REVUE DE STATISTIQUE APPLIQUÉE

THIERRY FOUCART

Tableaux symétriques et tableaux d'échanges

Revue de statistique appliquée, tome 33, nº 2 (1985), p. 37-54
<http://www.numdam.org/item?id=RSA_1985__33_2_37_0>

© Société française de statistique, 1985, 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/

TABLEAUX SYMETRIQUES ET TABLEAUX D'ECHANGES

Thierry FOUCART

Département de mathématiques
Université d'Orléans
45100 Orléans-La-Source

RESUME

L'analyse factorielle des correspondances d'un tableau symétrique possède des propriétés particulières bien connues. Nous en déduisons une méthode élémentaire pour modifier les termes diagonaux du tableau de façon que leur moyenne soit égale à celle des autres termes et que les facteurs restent invariants.

La description d'un tableau d'échanges consiste alors à effectuer l'analyse factorielle des correspondances du tableau symétrisé et à projeter sur les axes obtenus les lignes et les colonnes du tableau initial après en avoir modifié la diagonale : chaque zone est représentée par trois points dont l'un est le centre de gravité des deux autres pondérés par les exportations et les importations de la zône. La procédure peut s'étendre au cas de plusieurs tableaux d'échanges et simplifie les techniques de prévision basées sur la projection des facteurs. Nous donnons un exemple dans lequel nous montrons que ces méthodes sont moins efficaces qu'un simple ajustement du dernier tableau connu aux marges fixées.

ABSTRACT

Correspondance factor analysis of a symmetrical board possesses its own specific and well known properties. We thence infer an elementary method to change the board diagonal terms so that their average should be equal to the others' and leaving the factors invariable.

The description of an exchange board therefore consists in carrying out a correspondance factor analysis of the symmetrised board and projecting the lines and columns of the initial board on the then obtained axes after modifying the diagonal: each area is represented by three points, one of which is the center of gravity of the two others pondered by the exports and imports of the area. The procedure can be extended to the case of several input-output tables and simplifies the forecasting methods based on the factors' projection. An example is given, in which we show that these methods are less efficient than a simple adjustment of the last known board to the fixed margins.

1. ANALYSE DES CORRESPONDANCE D'UN TABLEAU SYMETRIQUE

Soit N_{IJ} un tableau symétrique : I est l'ensemble des lignes et J l'ensemble des colonnes. Bien que ces deux ensembles soient égaux, nous garderons les notations I et J. On note N la somme du tableau, P_{IJ} le tableau de probabilité associé, et n le nombre de lignes (et de colonnes) du tableau. On pose :

$$\forall i = 1,n \quad \forall j = 1,n \quad p_{i,j} = \frac{n_{i,j}}{N} \quad n_{i.} = \sum_{j=1}^{n} n_{i,j} \quad n_{.j} = \sum_{i=1}^{n} n_{i,j}$$

$$p_{i.} = n_{i.}/N \quad p_{.j} = n_{.j}/N$$

$$p_j^i = n_{ij}/n_{i.} \quad p_i^j = n_{i,j}/n_{.j}$$

Mots-clés : **Analyse factorielle, Correspondance symétrique, Ajustement aux marges, Tableaux d'échanges.**

La symétrie des données a une conséquence immédiate : les facteurs φ_ℓ définis sur I sont également les facteurs ψ_ℓ définis sur J, au signe près :

Définition

On appelle facteur direct (resp. inverse) un facteur φ_ℓ sur I égal au facteur ψ_ℓ de même rang sur J (resp. opposé).

La nature des facteurs dépend essentiellement des termes diagonaux : suivant que ces derniers sont nuls ou élevés, les facteurs auront tendance a être inverses ou directs. Le lien entre la nature des facteurs et la diagonale du tableau a été étudié complétement par J.P. BENZECRI (1973).

Soient u, v, w trois réels de somme 1. On pose :

$$P_{IJ}^{u,v} = u\, P_{IJ} + v\, P_{IJ}^{d} + w\, P_{IJ}^{p}$$

où P_{IJ}^{d} et P_{IJ}^{p} sont les tableaux ci-dessous :

$$p_{IJ}^{d} = \begin{vmatrix} p_{1.} & 0 \ldots 0 \ldots 0 \\ 0 & \cdot \qquad\qquad \cdot \\ & \cdot \qquad\qquad \cdot \\ \cdot & \cdot \qquad\qquad \cdot \\ 0 \ldots & p_{i.} \qquad 0 \\ \cdot & \cdot \qquad \cdot \\ \cdot & \cdot \qquad \cdot \\ 0 & p_{n.} \end{vmatrix} \qquad p_{IJ}^{p} = \begin{bmatrix} p_{1.}^{2} & \cdots p_{1.}\ p_{.j} \cdots p_{1.}\ p_{.n} \\ \cdot & \\ \cdot & \\ \cdot & \\ p_{i.}\ p_{.1} \cdots & p_{i.}\ p_{.j} \cdots p_{i.}\ p_{.n} \\ \cdot & \\ \cdot & \\ \cdot & \\ p_{n.}\ p_{.1} \cdots & p_{n.}\ p_{.j} \cdots p_{n.}\ p_{.n} \end{bmatrix}$$

avec : $\forall\, i = 1, n$

$$p_{i.} = p_{.i} = \sum_{j=1}^{n} p_{i,j}.$$

Le tableau P_{IJ}^{d} caractérise la liaison fonctionnelle entre les lignes et les colonnes, le tableau P_{IJ}^{p} l'indépendance sous contrainte de marges. Les marges du tableau $P_{IJ}^{u,v}$ sont évidemment égales à P_I et P_J.

