.1370	.1330	.1064	.1251	.1177	.1183	.1092	.1172	.1257	.0423
Avec interactions	.1253	.1301	.1085	.1261	.1163	.1205	.1059	.1208	.1298	.0572

Taille du bien immobilier (cf. tableau 2) : Le facteur taille estimé est fortement corrélé avec toutes les variables de taille, ce qui nous donne le droit de l'interpréter comme tel dans tous les cas de figure. Il est en général plus proche de la surface habitable que de n'importe quelle autre variable. La variable la moins pertinente en matière de taille est le nombre de cuisines (une variable presque constante : dans la grande majorité des cas, il y a une seule cuisine, les rares exceptions étant les chambres isolées, qui n'en ont pas, et certaines villas de luxe qui en comportent deux).

² L'interprétation en termes de liaison partielle n'est possible qu'en l'absence de multicollinéarité. S'il y en a une, les coefficients obtenus ne peuvent être considérés que comme des pondérations des variables dans le calcul du facteur. Le lecteur vigilant aura noté que c'est ici le cas, le nombre total de pièces étant égal à la somme des nombres de pièces d'un type donné. Les séries de coefficients sont donc sous-déterminées. Les coefficients fournis ici ont été déterminés par la régression PLS, laquelle utilise une contrainte sur la norme du vecteur des coefficients. La conséquence en est que les effets sont répartis de la façon la plus équilibrée possible.

Concernant les coefficients des variables dans le calcul du facteur, de grosses disparités entre les différentes estimations apparaissent. Dans les estimations sans résultantes, les signes et valeurs absolues des coefficients sont disparates : des « transferts d'effets » ont manifestement lieu entre ces variables fortement corrélées. C'est ce que les estimations régularisées *via* les résultantes parviennent à éviter. Elles fournissent des coefficients de même signe et très équilibrés, la seule variable ayant un coefficient trois fois plus faible que les autres étant le nombre de cuisines, dont on a déjà noté le peu de pertinence.

Qualité du bien immobilier (cf. tableau 3) : Les estimations sans résultantes fournissent un facteur médiocrement (et parfois même négativement) corrélé avec ces variables, supposées augmenter la qualité du bien. Les coefficients de ces variables dans le calcul du facteur sont de signes et de modules variables. L'interprétation du facteur est, dans de telles circonstances, malaisée. La prise en compte des interactions améliore très nettement la situation : le facteur obtenu est positivement et souvent bien corrélé aux variables.

TABLEAU 3
Facteur qualité du bien immobilier

Corrélations	Villa	Standing	État	Jardin	Cour	Piscine	Garage	Hte Techno.
<i>Sans résultantes</i>								
Sans interactions	0.6477	0.0228	0.4207	-0.1566	0.0583	-0.0672	0.6134	0.4308
Avec interactions	0.8403	0.5410	0.6309	0.4449	0.5779	0.3126	0.9018	0.5913
<i>Avec résultantes</i>								
Sans interactions	0.5573	0.7664	0.6790	0.6881	0.2466	0.5168	0.7994	0.8089
Avec interactions	0.7178	0.7063	0.6892	0.6273	0.3959	0.5184	0.8781	0.7162
Coefficients								
<i>Sans résultantes</i>								
Sans interactions	.6529	-.1868	.4305	-.6024	-.3839	-.2697	.2000	.4354
Avec interactions	0.3411	0.0951	0.1427	0.0583	0.2815	-0.1024	0.3895	0.1081
<i>Avec résultantes</i>								
Sans interactions	.0649	.2695	.2442	.1627	-.0257	.1141	.1757	.3540
Avec interactions	.2167	.1937	.2458	.1467	.0883	.1109	.2432	.1924

En outre, dans l'estimation régularisée avec interactions les coefficients des variables dans le calcul du facteur sont tous positifs.

Qualité du quartier (cf. tableau 4) : Ici, toutes les variables mesurent *a priori* une commodité, sauf peut-être la distance au centre-ville. Les estimations fournissent un facteur corrélé négativement à la distance au centre et positivement à l'indicatrice du quartier d'affaires (qui se trouve au centre, précisément).

