

HENRI ARIS

PATRICK RIO

Une méthode d'analyse tridimensionnelle basée sur un rapprochement des tableaux considérés

Journal de la société statistique de Paris, tome 133, n° 1-2 (1992),
p. 106-124

http://www.numdam.org/item?id=JSFS_1992__133_1-2_106_0

© Société de statistique de Paris, 1992, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE BASÉE SUR UN RAPPROCHEMENT DES TABLEAUX CONSIDÉRÉS

par Henri ARIS
*Laboratoire de Biométrie,
ENSA-INRA-Université, Montpellier II*

et Patrick RIO
*Laboratoire d'Économie et Sociologie Rurale,
INRA, Montpellier*

Résumé

Nous utilisons une méthode qui permet l'étude de l'évolution d'individus repérés dans des axes qui sont soumis à des perturbations. Bien que cette méthode ait déjà été proposée dans des buts voisins en analyse Procuste Généralisée, il nous a semblé utile d'en faire un exposé, de donner ses liens avec l'Analyse en Composantes Principales et l'Analyse Canonique Généralisée, et de la mettre en œuvre sur des données économiques.

Abstract

We use a method which allows studying units while axis have some change. Although this method has been developed in similar purposes in Generalized Procrustes Analysis, we give its mathematical justification, its links with Principal Component Analysis, with Generalized Canonical Analysis, and an application to economic data.

Keywords

Generalized Canonical analysis, Three mode data analysis, Principal Component analysis.

INTRODUCTION

La méthode que nous exposons ci-après permet de décrire les données qui se présentent sous la forme de plusieurs tableaux $n \times p$. Elle peut s'introduire à partir de plusieurs points de vue :

- Le point de vue le plus voisin est celui de l'analyse canonique généralisée où l'on cherche les variables et les sous-espaces qui décrivent le mieux

l'ensemble des tableaux. J. Meulman dans [8] propose une telle utilisation de l'analyse canonique en analyse tridimensionnelle.

- Le point de vue de l'analyse Procruste Généralisée (J. C. Gower [4]) qui, en analyse factorielle permet de rapprocher des tableaux de facteurs pour mettre en lumière l'«invariance factorielle», c'est-à-dire une même structure sur les facteurs.

L'analyse Procruste Généralisée consiste à déterminer un tableau compromis à partir de tableaux initiaux au moyen de transformations orthogonales. Mc Keon [7] et Mc Donald [6] ont examiné le cas des transformations quelconques, ce qui est le cas de la méthode exposée ci-après, en dehors du cadre procuste.

- Le point de vue analyse tridimensionnelle : l'objectif recherché ici n'est pas le rapprochement de «facteurs», mais bien celui de tableaux initiaux, rapprochement à la suite duquel on peut décrire le comportement des individus en diverses «occasions», les évolutions des variables, les ressemblances des occasions ; nous retrouverons ici les étapes de «compromis», «intrastructure» et «interstructure» introduites par Y. Escoufier [3] dans la méthode Statis (C. Lavit [5]).
- Le point de vue analyse de variance : on cherche des transformations linéaires qui réduisent le plus possible une variance résiduelle.
- Le point de vue algébrique : la méthode peut être interprétée comme la recherche d'un centre de gravité généralisé dans le cadre d'un module sur un anneau de matrices. Cela donne lieu à une interprétation géométrique simple : supposons que m observateurs relèvent séparément les positions de n individus dans un espace R^p , que ces observateurs commettent de légères erreurs tant sur la position des individus que sur celle des axes, et que l'on désire une position «moyenne» des individus ; la méthode exposée ci-après donne une solution qui est indépendante des repères choisis par les divers observateurs. On pourrait appeler cette méthode : «Rapprochement des occasions» (R.D.O.).

Enfin, nous avons analysé par cette méthode les tableaux donnant la répartition de l'épargne d'une population de ménages entre divers produits bancaires, et ceci pour les années «occasions» de la période 1979-1987 (section VII).

1. Le problème

On considère m ensembles de n points dans l'espace affine R^p muni de sa base canonique. On suppose $m \geq 2$ et $p \leq n$. Chaque ensemble de n points donne lieu à une matrice X_i de dimensions $n \times p$, matrice des coordonnées de ses points dans la base canonique de R^p , i étant un indice qui varie de 1 à m .

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

Notre but est de chercher m transformations affines A_i , $i = 1, \dots, m$ de R^p de façon à ce que la somme résiduelle des carrés des distances entre les points transformés des divers ensembles soit minimale, ce qui revient à minimiser la quantité R (dans laquelle le «prime» désigne l'opération de transposition et «tr» l'opérateur trace) :

$$R = \sum_{i < j} \text{tr}[(A_i(X_i) - A_j(X_j))'(A_i(X_i) - A_j(X_j))] \quad (1)$$

Il est clair que pour éviter la solution nulle qui est triviale, il nous faudra imposer une contrainte aux transformations A_i .

Mais auparavant on sait que les transformations affines A_i peuvent se décomposer en le produit d'une translation T_i et d'une transformation linéaire S_i .

Pour le critère choisi, on montre assez facilement que chercher les translations T_i qui minimisent R revient tout simplement à centrer chacune des colonnes des matrices X_i , ce qui revient à déplacer tous les nuages de points de façon à ce que leurs centres de gravités soient confondus avec l'origine des axes (voir Gower [14]). Les colonnes des matrices X_i seront donc, dans une première étape, centrées.