Théorème

Si tous les termes du tableau $P_{IJ}^{u,v}$ sont positifs ou nuls, les facteurs $\varphi_\ell^{u,v}$ de variance 1 du tableau $P_{IJ}^{u,v}$ sont égaux aux facteurs de P_{IJ} et les valeurs propres $\lambda_\ell^{u,v}$ se déduisent des valeurs propres λ_ℓ par la formule suivante :

$$[\lambda_\ell^{uv}]^{1/2}\, \epsilon^{uv}(\varphi_\ell) = u\, [\lambda_\ell]^{1/2}\, \epsilon(\varphi_\ell) + v$$

$\epsilon(\varphi_\ell)$ (resp. $\epsilon^{uv}(\varphi_\ell)$) étant égaux à 1 lorsque φ_ℓ est direct pour P_{IJ} (resp. pour $P_{IJ}^{u,v}$), et à -1 lorsque φ_ℓ est inverse.

Conséquence : les facteurs d'un tableau restent invariants lorsqu'on ajoute à ses termes diagonaux des nombres proportionnels aux termes marginaux correspondants.

Il suffit de choisir en effet : $u + v = 1$.

On notera toutefois que l'ordre et la nature des facteurs peuvent changer : un facteur inverse de rang ℓ de P_{IJ} peut devenir un facteur direct de rang m de $P_{IJ}^{u,v}$.

2. ETUDE DES TERMES DIAGONAUX

2.1. Maximisation de l'inertie expliquée par les k premiers axes (KAZMIERCZAK, 1978)

En utilisant les propriétés précédentes, on peut déterminer les paramétres u et v de façon que l'inertie expliquée par les k premiers axes (k fixé) soit maximale. Plus précisément, en supposant u et v positifs strictement, on a :

$$P_{IJ} = \frac{1}{u} P_{IJ}^{uv} - \frac{v}{u} P_{IJ}^{d} - \frac{w}{u} P_{IJ}^{p}.$$

D'où la formule de reconstitution des données :

$$p_{i,j}^{*} = p_{i.}\, p_{.j} \left(1 + \frac{v}{u} + \sum_{\ell=1}^{k} \left(\epsilon(\varphi_\ell)\, \lambda_\ell^{1/2} + \frac{v}{u}\right) \varphi_\ell(i)\, \varphi_\ell(j)\right) - \frac{v}{u} p_{i.}\, \delta_{ij}$$

avec

$$(\forall i \neq j \;\; \delta_{i,j} = 0 \,;\, \delta_{i,i} = 1).$$

Cette formule ne dépend que du facteur $\rho = v/u$. En notant A l'ensemble des facteurs non pris en compte et P_{IJ} le tableau reconstruit, on pose :

$$E_\rho(A^*) = \sum_{i,j} (p_{i,j} - P_{i,j}^{*})^2 / p_{i.}\, p_{.j}.$$

L'erreur de reconstruction mesurée par $E_\rho(A^*)$ est minimale pour :

$$\rho_{min} = - \frac{\sum_{\alpha \in A^*} \epsilon(\varphi_\alpha)\, \lambda_\alpha^{1/2}}{\mathrm{Card}\,(A^*)}$$

La procédure consiste à minimiser par rapport à A* l'erreur $E_{\rho\,min}(A^*)$ et à appliquer un algorithme de données manquantes.

Les résultats numériques obtenus par KAZMIERCZAK sur un tableau de migrations montrent l'intérêt de cette méthode qui permet de diminuer notablement l'erreur de reconstruction pour un nombre de facteurs fixés. On peut regretter toutefois la complexité de sa mise en œuvre : il faut calculer tous les facteurs et toutes les valeurs propres pour en déduire le meilleur système.

2.2. Pondération de la diagonale

Pour diminuer l'influence des termes diagonaux dans la taille des valeurs propres lorsqu'ils sont nuls ou très élevés, on peut penser à les pondérer (STEMMELEN, 1977). Tenant compte de cette idée, nous allons procéder de deux façons suivant que la diagonale du tableau est nulle ou non.

– dans le cas où la diagonale est nulle, on pose :

$$\forall i \in I \quad n^1_{i,i} = \frac{1}{n-1} n_{i.}$$

Cette transformation laisse les facteurs invariants et la moyenne des termes diagonaux ainsi définis est égale à la moyenne des termes extra-diagonaux puisque : $n_{i.} = \sum_{j, j \neq i} n_{i,j}$.

Elle s'interprête facilement : le terme diagonal défini est égal à la moyenne des termes de la même ligne.

– dans le cas où la diagonale est chargée, on pose :

$$\forall i \in I : n^1_{i,i} = n_{i,i} - \alpha\, n_{i.}.$$

Il suffit de déterminer le coefficient α de façon que :

$$\frac{1}{n} \sum_{i \in I} n^1_{i,i} = \frac{1}{n(n-1)} \sum_{\substack{i \neq j \\ i,j}} n_{i,j}$$

On obtient facilement :

$$\alpha = \frac{\sum_i n_{i,i} - \frac{1}{n-1} \sum_{i \neq j} n_{i,j}}{\sum_{i,j} n_{i,j}}$$

Notons que la formule revient à la précédente dans le cas d'une diagonale nulle. Par ailleurs, il peut apparaître des termes négatifs ; on pourra alors choisir pour α la plus grande valeur telle que :

$$\forall i = 1, n \quad n^1_{i,i} \geqslant 0$$

mais l'égalité des moyennes telle qu'elle est cherchée par STEMMELEN ne peut plus être obtenue par cette méthode.