L'estimation non régularisée et sans interactions conduit à un facteur mal corrélé aux variables, et auquel les contributions de celles-ci n'ont pas toujours le signe attendu (la distance au centre-ville a par exemple un coefficient du même signe que celui du quartier d'affaires). On remarquera que l'estimation non régularisée avec interactions donne ici des résultats nettement plus facilement interprétables.

TABLEAU 4
Facteur qualité du quartier

Corrélations	Distance centre-ville	Zone commerce	Bord de mer	Restaurants Hôtels	Axe routier	Standing quartier	Quartier d'affaires
<i>Sans résultantes</i>							
Sans interactions	−0.4315	0.5654	0.1736	0.6077	0.5175	0.4923	0.5714
Avec interactions	−0.6374	0.2702	0.0267	0.5719	0.5392	0.3879	0.8872
<i>Avec résultantes</i>							
Sans interactions	−0.8647	0.5136	−0.2730	0.6405	0.3433	0.3014	0.7933
Avec interactions	−0.8296	0.3912	−0.2277	0.6306	0.4251	0.3918	0.8348
Coefficients							
<i>Sans résultantes</i>							
Sans interactions	.2124	.7041	.2024	.2273	.2498	.2578	.4622
Avec interactions	−.1153	.1801	.2193	.1599	.2493	.0125	.7226
<i>Avec résultantes</i>							
Sans interactions	−.3691	.2627	−.0330	.2593	.1513	.1122	.3593
Avec interactions	−.3537	.1566	−.0411	.2304	.1851	.1683	.4144

Toutefois, la régularisation améliore en valeur absolue la corrélation entre le facteur et les variables, et équilibre les contributions de ces dernières dans le calcul du facteur.

Conclusion

L'estimation tenant compte des interactions a fourni un ajustement meilleur que celle qui les ignore. D'autre part, l'estimation régularisée, en rapprochant les facteurs des structures fortes, les a rendus interprétables, ainsi que le rôle de chacune des variables dans son calcul. On retiendra donc le modèle issu de cette estimation. Le modèle estimé, entre variables centrées réduites, est :

$$L = -.1971 + .5094T + .1979B + .5284Q + .2590TB + .4349TQ$$

2. Modèle à plusieurs équations

2.1. Modèle

• Considérons R groupes $Y_1, \dots, Y_R$ respectivement pondérés par les métriques $M_1, \dots, M_R$ et supposons qu'il existe dans chaque groupe Y_r une variable latente F_r telle que l'ensemble de ces variables satisfasse un modèle linéaire à plusieurs équations (cf. figure 10). Le degré de prise en compte des structures internes à Y_r , si l'on utilise le programme $\mathbf{R} = \mathbf{P}$, est égal à α_r .

• Notations : Nous noterons E_r l'ensemble des facteurs explicatifs (y compris les produits de facteurs interactifs) d'un facteur F_r (et noterons de même la matrice formée par ces facteurs). Généralement, le facteur F_r du groupe Y_r dépend linéairement de E_r , mais fait aussi partie des prédicteurs d'autres facteurs $F_s, \dots, F_t$, l'ensemble de ces derniers étant noté K_r . L'ensemble des facteurs interagissant avec F_r dans le modèle de F_s est noté J_r^s . L'ensemble E_s privé de F_r et de $\{F_r F_q / F_q \in J_r^s\}$ est noté E_s^{-r} . On notera p_r^s la variable $\Pi_{<Y_r>}^{E_s^{-r}} F_s$ normée. La valeur de n'importe quel élément a à l'itération k est notée $a(k)$.

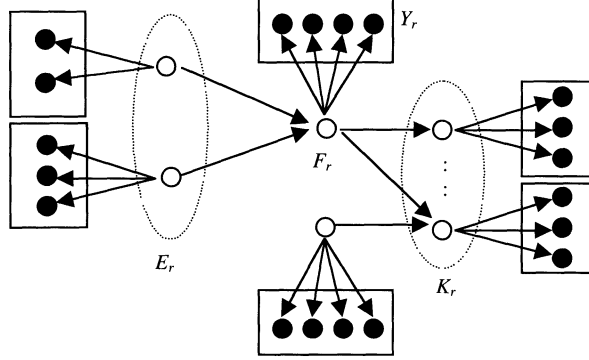


FIGURE 10

Système d'équations à variables latentes

2.2. Estimation du modèle sans interactions : l'algorithme ACTS

• H. Wold a proposé une technique empirique, PLS Path Modelling, pour estimer les variables et le système qui les relie (cf. [Wold 1985], [Tenenhaus 1999], [Bry 2003]). Dans le même objectif, nous proposons ici un algorithme assez proche de la variante de PLS-PM de Lohmöller ([Lohmöller 1989]) mais intégrant le principe de l'ACT. Cet algorithme est équivalent à l'étape 1 de l'ACT (facteurs de rang 1) dans le cas d'une équation simple.