Le problème devient maintenant :

Etant donné r un entier compris entre 1 et p , trouver m transformations linéaires, soit en pratique m matrices S_i $p \times r$ avec $r \leq p$ qui minimisent :

$$R = \sum_{i < j} \text{tr}((X_i S_i - X_j S_j)'(X_i S_i - X_j S_j)).$$

Sans contrainte sur les S_i , le minimum de R est zéro, obtenu pour $S_i = 0$. On impose alors la contrainte suivante, dite d'orthogonalité faible (Meulman [8]) :

$$\sum_{i=1}^m S_i' X_i' X_i S_i = K I_r \quad (2)$$

K étant un réel positif et I_r la matrice identité d'ordre r .

Il est évident que si les matrices S_i , $i = 1, \dots, m$ forment une solution du problème, alors, si M est une matrice orthogonale $r \times r$, les matrices $S_i M$, $i = 1, \dots, m$ forment aussi une solution de ce problème.

Pour simplifier la présentation, on prendra $K = 1$.

2. Solution

D'abord on développe R :

$$R = \sum_{i < j} [\text{tr}(S_i' X_i' X_i S_i) + \text{tr}(S_j' X_j' X_j S_j) - 2 \text{tr}(S_i' X_i' X_j S_j)]$$

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

et, comme pour m nombres réels $a_i, i = 1, \dots, m$, on a l'égalité :

$$\sum_{i < j} (a_i + a_j) = (m - 1) \sum_{i=1}^m a_i$$

R devient alors :

$$\begin{aligned} R &= (m - 1) \sum_{i=1}^{i=m} \text{tr}(S_i' X_i' X_i S_i) - 2 \sum_{i < j} \text{tr}(S_i' X_i' X_j S_j) \\ &= m \sum_{i=1}^{i=m} \text{tr}(S_i' X_i' X_i S_i) - \sum_{i,j} \text{tr}(S_i' X_i' X_j S_j). \end{aligned}$$

Si on pose \bar{X} :

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i S_i$$

on obtient aussi deux expressions intéressantes pour R :

$$\begin{aligned} R &= m \sum_{i=1}^{i=m} \text{tr}(S_i' X_i' X_i S_i) - m^2 \text{tr}(\bar{X}' \bar{X}) \\ &= m \sum_{i=1}^{i=m} \text{tr}((X_i S_i - \bar{X})' (X_i S_i - \bar{X})) \end{aligned}$$

Si nous notons

$$\begin{aligned} V_{ij} &= X_j' X_j \\ X &= (X_1 | X_2 | \dots | X_n) \quad (X \text{ est } n \times mp) \end{aligned}$$

$$S = \begin{pmatrix} S_1 \\ \dots \\ S_m \end{pmatrix} \quad (S \text{ est } mp \times r)$$

$$B = (V_{ij}) = X' X$$

$$A = (\text{bloc - diagonale } V_{ii})$$

nous avons :

$$\sum_{i=1}^m \text{tr}(S_i' X_i' X_i S_i) = \text{tr}(S' A S)$$

$$\sum_{i,j} \text{tr}(S_i' X_i' X_j S_j) = \text{tr}(S' B S)$$

Notre problème devient alors :

$$\left[\begin{array}{l} \text{trouver } S \text{ } mp \times r \text{ qui minimise } R \\ R = m \text{ Tr}(S' A S) - \text{Tr}(S' B S) \\ \text{en respectant la contrainte} \\ S' A S = I_r \end{array} \right. \quad (3)$$

ou

$$\left[\begin{array}{l} \text{trouver } S \text{ } mp \times r \text{ qui maximise} \\ \text{Tr}(S'BS) \\ \text{en vérifiant la contrainte } S'AS = I, \end{array} \right.$$

Lemme. $\text{Ker } A \subset \text{Ker } B$.

Preuve. Le i -ième bloc diagonal $p \times p$ de A est V_{ii} .

Soit :

$$j_i = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad i\text{-ième ligne} \quad i = 1, \dots, m$$

il est évident que :

$$v \in \text{Ker}(V_{ii}) \Rightarrow j_i \otimes v \in \text{Ker}(A)$$

et que les vecteurs :

$$\{j_i \otimes v \mid v \in \text{Ker}(V_{ii}), i = 1, \dots, m\}$$

forment un système de générateurs de $\text{Ker}(A)$; mais :

$$\begin{aligned} V_{ii}v = 0 &\Leftrightarrow X_i'X_iv = 0 \\ &\Rightarrow v'X_i'X_iv = 0 \\ &\Rightarrow X_iv = 0 \\ &\Rightarrow \forall j \quad X_j'X_iv = 0 \\ &\Rightarrow \forall j \quad V_{ij}v = 0 \\ &\Rightarrow j_i \otimes v \in \text{Ker}(B) \end{aligned}$$

et donc $\text{Ker}(A) \subset \text{Ker}(B)$.

La matrice A est bloc-diagonale, formée de blocs $p \times p$ semi définis positifs V_{ii} , elle est donc elle-même semi-définie positive.

Nous pouvons écrire :

$$A = \sum_{k=1}^y \mu_k \gamma_k \gamma_k'$$

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

où $v = \text{rang}(A)$
 μ_k : la k -ième valeur propre de A
 γ_k : le vecteur propre associé à μ_k et normalisé à 1.