L'avantage de ces deux procédés est surtout d'être simples d'emploi ; ils ne nécessitent pas le calcul préalable des facteurs, et, dans l'optique de la description d'une suite de tableaux d'échanges, ils écartent les valeurs propres les unes des autres ; l'invariance des facteurs dans la transformation de la diagonale permet de déterminer les facteurs de même interprétation dans chaque tableau indépendamment du rang des valeurs propres associées ; cette propriété résoud, au moins en partie, une difficulté rencontrée par STEMMELEN et MADRE (MADRE, 1980).

3. ANALYSE FACTORIELLE D'UN TABLEAU D'ECHANGES

Un tableau d'échanges N_{IJ} est constitué par définition par les échanges d'une zone, d'un pays ou d'une région par exemple, que nous appelerons i, et d'une zone j. En règle générale, les zones exportatrices sont en lignes et les zones importatrices en colonnes. On a I = J, mais le tableau n'est pas symétrique.

3.1. Procédures classiques

On peut effectuer l'analyse factorielle des correspondances de plusieurs façons :

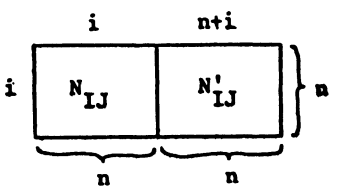
(i) le tableau transposé étant noté N'_{IJ}, il est classique d'en déduire les tableaux symétrique et antisymétriques en posant :

$$N^s_{IJ} = \frac{1}{2}(N_{IJ} + N'_{IJ}).$$

$$N^a_{IJ} = \frac{1}{2}(N_{IJ} - N'_{IJ})$$

En effectuant l'analyse des correspondances du tableau symétrisé N^s_{IJ}, on retrouve évidemment les propriétés exposées dans le paragraphe 1.

(ii) une autre procédure consiste à juxtaposer le tableau d'échanges et son transposé :



La symétrie de I et J est respectée. chaque zone est représentée par trois points : la ligne i correspond à la suite des exportations et importations, les colonnes i et i + n aux importations et aux exportations respectivement.

(iii) d'autres méthodes d'analyse factorielle existent ; citons les travaux de CONSTANTINE et GOWER (1978), qui portent sur les matrices symétriques et antisymétriques, d'ESCOUFIER et GRORUD (1979), qui font appel à l'analyse des données complexes, etc.

3.2. Analyse factorielle du couple (N_{IJ}, N'_{IJ})

La procédure développée dans ce paragraphe s'inscrit dans le cadre général de la description d'une suite de tableaux de probabilité (E. YAGOLNITZER, 1977 ; T. FOUCART, 1978 ; B. ESCOFIER, 1981). La démarche est la suivante :

(i) on effectue tout d'abord l'analyse des correspondances du tableau symétrisé N^s_{IJ}, obtenant ainsi un référentiel possédant toutes les propriétés énoncées dans le paragraphe 1 : facteurs directs, inverses, etc.

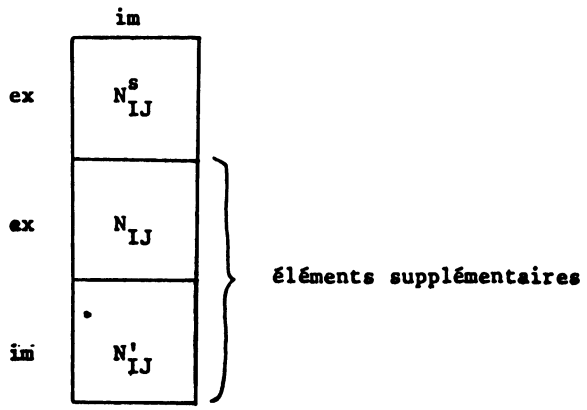
(ii) ce référentiel est utilisé pour représenter les lois conditionnelles en ligne du tableau initial N_{IJ} et de son transposé. La projection des lois conditionnelles en colonnes présente ici un intérêt limité car elle est redondante avec la précédente. En effet, si F_α correspond aux facteurs lignes et G_α aux facteurs colonnes, on a, comme on le vérifie facilement avec les formules de transition :

$$F_\alpha\,(\text{i exp}) = \epsilon_\alpha\, G_\alpha\,(\text{i imp}) \quad F_\alpha\,(\text{i imp}) = \epsilon_\alpha\, G_\alpha\,(\text{i exp})$$

où ϵ_α désigne la parité du facteur de rang α, i exp et i imp la zone i définie par ses exportations et ses importations.

(iii) on peut également projeter en éléments supplémentaires les marges des tableaux et les axes principaux obtenus par l'analyse des correspondances de chacun d'entre eux.

Cette procédure se résume par le schéma ci-dessous :



Chaque zone est représentée par trois points ; le premier la caractérise par ses échanges totaux (importations plus exportations), le second par ses exportations et le troisième par ses importations.

Le tableau N^s_{IJ} n'est autre que le tableau moyen des tableaux N_{IJ} et N'_{IJ}. On retrouve donc les principes d'interprétation déjà explicités dans l'étude d'une suite de tableaux de contingence ; on peut démontrer une propriété barycentrique qui prend, dans le cas d'un tableau d'échanges, un intérêt particulier.