• Lorsqu'une même variable latente intervient dans n équations, l'étape 1 de l'ACT présentée plus haut aboutira à n estimations de cette variable. Il est nécessaire de trouver un compromis entre ces estimations. Le programme **R** (i.e. **P** ou **Q**) fournit une manière naturelle de calculer un tel compromis. En effet :

– Dans le cas où F_r est un facteur purement dépendant ($K_r = \emptyset$), son modèle constitue une équation unique. F_r jouant le rôle de G dans l'étape 1 de l'ACT, il est aisé de voir que la $k+1$ -ième itération calcule $F_r(k+1)$ comme la solution normée du programme :

$$\mathbf{R}(Y_r, M_r; < E_r(k) >) = \mathbf{R}(Y_r, M_r; E_r(k), (E_r(k)' E_r(k))^{-1}).$$

– Lorsque F_r n'est qu'un facteur explicatif ($E_r = \emptyset$) intervenant dans une seule équation modélisant un facteur F_s (qui joue donc ici le rôle de G), cette

même itération calcule $F_r(k+1)$ comme solution normée du programme :

$$\mathbf{R1}(Y_r, M_r; p_r^s(k))$$

Ces deux cas sont les seuls possibles quand on estime une équation unique par la première étape de l'ACT. Leur traitement par l'ACT s'avère être un cas particulier de l'algorithme suivant.

L'algorithme ACTS (ACT- Système)

Itération 0 (initialisation)

On prend des valeurs initiales pour tous les facteurs thématiques, *e.g.* en résolvant $\mathbf{R}(Y_r, M_r; L_r, N_r)$, où $L_r = \bigcup_{s/F_s \in E_r \cup K_r} Y_s$ est la réunion des groupes en relation avec Y_r (explicatifs ou dépendants), et $N_r = \text{Diag}(M_s)_{s/F_s \in E_r \cup K_r}$ la matrice bloc-diagonale des métriques associées à ces groupes.

Itération $k+1$, $k \geq 0$

$F_r(k+1)$ est pris égal à la solution normée du programme :

$$\mathbf{R}(Y_r, M_r; [E_r(k), \{p_r^s(k)\}_{F_s \in K_r}], \text{Diag}((E_r(k)' E_r(k))^{-1}, I))$$

Note. – Si $\mathbf{R} = \mathbf{P}$, on peut moduler la prise en compte des structures internes au deuxième groupe en utilisant sa résultante de degré γ , où $\gamma \in \mathbf{R}^+$.

Si les valeurs courantes des facteurs sont assez proches de celles calculées à l'itération précédente, on s'arrête.

Note. – Lorsque $\text{card}(E_r \cup K_r) = 1$, $\mathbf{R}$ sera naturellement remplacé par $\mathbf{R1}$.

2.3. Estimation du modèle avec interactions : l'algorithme ACTIS

2.3.1. L'algorithme

Ainsi que nous avons dérivé l'algorithme ACTS de l'ACT pour estimer un système sans interactions, nous allons intégrer les principes de l'ACT-I dans un nouvel algorithme pour estimer un système avec interactions. On devra y retrouver l'étape 1 de l'ACT-I dans le cas d'une équation unique comportant une interaction.

Notations. – On reprendra celles de l'algorithme ACTS. En outre, soit F_r un facteur explicatif interagissant dans le modèle d'un facteur F_s , J_r^s l'ensemble des facteurs interagissant avec lui dans ce modèle, et soit A_r une variable du type $\alpha + \sum_{q/F_q \in J_r^s} \beta_q F_q$. On notera $i_t^s(A_r)$ la variable $(\text{diag}(A_r))^{-1} \Pi_{<A_r \otimes Y_r>}^{<E_s^{-r}} F_s$ normée.