Nous pouvons alors définir :

— une inverse généralisée A^{-1} of A :

$$A^{-1} = \sum_{k=1}^v (\mu_k)^{-1} \gamma_k \gamma_k'$$

— la racine carrée $A^{1/2}$ of A :

$$A^{+1/2} = \sum_{k=1}^v (\mu_k)^{1/2} \gamma_k \gamma_k'$$

— l'inverse généralisée $A^{-1/2}$

$$A^{-1/2} = \sum_{k=1}^v (\mu_k)^{-1/2} \gamma_k \gamma_k'$$

Avec ces notations et en tenant compte du fait que $\text{Ker } A \subset \text{Ker } B$, notre problème peut s'écrire, si on pose : $Y = A^{1/2}S$

$$\left[\begin{array}{l} \text{Trouver } Y \text{ } mp \times r \text{ qui maximise} \\ \text{Tr}(Y'A^{-1/2}BA^{-1/2}Y) \\ \text{sachant que} \\ Y'Y = I_r \end{array} \right.$$

Nous remarquons que :

$$(YY')(YY') = YY'$$

et que : $(YY')' = YY'$

donc YY' est un «projecteur» de R^{mp} de rang r . Comme :

$$\text{Tr}(Y'A^{-1/2}BA^{-1/2}Y) = \text{Tr}(A^{-1/2}BA^{-1/2}(YY'))$$

notre problème revient à trouver un projecteur P de rang r qui maximise :

$$\text{Tr}(A^{-1/2}BA^{-1/2}P).$$

On montre (R. Sabatier [9], [10]) que le maximum d'une telle quantité est égale à la somme des $q = \text{Min}(r, \text{rang } A^{-1/2}BA^{-1/2})$ plus grandes valeurs propres de la matrice :

$$A^{-1/2}BA^{-1/2}$$

chaque valeur propre étant prise en compte avec son ordre de multiplicité et que les q premiers vecteurs propres associés à ces valeurs et normalisés à 1 donnent une solution pour les colonnes de Y .

Si nous notons cette solution \tilde{Y} , nous avons alors :

$$A^{-1/2}\tilde{Y} = \tilde{S} = \begin{pmatrix} \tilde{S}_1 \\ \dots \\ \dots \\ \tilde{S}_m \end{pmatrix}$$

et

$$\tilde{X}_i = X_i \tilde{S}_i$$

En supposant que les valeurs propres λ_k de $A^{-1/2}BA^{-1/2}$ sont rangées en ordre décroissant

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \quad t = \text{rang}(A^{-1/2}BA^{-1/2})$$

le minimum de R obtenu pour cette solution optimale est alors :

$$R = mq - \sum_{i=1}^q \lambda_i \quad q = \text{Min}(r, t).$$

Remarque. Si K est distinct de 1, on trouve :

$$S = K^{-1/2}A^{-1/2}Y$$

et

$$R = K(mq - \sum_{i=1}^q \lambda_i) \quad q = \text{Min}(r, t).$$

3. Propriétés de cette solution

Soit \bar{X} :

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i \tilde{S}_i$$

Lemme 1.

Les colonnes de \bar{X} sont orthogonales.

Preuve.

$$\begin{aligned} \bar{X} &= \frac{1}{m} X\tilde{S} \\ \bar{X}'\bar{X} &= \frac{1}{m^2} \tilde{S}'X'X\tilde{S} \\ &= \frac{1}{m^2} \tilde{S}'B\tilde{S} \\ &= \frac{1}{m^2} \Lambda_q \end{aligned}$$

avec Λ_q diagonale et $q = \text{Min}(r, \text{rang } B)$.

Lemme 2.

Si m est valeur propre de $A^{-1/2}BA^{-1/2}$, si γ est un vecteur normalisé associé à m et si l'on note :

$$s = \begin{pmatrix} s_1 \\ \dots \\ \dots \\ s_m \end{pmatrix} = A^{-1/2}\gamma$$

alors

$$X_1s_1 = X_2s_2 = \dots = X_ms_m.$$

Preuve. Nous avons seulement à estimer la valeur de R quand $r = 1$

$$\begin{aligned} R &= \sum_{i < j} \text{tr}((X_i s_i - X_j s_j)' (X_i s_i - X_j s_j)) \\ &= m \text{tr}(s'As) - \text{tr}(s'Bs) \\ &= m - m \\ &= 0 \end{aligned}$$

La méthode nous donne les sous-espaces dans lesquels la position des individus est la même pour tous les observateurs.

Lemme 3.

Soit Z_i m matrices $p \times p$ de rang p , et

$$\hat{X} = X_i Z_i \quad i = 1, \dots, m.$$

L'application de la méthode aux \hat{X}_i , $i = 1, \dots, n$ (avec $\hat{K} = 1$) donne les mêmes résultats \bar{X} et \tilde{X}_i , $i = 1, \dots, m$.

Preuve. Soit Z la matrice bloc diagonale $mp \times mp$ dont le i -ième bloc est Z_i . Nous avons

$$\begin{aligned} \hat{X} &= XZ \\ \hat{B} &= Z'X'XZ = Z'BZ \\ \hat{A} &= Z'AZ. \end{aligned}$$

Il nous faut noter que :

$$A^{-1/2}BA^{-1/2}Y = Y\Lambda_r \quad \text{et} \quad Y'Y = I_r$$

peuvent s'écrire, si A^{-1} est une inverse généralisée de A :

$$A^{-1}BA^{-1/2}Y = A^{-1/2}Y\Lambda_r$$

$$A^{-1}BS = S\Lambda_r \quad \text{et} \quad S'AS = I_r$$

de la même façon : $\hat{A}^{-1}\hat{B}\hat{S} = \hat{S}M_r$ et $\hat{S}'\hat{A}\hat{S} = I_r$

deviennent : $Z^{-1}A^{-1}(Z)^{-1}ZBZ\hat{S} = \hat{S}M_r$
 $Z^{-1}A^{-1}BZ\hat{S} = \hat{S}M_r$
 $A^{-1}B(Z\hat{S}) = (Z\hat{S})M_r$ et $\hat{S}'Z'AZ\hat{S} = I_r$

nous pouvons alors identifier :

$$S = Z\hat{S} \quad \text{et} \quad \Lambda_r = M_r$$

soit $S_i = Z_i\hat{S}_i$

et donc : $\tilde{X}_i = \hat{X}_i\hat{S}_i = X_iZ_iZ_i^{-1}S_i = X_iS_i = \tilde{X}_i$.