Théorème

La loi conditionnelle sachant i du tableau symétrisé N^s_{IJ} est le centre de gravité des lois conditionnelles sachant i de N_{IJ} et N'_{IJ} pondérées par les coefficients $n_{i.}$ et $n_{.i} = n'_{i.}$.

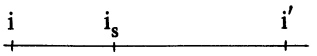
On a en effet :

$$(i,j) \in I \times J \quad p_j^{si} = \frac{n_{i,j}^s}{n_{i.}^s} = \frac{n_{i,j} + n'_{i,j}}{n_{i.} + n'_{i.}}$$

$$= \frac{n_{i.}}{n_{i.} + n'_{i.}} \frac{n_{i,j}}{n_{i.}} + \frac{n'_{i.}}{n_{i.} + n'_{i.}} \frac{n'_{i,j}}{n'_{i.}}$$

$$= \frac{n_{i.}}{n_{i.} + n'_{i.}} p_j^i + \frac{n'_{i.}}{n_{i.} + n'_{i.}} p_j'^i.$$

Les points i_s, i et i′ représentant les lois conditionnelles P_J^{si}, P_J^i, $P_J'^i$ se placent suivant le cas de l'une des façons ci-dessous :



: $n_{i.} > n'_{i.} = n_{.i}$; la zone i exporte plus qu'elle n'importe.

: $n_{i.} < n'_{i.} = n.i$; la zone i importe plus qu'elle n'exporte.

Cette propriété barycentrique n'est pas vérifiée si l'on modifie la diagonale du tableau N_{IJ}^s sans modifier aussi celles des tableaux N_{IJ} et N'_{IJ}. Pour qu'elle le soit, on modifiera les diagonales de ces tableaux de la façon suivante :

(i) dans le cas de termes diagonaux nuls, on pose :

$$\forall i \in I \quad n_{i,i} = \frac{1}{n-1} \sum_{j \neq i} n_{i,j}$$

$$n'_{i,i} = \frac{1}{n-1} \sum_{j \neq i} n_{j,i}.$$

Cela revient simplement à définir comme terme diagonal de N_{IJ} (resp N'_{IJ}) la moyenne des exportations de chaque zone (resp. des importations). On notera que les deux tableaux N_{IJ} et N'_{IJ}, une fois transformés, n'ont pas la même diagonale et ne sont plus transposés l'un de l'autre.

(ii) dans le cas de termes diagonaux non nuls, on pose :

$$\forall i \in I \quad n_{i,i} = n_{ii} - \alpha n_{i.}$$

$$n'_{i,i} = n_{i,i} - \alpha n_{.i}.$$

Le coefficient α est défini comme cela est indiqué dans le paragraphe 2.2 ; comme dans le cas précédent, les deux tableaux ne sont plus transposés l'un de l'autre après transformation.

Il existe évidemment d'autres méthodes de pondération de la diagonale d'un tableau d'échanges. On peut par exemple modifier le tableau de façon que les termes diagonaux soient égaux au produit des termes marginaux correspondants (Le FOLL, BURTSCHY, 1981). Cela revient à reconstituer la diagonale considérée comme manquante au 1er ordre, et la contribution des termes diagonaux à l'inertie totale du tableau devient nulle. Nous voyons deux inconvénients à cette méthode :

– tout d'abord, les marges sont modifiées : chaque zone est affectée d'un poids difficile à interpréter et les facteurs ne sont plus les mêmes.

– ensuite, on ignore si la reconstruction des termes extra-diagonaux est améliorée puisque l'inertie totale n'est pas invariante dans la transformation.

3.3. Analyse d'une suite de tableaux d'échanges

On peut sans difficulté généraliser la méthode proposée pour étudier une suite de T tableaux d'échanges, indexés par le temps par exemple. La démarche est alors la suivante :

(i) on transforme les diagonales de chaque tableau de la suite suivant la méthode expliquée précédemment. Chaque tableau est ainsi dédoublé en N^t_{IJ} et N'^t_{IJ} qui se distingue du transposé par les termes diagonaux.

(ii) on calcule le tableau moyen de la suite des 2T tableaux obtenus: il n'est pas nécessaire de pondérer sa diagonale qui vérifie déjà les critéres du paragraphe 2.2.

(iii) on projette ensuite les lois conditionnelles en lignes des tableaux N^t_{IJ} et N'^t_{IJ} sur les axes obtenus par l'analyse des correspondances du tableau moyen.

(iv) on peut aussi projeter la suite des marges.

Chaque zone étant représentée pour chaque tableau par trois points, la représentation risque d'être difficile à interpréter, mais il peut apparaître des évolutions régulières qui incitent à effectuer une projection dans le temps si les données sont chronologiques.

Une méthode de projection très simple peut être envisagée (STEMMELEN, 1977) : elle consiste à calculer les droites des moindres carrés par rapport au temps sur les coordonnées des facteurs obtenus par l'analyse des correspondances de chaque tableau. En ne gardant que les axes significatifs, on diminue le nombre de régressions à effectuer en conservant le maximum d'information et en éliminant plus ou moins les perturbations aléatoires correspondants éventuellement aux derniers axes.

	tableau 1	tableau k	prévision
facteurs retenus	$\varphi_1(1), \psi_1(1)$	$\varphi_1(k), \psi_1(k)$	$\varphi_1(h), \psi_1(h)$
	.		
	.		
	.		
	$\varphi_\ell(1), \psi_\ell(1)$	$\varphi_\ell(k), \psi_\ell(k)$	$\varphi_\ell(h), \psi_\ell(h)$.