Enfin, K_r^+ désignera l'ensemble des facteurs dont F_r est un facteur explicatif non interactif, et on notera $K_t^\times = K_r \setminus K_r^+$.

L'algorithme ACTIS (ACT avec Interactions et Système)

Itération 0 (initialisation)

On prend des valeurs initiales pour tous les facteurs thématiques, *e.g.* en résolvant $\mathbf{R}(Y_r, M_r; L_r, N_r)$, où $L_r = \bigcup_{s/F_s \in E_r \cup K_r} Y_s$ est la réunion des groupes en relation avec Y_r (explicatifs ou dépendants), et $N_r = \text{Diag}(M_s)_{s/F_s \in E_r \cup K_r}$ la matrice bloc-diagonale des métriques associées à ces groupes.

Itération $k+1$, $k \geq 0$

- Pour tout F_r et tout $F_s \in K_r^\times$:

1 – On régresse $F_s(k)$ sur $E_s(k)$. On note α le coefficient de $F_r(k)$ dans cette régression, et β_q celui de $F_r(k) \cdot F_q(k)$, où $F_q \in J_r^s$. On pose $A_r = \alpha + \sum_{q/F_q \in J_r^s} \beta_q F_q$.

2 – On calcule $i_r^s(A_r)$ puis on régresse $F_s(k)$ sur $\{i_r^s(A_r)\} \cup \{i_r^s(A_r) \cdot F_q(k) / F_q \in J_r^s\} \cup E_s^{-r}(k)$. On prend pour nouvelle valeur de α et de β_q les coefficients respectifs de $i_r^s(A_r)$ et $i_r^s(A_r) \cdot F_q(k)$ dans cette régression, et l'on recalcule A_r .

Cette étape 2 est recommencée jusqu'à stabilité suffisante de $i_r^s(A_r)$. On en note $i_r^s(k)$ la valeur finale.

- Pour tout r , $F_r(k+1)$ est pris égal à la solution normée du programme :

$$\mathbf{R}(Y_r, M_r; [E_r(k), \{p_r^s(k)\}_{F_s \in K_r^+}, \{i_r^s(k)\}_{F_s \in K_r^\times}], \text{Diag}((E_r(k)' E_r(k))^{-1}, I))$$

Note. – Si $\mathbf{R} = \mathbf{P}$, on peut moduler la prise en compte des structures internes au deuxième groupe en utilisant sa résultante de degré γ , où $\gamma \in \mathbf{R}^+$.

Si les valeurs courantes des facteurs sont assez proches de celles calculées à l'itération précédente, on s'arrête.

Note. – Lorsque $\text{card}(E_r \cup K_r) = 1$, $\mathbf{R}$ sera remplacé par $\mathbf{R1}$.

Remerciements

Je tiens à remercier vivement Pierre Cazes et Serigne Touba Diassé pour leur relecture attentive et leurs suggestions.

Références

- BRY X. (2003), *Une méthode d'estimation empirique d'un modèle à variables latentes : l'Analyse en Composantes Thématiques*, RSA vol. 51, n° 2, pp. 5-45.
- BRY X. (2001), *Une autre approche de l'Analyse Factorielle : l'Analyse en Résultats Covariantes*, RSA vol. 49, n° 3, pp. 5-38.
- CAZES P. (1997), *Adaptation de la régression PLS au cas de la régression après Analyse des Correspondances Multiples*, RSA vol. 45, n° 2, pp. 89-99.
- LOHMÖLLER J.-B. (1989), *Latent Variables Path Modelling with Partial Least Squares*, Physica-Verlag, Heidelberg.
- TENENHAUS M. (1998), *La régression PLS, théorie et pratique*, Technip.
- TENENHAUS M. (1999), *L'approche PLS*, RSA vol. 47, n° 2, pp. 5-40.
- VIVIEN M. (2002), *Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux : théorie et applications*, Thèse de doctorat, Université Montpellier I.
- WOLD H. (1985), *Partial Least Squares*, Encyclopedia of Statistical Sciences, John Wiley & Sons, pp. 581-591.

Annexe : données immobilières

Variables :

Loyer mensuel, en milliers de F CFA.