4. Interprétation sous la forme d'une analyse en composante principale

Nous utilisons la présentation de l'analyse en composante principale donnée par Y. Escoufier dans [3]. Nous avons m matrices X_i $n \times p$, les colonnes de chacune des matrices X_i étant centrées, et toutes ces matrices étant supposées de rang p . On note :

$$\begin{aligned} X &= (X_1 \mid \dots \mid X_m) \\ B &= X'X \\ A &= \text{la matrice dont le } i\text{-ième bloc diagonal } p \times p \text{ est } X_i'X_i \\ D &= \frac{1}{n} I_n \\ Q_1 &= \frac{1}{n} A \\ Q &= Q_1^{-1} = nA^{-1}. \end{aligned}$$

On considère l'analyse en composantes principales de la matrice X utilisant la métrique D $n \times n$ pour les variables, et la métrique Q $mp \times mp$ pour les individus. Un axe principal se note u et vérifie les équations suivantes :

$$\begin{aligned} X'DXQu &= \lambda u & u'Qu &= 1 \\ BA^{-1}u &= \lambda u & u'(nA^{-1})u &= 1 \end{aligned}$$

qui peuvent s'écrire :

$$A^{-1/2}BA^{-1/2}(A^{-1/2}u) = \lambda(A^{-1/2}u) \quad \text{et} \quad (n^{1/2}A^{-1/2}u)'(n^{1/2}A^{-1/2}u) = 1$$

dans le précédent paragraphe, nous avons, si \tilde{y} est une colonne de \tilde{Y} :

$$A^{-1/2}BA^{-1/2}\tilde{y} = \lambda\tilde{y} \quad \text{et} \quad \tilde{y}'\tilde{y} = 1.$$

Ainsi

$$\tilde{y} = \sqrt{n} A^{-1/2} u$$

et pour les matrices :

$$\tilde{Y} = \sqrt{n} A^{-1/2} U.$$

Nous avons une interprétation de \bar{X} :

$$\begin{aligned} \bar{X} &= \frac{1}{m} X\tilde{S} \\ &= \frac{1}{m} XA^{-1/2}\tilde{Y} \\ &= \frac{1}{m} Xn^{1/2}A^{-1}U \\ &= \frac{1}{m\sqrt{n}} X(nA^{-1})U \\ &= \frac{1}{m\sqrt{n}} XQU. \end{aligned}$$

Ainsi les colonnes de \bar{X} sont identiques aux composantes principales du triplet (X, nA^{-1}, D) multipliées par la constante :

$$\frac{1}{m\sqrt{n}} \quad (\text{ou } \frac{\sqrt{K}}{m\sqrt{n}} \text{ si } K \text{ est différent de } 1)$$

5. Comparaison avec l'analyse canonique généralisée de Carroll [1]

Dans cette méthode, Carroll cherche un vecteur z $n \times 1$ centré et m vecteurs C_i , $p \times 1$ tels que

$$R^2 = \sum_{i=1}^m w_i [\rho(z, X_i C_i)]^2$$

est maximum, ou ρ est le coefficient de corrélation linéaire, et w_i un poids pour la i -ième matrice. Nous prendrons $w_i = 1$ pour tout i .

Il est bien connu que le maximum de ρ pour un z donné est (avec $D = \frac{1}{n} I_n$) :

$$\frac{z'D_i(X_i'DX_i)^{-1}X_i'Dz}{z'Dz}$$

R^2 s'écrit alors :

$$R^2 = \frac{1}{z'Dz} \sum_{i=1}^m z'DX_i(X_i'DX_i)^{-1}X_i'Dz$$

nous notons

$$M = \sum_{i=1}^m DX_i (X_i' DX_i)^{-1} X_i' D$$

le maximum de $R^2 = \frac{z' M z}{z' D z}$

est obtenu pour $z = D^{-1/2} v$, ou v est le vecteur propre de $D^{-1/2} M D^{-1/2}$ correspondant à la plus grande valeur propre.

$$D^{-1/2} M D^{-1/2} v = \lambda v \quad \text{et} \quad v' v = 1$$

soit $D^{-1} M z = \lambda z \quad \text{et} \quad z' D z = 1$

avec les notations du précédent paragraphe, on peut écrire :

$$M = DX (nA^{-1}) X' D$$

et donc : $D^{-1} (DX (nA^{-1}) X' D) z = \lambda z \quad \text{et} \quad z' D z = 1.$

Dans l'analyse en composantes principales de X avec la métrique nA^{-1} pour les individus et $D = \frac{1}{n} I_n$ pour les variables, la première composante principale est :

$$X (nA^{-1}) X' D c = \lambda c \quad \text{avec} \quad c' D c = \lambda$$

on en déduit donc que : $c = \sqrt{\lambda z}.$

Conclusion

Si nous notons (\bar{X}_j) la j -ième colonne de \bar{X} , c_j la j -ième composante principale, et z_j le j -ième facteur canonique dans l'analyse de Carroll, nous avons :

$$(\bar{X}_j) = \frac{1}{m \sqrt{n}} \quad c_j = \frac{\sqrt{\lambda_j}}{m \sqrt{n}} \quad z_j.$$

Remarque

Si K est différent de 1, on a :

$$\bar{X}_j = \frac{1}{m} \sqrt{\frac{K}{n}} \quad c_j = \frac{1}{m} \sqrt{\frac{K \lambda_j}{n}} \quad z_j.$$

6. Propositions pour une méthode d'analyse tridimensionnelle

La méthode proposée présente des étapes similaires à celles de la méthode «Statis» mise au point par Y. Escoufier [3] et exposée par C. Lavit [5].