Une fois les facteurs projetés, on applique la formule de reconstitution des données pour reconstruire un tableau qui est ensuite utilisé comme tableau initial pour ajuster les marges (que l'on suppose en général connues) à l'aide de l'algorithme R.A.S. par exemple. On peut aussi déduire des projections des facteurs une base (orthonormée pour la métrique du χ^2 définie par les marges fixées) à l'aide de la méthode d'orthonormalisation de SCHMIDT (MADRE, 1980). Cette procédure possède l'avantage de conserver, pour chaque valeur de l'entier k, les sous-espaces engendrés par les k premiers vecteurs.

Cette méthode de projection n'est applicable que sous certaines hypothèses. Sans parler de celles qui concernent la nature temporelle des données (elles n'ont

pas été abordées par les auteurs cités), on voit immédiatement que l'ordre des facteurs de même interprétation n'est pas toujours le même dans tous les tableaux. Il est donc indispensable de pouvoir procéder à des regroupements indépendamment de la taille des valeurs propres.

Dans cette optique, les propriétés développées précédemment sont intéressantes : la pondération de la diagonale d'un tableau symétrique a pour effet d'écarter les valeurs propres les unes des autres sans modifier les facteurs. On peut donc envisager la procédure suivante :

(i) considérer la suite des tableaux symétrisés

(ii) transformer les diagonales suivant l'une des deux méthodes proposées

(iii) retenir les facteurs significatifs obtenus par l'analyse des correspondances de chaque tableau.

(iv) régresser ces facteurs, reconstruire le tableau symétrique et l'ajuster aux marges fixées pour obtenir une prévision du tableau d'échanges.

Notons en outre que, les tableaux étant symétriques, les facteurs φ_ℓ et ψ_ℓ sont égaux au signe prés ; cette propriété divise par deux le nombre de régressions à effectuer.

4. APPLICATIONS

4.1. Description des échanges commerciaux intra-CEE de 1981

Nous avons étudié par la méthode décrite dans le paragraphe 3.2. le tableau d'échanges intra-CEE de 1981 (tableau 1); sa diagonale éant initialement nulle, nous avons procédé au calcul préalable des termes diagonaux suivant l'approche expliquée dans le paragraphe 2.2. Après avoir examiné les conséquences de cette transformation sur la qualité de la représentation, nous présentons les résultats de l'analyse factorielle des correspondances du couple (N_{IJ}, N'_{IJ}).

TABLEAU 1

Echanges intra CEE de 1981

	RFA	FRA	ITA	P-B	BLX	R-U	IRL	DAN	TOTAL
RFA	0	18 303	12 820	12 682	10 450	10 420	651	3 019	68 345
FRA	16 014	0	10 245	3 796	7 597	7 249	463	604	45 968
ITA	10 916	9 796	0	1 763	1 873	4 169	203	368	29 088
P-B	18 518	7 064	3 396	0	9 173	5 749	368	1 111	45 379
BLX	10 108	9 158	2 539	6 719	0	4 479	207	575	33 785
R-U	11 030	6 146	3 177	5 068	3 380	0	5 116	1 855	35 772
IRL	686	592	197	348	197	3 169	0	47	5 236
DAN	2 415	706	661	512	237	2 118	73	0	6 722
TOTAL	69 687	51 767	33 035	30 888	32 097	37 353	7 081	7 579	270 295

RFA : République Fédérale d'Allemagne, FRA : France, ITA : Italie, P.B. : Pays-Bas, BLX : Bénélux, R.U. : Royaume-Uni, IRL : Irlande. Dan : Danemark.

Choix des axes principaux

En nous limitant tout d'abord à l'étude du tableau symétrisé, nous nous proposons d'étudier ici les conséquences de la pondération de la diagonale sur la qualité de la reconstruction des échanges suivant différents critères.

On dispose évidemment de la mesure classique donnée par la somme des valeurs propres non prises en comptes dans la formule de reconstitution ; on sait que cette somme est égale au carré de la distance du χ^2 entre le tableau de données et le tableau reconstruit. Dans le cas d'un tableau symétrique dont les termes diagonaux n'ont qu'une importance relative, cet indice n'est pas très adapté puisqu'il tient compte de la reconstruction de la diagonale. Nous proposons donc, en complément, le même indice limité aux termes extra-diagonaux, en décomposant la formule de reconstitution des données de la façon suivante (NIJKAMP, PAELINCK, 1976) :

$$\| P_{IJ} - P^*_{IJ} \|^2 = \sum_{i=1}^{n} (p_{i,i} - p^*_{i,i})^2 / p^2_{i.} + \sum_{i \neq j} (p_{i,j} - p^*_{i,j})^2 / (p_{i.}\, p_{j.})$$

où $p^*_{i,j}$ est l'échange (i , j) du tableau reconstruit par les k premiers axes.

Rappelons que la pondération de la diagonale respecte les lois marginales et ne modifie que les valeurs propres ainsi que l'ordre et la parité des facteurs. On a les résultats suivants :

Diagonale nulle inertie totale $\Sigma \lambda_i = 0.415$	diagonale estimée $\Sigma \mu_i = 0.208$
$\varphi_1 \ \epsilon (\varphi_1) = -1 \ \lambda_1 = 0.150 (36\,\%)$	$\varphi_4 \ \epsilon (\varphi_4) = \ \ 1 \ \mu_1 = 0.081 \ (39\,\%)$
$\varphi_2 \ \epsilon (\varphi_2) = -1 \ \lambda_2 = 0.111 (28\,\%)$	$\varphi_1 \ \epsilon (\varphi_1) = -1 \ \mu_2 = 0.046 \ (22\,\%)$
$\varphi_3 \ \epsilon (\varphi_3) = -1 \ \lambda_3 = 0.089 (21\,\%)$	$\varphi_2 \ \epsilon (\varphi_2) = -1 \ \mu_3 = 0.028 \ (13\,\%)$

Dans un premier temps, nous avons reconstitué les données à l'aide de deux premiers axes ; la qualité de la reconstruction est bien meilleure après modification de la diagonale : les distances du χ^2 entre tableau initial et tableau reconstruit passent de 0.154 a 0.081. Le second critère est aussi largement amélioré : la contribution absolue des termes extra-diagonaux à ces distances passent de 0.109 à 0.060.