Caractéristiques de taille du bien immobilier :

- surface du terrain (m²) - surface habitable (m²)
- surface habitable rapportée au nombre de pièces résidentielles, *i.e.* chambres et salons (m²)
- nombre total de pièces (chambres, salons, cuisines, salles de bain, WC)
- nombre de salles de bain – nombre de chambres (chambres à coucher, salle de jeu, bureau...)
- nombre de salons (ou salles à manger) – nombre de WC (séparés) – nombre de cuisines

Caractéristiques qualitatives du bien immobilier :

- Villa (0 = appartement ; 1 = villa) – Standing (0 = non ; 1 = oui)
- État général (0 = mauvais ; 1 = moyen ; 2 = bon ; 3 = neuf) – Jardin (0 = non ; 1 = oui) – Cour (0 = non ; 1 = oui) – Piscine (0 = non ; 1 = oui) – Garage (0 = sans ; 1 = simple ; 2 = double)
- Installations de haute technologie, telles que panneaux ou chauffe-eau solaire, antenne parabolique... (nombre d'installations).

Caractéristiques de situation (quartier) :

- Distance au centre-ville (0 = centre-ville ; 1 = moins de 2 km du c.-v. ; 2 = entre 2 et 10 km du c.-v. ; 3 = plus de 10 km du c.-v.)
- Présence d'une zone commerciale (0 = à plus d'1 km ; 1 = à moins d'1km)
- Situation par rapport au bord de mer (0 = plages à plus d'1 km ; 1 = à moins d'1km)
- Présence de distractions de type hôtelier (restaurants, clubs, casinos, hôtels) (0 = plus de 2 km ; 1 = moins de 2 km)
- Accès à l'un des 4 axes routiers (Corniche, VDN, autoroute, route de Rufisque) reliant le centre-ville à la périphérie (0 = plus d'1 km ; 1 = moins d'1km)
- Standing du quartier (0 = irrégulier ; 1 = populaire régulier ; 2 = classes moyennes ; 3 = bourgeois)
- Quartier d'affaires (0 = non ; 1 = oui)

Données :