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

La méthode donne une matrice compromis \bar{X} et des matrices transformées

$$\tilde{X}_i = X_i S_i.$$

D'abord dans le but de comparer les m occasions qui ont donné les matrices X_i , nous utilisons l'A.C.P. de la matrice :

$$T = (T_{ij}) = (\text{tr}(\tilde{X}_i \tilde{X}_j'))$$

ce qui nous donne l'«interstructure».

En second lieu nous cherchons une représentation des individus.

Dans \bar{X} , les individus sont repérés d'une manière «consensuelle» entre les juges (ou occasions). En utilisant \tilde{X}_i et \bar{X} , nous comparons les positions des individus vus par le juge n° i (année n° i) après réajustement des variables et vus par le juge «consensus» (position «moyenne» sur la période). Les axes utilisés sont ceux de l'A.C.P. du triplet (X, nA^{-1}, D) . L'ensemble de ces comparaisons forme l'«intrastructure».

De façon similaire, il est possible de comparer les variables initiales au moyen des matrices de corrélation :

$$D (1/s)_i X_i' D \bar{X} (D_{1/s})$$

où $(D_{1/s})_i$ et $(D_{1/s})$ sont respectivement les matrices diagonales des inverses des écarts-types de X_i et de \bar{X} , tandis que D est $(1/n)I_n$.

7. Un exemple

Monsieur P. Rio, économiste à l'INRA de Montpellier, nous a proposé d'analyser ses données avec la méthode précédente.

Les données se composent de neuf tableaux X_i , $i = 1, \dots, 9$ formés par les valeurs de six variables pour une population de 250 agriculteurs.

Les six variables sont les avoirs de ces personnes en dépôt le 31 décembre des années 1979 à 1987 dans six types de comptes bancaires : Titres, Dépôts à vue, Livret, Compte d'épargne logement, Plan d'épargne logement, Dépôts à terme.

La méthode fournit les tableaux transformés \tilde{X}_i et le tableau compromis \bar{X} .

Dans cet exemple, les axes varient peu d'une année sur l'autre, car les transformations S_i ne font passer le rapport inertie résiduelle sur inertie centrale que de 0,49 à 0,55.

La position des tableaux \tilde{X}_i donne l'«interstructure».

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

L'observation d'ensemble dans le plan 1-2 (figure 1, axe 1 : 56 %, axe 2 : 16 % de l'inertie totale) fait apparaître 3 périodes nettement distinctes :

- les années 79 à 81 se situent sur un axe de l'origine vers l'angle N-O du graphique ;
- les années 85 à 87 selon un axe de l'angle S-O vers l'origine ;
- les années 82 à 84 assurent la transition de l'angle N-O à l'angle S-O, l'année 83, plus rapprochée de l'origine, jouant un rôle de charnière.

Pour interpréter ces évolutions nous avons étudié les variances des colonnes des tableaux transformés ainsi que les moyennes et les variances des colonnes des tableaux initiaux :

- On remarque tout d'abord que les tableaux se disposent le long de l'axe 1 selon le critère de la variance totale du tableau transformé ; les plus grandes variances sont atteintes en 82, 84 et 85, les plus petites en 79 et en 87, une plus grande variance signifiant une plus grande instabilité pour certaines variables ce qui est ici le signe d'une évolution de la structure des portefeuilles. Il se trouve que ce critère est très proche de celui de la variance de la première colonne des tableaux transformés, colonne qui s'interprétera par la suite comme une colonne à dominance "titres".

Le premier axe s'interprète donc en termes d'instabilité ou de changement dans la structure des portefeuilles.

- D'autre part, le deuxième axe oppose les années 78-80-81-82 aux années 84-85-86-87 ; pour trouver l'interprétation de cet axe on examine le tableau des dépôts moyens. On constate une assez grande stabilité des dépôts moyens. Toutefois, en retenant les trois plus fortes valeurs moyennes de chaque type de dépôt, on observe que la majorité des maxima en dépôt dans les produits les plus liquides (dépôts à vue, livret et compte d'épargne logement) sont tous situés dans les années 79-83, alors que les années 84-89 concentrent la plupart des fortes valeurs des «dépôts de rendement» (dépôts à terme, plan d'épargne logement, titres).

Cette interprétation est confirmée par l'examen des variances des colonnes des tableaux transformés associées aux variables titres et dépôts à vue (que nous déterminons plus loin), variances qui opposent nettement les deux périodes.

Le deuxième axe oppose donc les années selon le critère de la plus ou moins grande liquidité du placement.

L'année 83 se caractérise cependant par l'association des maxima des dépôts en livret et compte d'épargne logement, une valeur encore élevée pour les dépôts à vue mais aussi une valeur élevée pour les titres. Les années 84 à 87 voient progressivement se succéder les maxima des dépôts à terme, des titres puis des plans d'épargne logement.

L'axe 3 (7 % de l'inertie totale, figure 2) oppose les années du début et de la fin de la période étudiée, soit 79-80 et 86-87 aux années de la transition 82 à 84,

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

les années 81 d'une part, 85 d'autre part assurant la liaison entre ces deux ensembles. Cette opposition s'interprète le mieux par l'opposition entre un poids élevé des titres dans l'ensemble des avoirs en début et en fin de période (plus de 37 % du total des dépôts sur ces périodes) à une importance réduite en 83-84 (inférieure à 34 %) et surtout 82 (moins de 29 %). Ces dernières années sont associées à un poids élevé des livrets, ce qui les distingue de l'année 1981 où les titres avaient déjà pris une moindre importance (32 %).