On aboutit aux mêmes conclusions en considérant les trois premiers axes.

Description des échanges intra-CEE en 1981

Compte-tenu des résultats du paragraphe précédent, nous nous limitons à l'étude du tableau à diagonale transformée. Le quatrième axe du tableau initial est ainsi placé au premier rang.

Les deux premiers facteurs expliquent 61 % de l'inertie, et la liaison entre les exportations et les importations esmble caractérisée surtout par deux propriétés mises en évidence dans la représentation graphique du plan 1 x 2 (Fig. 1).

Le premier facteur est direct : il oppose le Royaume-Uni et le Danemark au reste de la CEE. Le deuxième facteur est inverse : il est déterminé par les profils du commerce de l'Allemagne et des Pays-Bas. On retrouve l'Irlande et le Royaume-Uni sur le troisième axe qui est inverse. L'interprétation de cette représentation graphique étant classique, nous n'insisterons pas plus sur ce point.

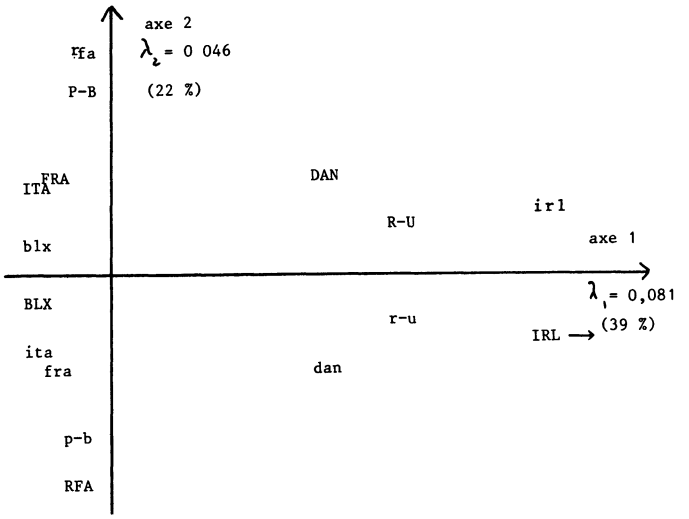


Figure 1. – Echanges moyens intra-CEE (1981)
Majuscules : exportations. Minuscules : importations (1er facteur direct, 2^{e} facteur inverse).

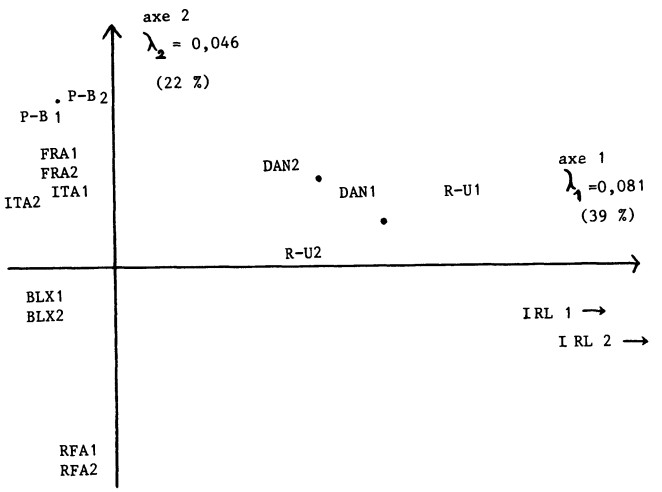


Figure 2. – Echanges INTRACEE (1981)
Exportations (codées 1)
Importations (codées 2)

La figure 2 est obtenue en représentant en éléments supplémentaires les lignes des tableaux N_{IJ} et N'_{IJ}. Nous avons représenté le tableau moyen par des points situés donc aux centres de gravité des identificateurs correspondants. Il apparait clairement sur ce schéma que les exportations (codées 1) et les importations (codées 2) d'un même pays ont des profils voisins: les clients sont également les fournisseurs. La plus grande différence concerne l'Irlande (non