Quartier	Loyer	SurfTerrain	SurfHabitable	SurfHab/PiecR esid	NbPieces	NbPiecesResi	NbSDB	NbChambres	NbSalons	NbWC	NbCuis	Villa	Standing inn	Eiat	Jardin	Cour	Piscine	Garage	Hte Techno.	DisCrVille	Zone Commer	BordMer	Hôtellerie	AxeRoutier	Standing Quartier	Quartier Affaires
Fass	20	19	19	19,0	3	1	1	1	0	1	0	0	0	1	0	0	0	0	0	1	1	0	1	0	1	0
Fass	30	55	55	27,5	5	2	1	1	1	1	1	0	0	1	0	0	0	0	0	2	1	0	1	0	0	0
Derkle	35	40	40	13,3	6	3	1	2	1	1	1	0	0	1	0	0	0	0	0	2	1	0	0	0	2	0
Colobane	48	58	58	19,3	7	3	1	2	1	2	1	0	0	0	0	1	0	0	0	2	1	0	0	1	1	0
Bopp	50	75	75	15,0	9	5	1	3	2	2	1	0	0	2	0	1	0	0	0	2	1	0	0	0	1	0
GrandYoff	50	80	80	20,0	9	4	2	3	1	2	1	0	0	1	0	1	0	1	0	3	0	0	0	1	1	0
LiberteI	50	36	36	18,0	5	2	1	1	1	1	1	0	1	3	0	1	0	0	0	2	1	0	0	0	2	0
Malika	55	99	129	25,8	9	5	1	3	2	2	1	1	0	1	0	1	0	0	0	3	0	1	0	0	1	0
Yoff	55	53	53	17,7	6	3	1	2	1	1	1	0	0	2	0	1	0	0	0	3	1	1	0	1	1	0
NiayeCoker	60	59	59	14,8	7	4	1	3	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	0	1	0
Pikine	60	60	60	15	7	4	1	3	1	1	1	0	0	2	0	1	0	0	0	3	1	0	0	0	1	0
JetdEau	70	69	69	17,3	8	4	1	3	1	2	1	0	1	2	1	1	0	1	1	2	1	0	0	0	2	0
Medina	90	70	70	17,5	9	4	2	3	1	2	1	0	0	1	0	1	0	0	0	1	1	0	0	1	1	0
Medina	90	50	50	16,7	6	3	1	2	1	1	1	0	0	2	0	1	0	0	0	1	1	0	1	1	1	1
Yoff	90	71	71	23,7	6	3	1	2	1	1	1	0	0	1	0	1	0	0	0	3	1	1	1	1	1	0
Castors	95	154	139	19,9	11	7	1	6	1	2	1	1	0	0	0	1	0	0	0	2	1	0	0	0	2	0
GueuleTapee	95	90	90	18	10	5	2	4	1	2	1	0	0	1	0	1	0	0	0	3	1	0	1	1	1	0
HLM	120	149	149	21,3	12	7	2	5	2	2	1	1	0	1	0	1	0	2	0	2	1	0	0	0	1	0
LiberteVI	145	147	176	22,0	14	8	3	6	2	2	1	1	0	2	0	1	0	2	0	2	1	0	0	0	2	0
SacreCoeurIII	150	205	159	26,5	11	6	2	4	2	2	1	1	0	3	0	1	0	2	0	2	0	0	0	1	2	0
Parcelles	190	148	292	22,5	20	13	3	10	3	3	1	1	0	2	0	1	0	2	0	3	1	1	0	0	1	0
Mermoz	200	90	90	22,5	9	4	2	2	2	2	1	0	1	0	1	0	0	1	0	2	1	1	1	1	3	0
FenetreMermoz	240	96	96	19,2	10	5	2	3	2	2	1	0	1	2	0	1	0	1	0	2	1	1	1	1	2	0
Foire	280	350	270	33,8	15	8	3	6	2	3	1	1	1	2	1	1	0	2	1	3	0	0	0	1	2	0
BelAir	290	310	195	32,5	12	6	2	4	2	3	1	1	1	2	0	1	0	2	0	2	0	1	1	0	2	0
SacreCoeur	290	255	204	34,0	11	6	2	4	2	2	1	1	0	2	0	1	0	2	0	2	1	0	1	0	2	0
Fass	310	387	174	29,0	11	6	2	4	2	2	1	1	0	2	1	1	0	1	0	1	1	0	1	0	1	0
Hann	350	293	220	31,4	13	7	2	5	2	3	1	1	1	2	1	1	0	2	0	3	0	1	0	1	3	0
Plateau	350	60	60	15,0	8	4	1	3	1	2	1	0	1	2	0	1	0	0	0	0	1	0	1	1	3	1
Mermoz	390	203	198	33,0	11	6	2	4	2	2	1	1	0	2	1	1	0	2	1	2	1	0	1	1	2	0
Fann	440	100	100	20,0	10	5	2	3	2	2	1	0	1	3	1	1	0	1	0	2	1	1	1	1	3	0
Plateau	590	95	95	19,0	10	5	2	3	2	2	1	0	0	2	0	1	0	0	0	0	1	0	1	1	3	1
PointE	600	589	251	35,9	14	7	3	6	1	3	1	1	1	1	1	1	1	2	0	2	1	0	1	1	3	0
PointE	690	250	304	30,4	17	10	3	7	3	3	1	1	1	2	0	1	0	3	1	2	1	0	1	1	3	0
Ngor	800	307	285	31,7	15	9	2	6	3	3	1	1	1	3	1	1	1	2	0	3	1	1	1	1	3	0
FannHock	840	252	270	24,5	18	11	3	8	3	3	1	1	1	2	0	1	0	3	0	2	1	1	1	1	3	0
Mamelles	850	598	294	29,4	17	10	3	6	4	2	2	1	1	3	1	1	0	2	1	3	0	1	1	1	3	0
Mamelles	970	396	396	39,6	18	10	3	7	3	4	1	1	1	2	0	1	1	3	1	3	0	1	1	1	3	0
Almadies	1000	500	346	28,8	20	12	3	8	4	4	1	1	1	3	1	1	1	3	2	3	0	1	1	1	3	0
Plateau	1370	154	154	25,7	12	6	2	4	2	3	1	1	1	2	0	1	0	2	0	0	1	0	1	1	3	1
FannResidence	2100	988	390	35,5	20	11	4	8	3	4	1	1	1	3	1	1	0	3	2	2	1	1	1	1	3	0