Étude des variables :

Ces résultats peuvent être synthétisés par l'étude des corrélations des variables avec les axes du compromis (qui sont les axes d'une ACP). On observe que le premier axe est fortement lié aux titres ($r=0,8$ à $0,9$), le second traduisant une liaison entre Livrets et Plans d'épargne logement, le troisième soulignant une opposition entre ces deux variables : corrélation positive avec les PEL et négative avec les Livrets. Le quatrième axe est associé aux Comptes d'épargne logement, les Dépôts à vue au cinquième axe.

On peut étudier sur des graphiques «cercles de corrélation» l'évolution des diverses variables dans le temps, par rapport aux axes du compromis. Ainsi la variable Titres, toujours fortement liée au premier axe évolue-t-elle régulièrement d'une corrélation négative vers une corrélation nulle avec la deuxième variable compromis représentant la liaison des Livrets et des Plans (figure 3).

Étude des individus :

C'est l'étape «intrastructure» : on peut tracer la trajectoire suivie par un individu ou par un groupe d'individus. Pour l'interprétation on tiendra compte des résultats observés dans l'analyse de la matrice des «juges», qui comportait l'extraction d'un plan résumant au mieux (% de la variance) les positions des tableaux entre eux, ce plan étant défini par l'évolution des titres d'une part, des liquidités (DAV) d'autre part. Il semble alors justifié d'engager l'étude des trajectoires des individus regroupés selon la valeur totale de leurs dépôts en 1979 dans le plan de l'espace «compromis» défini par le système d'axes le plus corrélé à ces variables (TITRES et DAV, qui portent 43 % de l'inertie totale du compromis).

Dans ce plan (figure 4), les petits portefeuilles se caractérisent par des valeurs faibles dans les deux dimensions. Les quatre premiers groupes se succèdent selon le gradient de liquidité. On observe un décrochement progressif, lorsque croît la taille moyenne du portefeuille, traduisant l'augmentation de la part des titres. Le cinquième groupe, le plus «riche», occupe une position originale, en raison de l'importance des titres bien sûr, mais aussi en raison d'une certaine réduction du volume des dépôts à vue qui s'établit au voisinage de la valeur moyenne pour l'ensemble de la population.

On observe que les trajectoires décrites par les différentes classes restent compatibles avec une individualisation de chaque classe au cours de la période envisagée, ce qui dénote une grande stabilité des différents groupes. Ce résultat n'est pas contredit pour les deux premières classes (les plus petits portefeuilles) qui sont cependant très proches l'un de l'autre.

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

L'effet de période conduit à un déplacement des centres de gravité par rapport au début de période vers le centre du graphique, ce qui peut s'interpréter comme une certaine homogénéisation des portefeuilles. Cette homogénéisation n'est toutefois pas complète puisque le groupe 4 manifeste à la fois une forte accumulation des liquidités et une faible croissance de la part des titres.

Les trajectoires des classes sont individuellement peu conformes au schéma moyen des tableaux. Le changement global (observé dans l'espace des juges) résulte donc de mouvements individuels dispersés. On peut interpréter ce résultat en disant que le tri des individus selon la dimension du portefeuille fournit un bon critère de dispersion des caractéristiques de ces portefeuilles mais que ce critère est peu instructif sur les arbitrages entre produits au cours du temps.

Les arguments permettant de rendre compte du choix de la combinaison d'actifs seraient donc distincts de ceux qui expliquent les changements, au cours du temps, de cette combinaison.

En regroupant les individus par classe d'âge (figure 5), les agents les plus âgés demeurent toujours plus éloignés des autres.

La position des centres de gravité des quatre classes d'âge évolue selon une forme en U déjà répertoriée dans la littérature économique. Les plus jeunes (c1) détiennent les portefeuilles où les dépôts à vue dominent (figure 5). Ceux-ci se réduisent progressivement au profit des titres. Ce mouvement est représenté sur le graphique par le déplacement des centres de gravité selon l'axe 1. Toutefois les individus ayant plus de 55 ans (c4) présentent des préférences plus marquées pour les produits liquides que ceux de la classe d'âge inférieure (c3) : il y a moindre diversification de portefeuille chez les ménages plus âgés.

Ce dernier résultat révèle un aspect important des données. Si l'évolution relativement continue des tableaux les uns par rapport aux autres au cours de la période étudiée domine la représentation obtenue, on met en évidence ici un effet d'âge dans le choix de la structure de la répartition des avoirs entre les différents types de dépôts.

En conclusion on peut dire que la déformation des tableaux entre eux n'est donc pas due seulement à l'évolution des variables, facilement interprétée par l'économiste en terme de variation des taux de rendements relatifs des différents produits, mais aussi au vieillissement des individus, aux variations des revenus ou de la richesse individuelle d'un individu d'une année sur l'autre.

BIBLIOGRAPHIE

CARROLL I., Douglas, Generalization of Canonical Correlation Analysis to Three or more Sets of Variables. *Proceedings*, 76th Annual Convention, APA, 1968.

ESCOUFIER Y., Principal Component Analysis with respect to Instrumental Variables, in *Methods for Multidimensional Data*, Lecture Notes from the European courses in Statistics.