TABLEAU 2

Migration dans la banlieue parisienne

	CHA	IVR	KRE	GEN	VIT	ALF	CHO	BON	VAL	ORL	RUN	FRE	THI	JOI	SUC
Charenton	6 238	269	45	14	204	824	57	250	70	76	16	36	0	403	189
Ivry	270	11 268	1 113	1 113	257	2 483	530	708	166	878	166	205	281	457	174
Kremlin	34	585	11 353	1 001	1 493	32	143	62	133	207	327	549	226	133	0
Gentilly	0	106	1 389	10 695	425	100	99	220	27	111	215	1 037	26	152	117
Vitry	186	667	894	281	11 263	1 009	1 577	148	123	1 021	154	265	860	314	90
Alfort	713	258	134	75	632	16 420	595	1 675	563	250	29	0	118	507	297
Choisy	0	181	78	41	763	148	5 590	24	396	964	104	38	745	25	87
Bonneuil	51	81	68	0	133	1 094	109	9 235	107	92	0	28	39	1 831	491
Valenton	31	34	34	28	34	316	271	148	6 161	628	0	0	59	83	228
Orly	14	108	492	177	353	104	528	209	568	6 461	315	408	551	191	130
Rungis	0	21	160	83	81	33	23	20	64	248	1 455	110	106	21	0
Fresnes	0	53	310	260	156	0	0	0	0	82	481	3 889	131	0	0
Thiais	0	66	21	0	151	40	421	24	43	248	26	0	1 498	25	0
Joinville	327	43	0	63	206	801	42	1 362	0	40	54	90	35	17 045	774
Sucy	0	0	0	26	26	20	28	159	591	102	0	0	0	403	5 624

représentée), le Royaume-Uni (R-U1 et R-U2) et le Danemark (DAN1 et DAN2). On remarque la particularité du commerce de l'Irlande et de l'Angleterre au bénéfice de cette dernière.

4.2. Etude des migrations dans la banlieue parisienne

Nous avons repris les données de KAZMIERCZAK (1978) pour étudier les effets de la modification des termes diagonaux dans le cas d'une diagonale chargée. Il s'agit des migrations entre 15 zones de la région parisienne ; le terme n_{ij} du tableau donne le nombre de départs de la zone j vers la zone i (le nom des zones est donné sur le tableau 2).

Il s'est révélé impossible d'obtenir des termes diagonaux de moyenne égale à celle des termes extra-diagonaux en respectant les marges. Nous avons donc déterminé le coefficient α tel que tous les termes de la diagonale sont positifs ou nuls.

Nous donnons ci-dessous les valeurs propres obtenues sur le tableau symétrisé initial (1ère ligne) et après transformation de la diagonale (2e ligne). Les cinq premiers facteurs sont invariants dans la transformation.

tableau	1	2	3	4	5	inertie totale
symétrique	0,783 13 %	0,648 11 %	0,601 10 %	0,545 9 %	0,467 8 %	5,846 100 %
transformé	0,633 25 %	0,427 17 %	0,364 15 %	0,283 11 %	0,197 8 %	2,504 100 %

La modification de la diagonale a donc pour effet une diminution de l'inertie totale et de chacune des cinq premières valeurs propres ; les pourcentages d'inertie expliquée par chaque axe augmentent et les valeurs propres s'écartent les unes des autres.

Le critère étant celui de la somme des valeurs propres non prises en compte dans la reconstruction, il apparait de façon évidente que la reconstruction est bien meilleure après la pondération des termes diagonaux (1,445 au lieu de 4,41 pour les deux premiers axes etc.). L'autre critère défini par la contribution absolue des termes extra-diagonaux conduit a la même conclusion : sa valeur passe de 0,855 à 0,429. Ce critère n'est pas optimisé par l'analyse puisque le plan 1×3 donne une moins bonne représentation des termes extra-diagonaux que le plan 1×4 (0,547 et 0,532).

La reconstruction obtenue par Kazmierczak est nettement meilleure (cf. paragraphe 2.1).

L'analyse factorielle du tableau des migrations donne en fait des résultats comparables que l'on transforme ou non la diagonale. Les facteurs restant invariants et ordonnés de la même façon, les coordonnées varient dans un rapport égal à la racine carrée du rapport des valeurs propres. Les caractères "·" représentent les zones géographiques du tableau symétrisé et sont les centres de gravité des identificateurs de ces zones pondérées par les termes marginaux. Le plan 1×2 (Fig. 3) met bien en évidence certains déséquilibres migratoires : les départs (codes 2) sont plus nombreux que les arrivées (codées 1) dans les zones Valenton, Fresnes, Rungis, et moins nombreux dans les zones Charenton, Thiais et Ivry.