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

- ESCOUFIER Y., *Three-mode Data Analysis : The "Statis Method"*. Lecture notes from the European courses in statistics.
- GOWER J.-C., Generalized Procrustes Analysis, *Psychometrika*, 1975, (40), 33-51.
- LAVIT Ch., *Analyse conjointe de tableaux quantitatifs*. Collection méthodes et programmes, Masson éd., 1988.
- MCDONALD R.-P., A Unified Treatment of the Weighting Problem, *Psychometrika*, 1968, 33, 351-381.
- MCKEON J.-J., *Canonical Analysis : some Relations between Canonical Correlation, Factor Analysis, Discriminant Analysis, and Scaling Theory*. *Psychometrika* monograph, 1966, 13.
- MEULMAN J., Distance Analysis in Reduced Canonical Spaces in *Multiway Data Analysis*. R. COPPI and S. BOLASCO (Editors). Elsevier Science Publishers B.V., 1989, 233-244.
- SABATIER R., *Méthodes factorielles en analyse de données : approximation et prise en compte de variables concomitantes*. Thèse d'Etat. Université des Sciences et Techniques du Languedoc. Montpellier, 1987.
- SABATIER R., JAN Y., ESCOUFIER Y., Approximation d'applications linéaires en analyse en composantes principales in *Data Analysis and Informatics*. North Holland, 1984, 569-580.

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

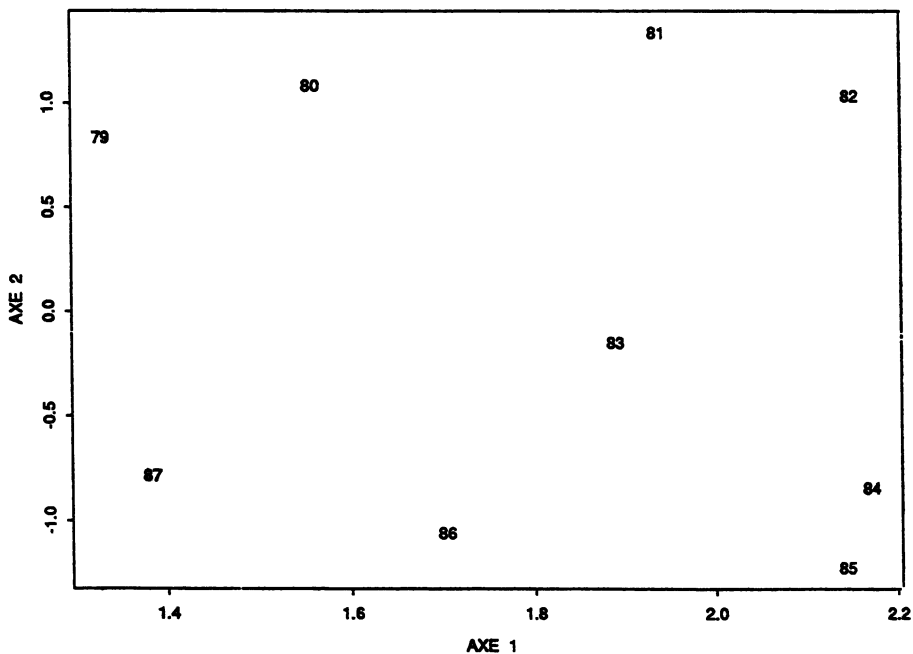


FIGURE 1. Interstructure : les années 1979 à 1987 vues dans le plan 1-2

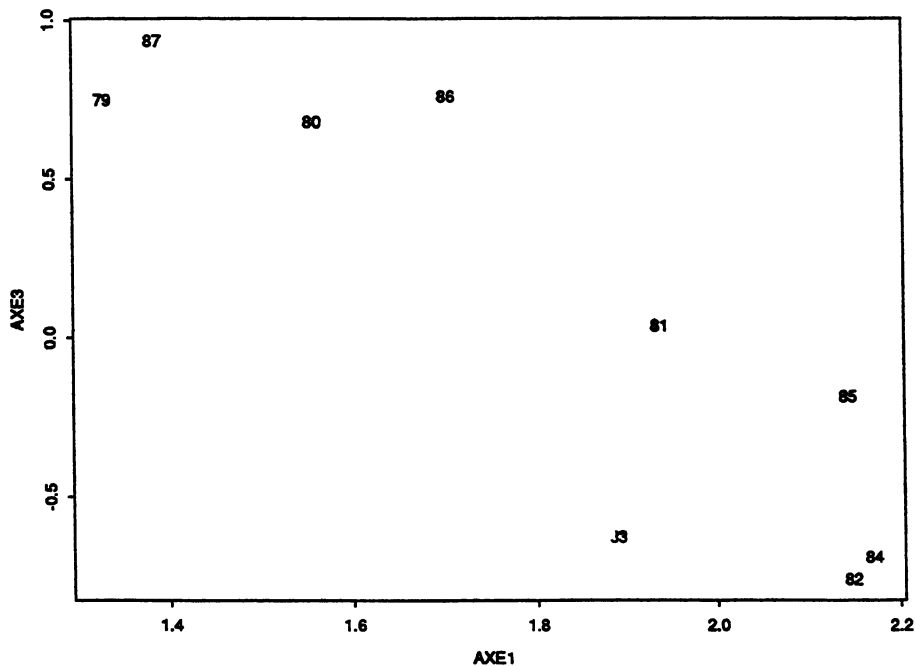


FIGURE 2. Interstructure : les années 1979 à 1987 vues dans le plan 1-3

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

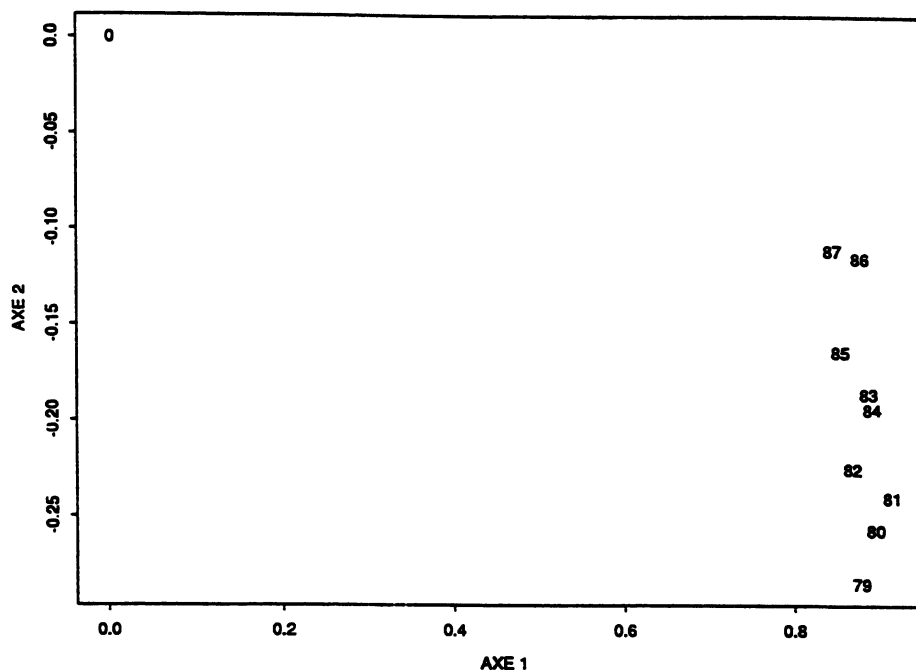


FIGURE 3. Cercle des corrélations. Le point 0 désigne l'origine.
Corrélations de la variable Titres avec les axes 1 et 2 du compromis

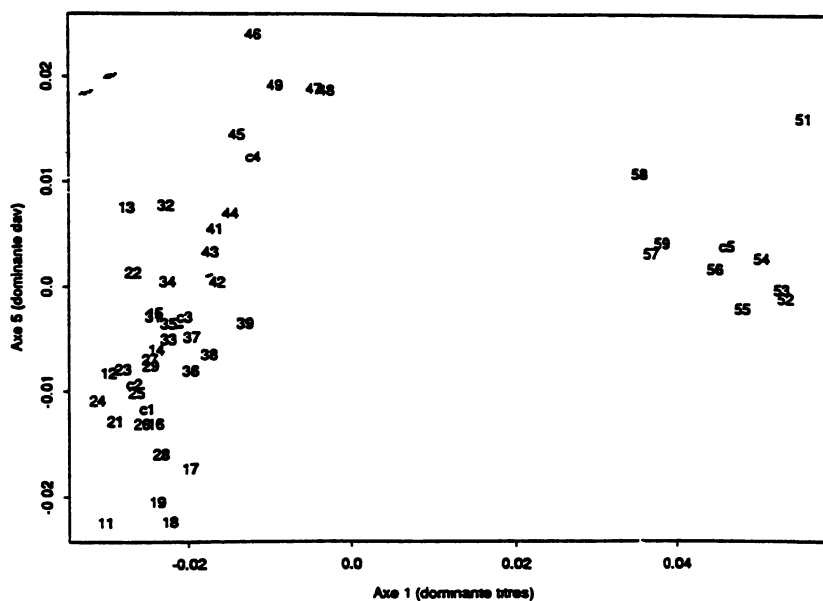


FIGURE 4. Positions relatives des centres de gravité des classes selon la taille du portefeuille :
 - le premier chiffre indique la classe de taille du portefeuille (de 1 à 5)
 - le deuxième chiffre est l'indice de l'année (1 pour 1979)

UNE MÉTHODE D'ANALYSE TRIDIMENSIONNELLE

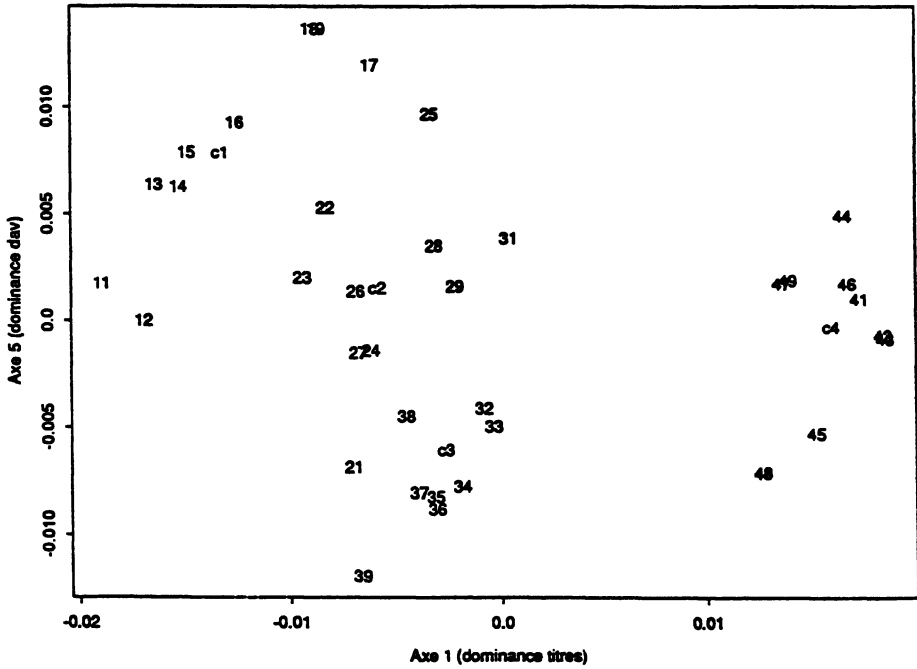


FIGURE 5. Positions relatives des centres de gravité des classes selon l'âge :

- le premier chiffre indique la classe d'âge (croissant de 1 à 4)
- le deuxième chiffre est l'indice de l'année (1 pour 1979)