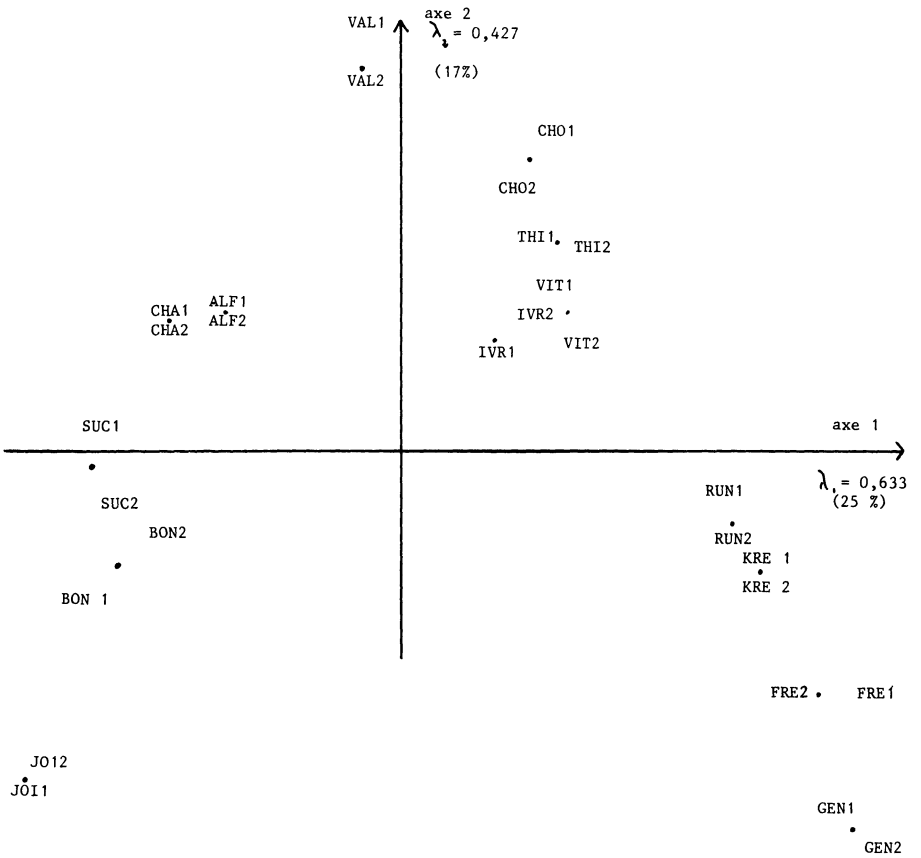


Figure 3. – Migrations interurbaines
{ 1 : arrivées
{ 2 : départs
{ • : points moyens
ORLY et VITRY sont confondus.

4.3. Analyse de l'évolution du commerce extérieur

Nous avons effectué l'analyse de la suite des tableaux du commerce extérieur intra-CEE de 1973 à 1980 pour tenter de reconstituer les échanges de 1981 en projetant les facteurs.

La figure 4 représente le plan principal 1×2 calculé sur le tableau moyen (contribution de 65 % a l'inertie totale). Pour faciliter la lecture du schéma, nous nous sommes limités à projeter les profils définis par les exportations de chaque tableau et numérotés de 1 (1973) à 8 (1980). Figurent aussi sur ce schéma les exportations réelles de 1981 (point 9), reconstruites par l'algorithme R.A.S. initialisé par le tableau de 1980 (point 10) et obtenues en régressant les facteurs (point 11). Les diagonales ont été évidemment modifiées.

Pour effectuer cette régression, nous avons conservé les trois premiers facteurs de chaque tableau et obtenons les vecteurs et valeurs propres suivantes :

	1	2	3
1	0,317	1,601	0,098
2	0,405	– 0,550	0,677
3	0,485	– 0,742	– 0,204
4	0,388	– 1,297	0,404
5	0,528	– 0,039	– 0,194
6	– 1,462	– 0,259	– 1,973
7	– 4,989	0,437	3,773
8	– 0,822	– 0,735	1,048
λ	0,0714	0,0444	0,0235

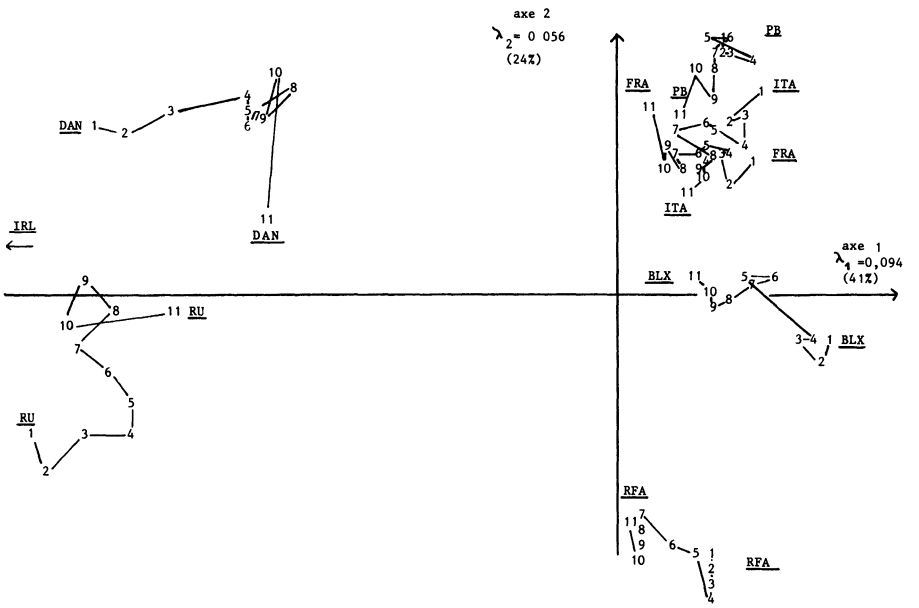


Figure 4. – Exportations INTRACEE de 1973 (1) à 1980 (8)
en suppl. { 10 : exportations en 1981 (RAS sur 1980)
11 : exportations en 1981 (Projections des facteurs)
9 : exportations réelles en 1981.

Effectuer une simple régression des moindres carrés pour obtenir une prévision des facteurs est évidemment discutable : aucune hypothèse n'est faite sur le caractère temporel des données (stationnarité par exemple). Compte tenu des contraintes d'orthonormalité qu'ils doivent vérifier, les facteurs ne peuvent être stables. Il ne faut toutefois pas perdre de vue que le but de la procédure n'est que de construire un "squelette" à partir duquel on reconstruira un tableau d'échanges possédant des marges fixées.

Comme nous l'avons déjà précisé, la formule de reconstruction donne un tableau dont les termes diagonaux ne sont pas nuls; pour obtenir la prévision définitive, nous les avons mis à zéro avant d'appliquer l'algorithme R.A.S., les marges

TABLEAU 3

Echanges commerciaux intra CEE prévus par projection des facteurs

	RFA	FRA	ITA	P-B	BLX	R-U	IRL	DAN	Total
RFA	0	19 095	12 084	12 905	10 217	10 951	782	2 312	68 345
FRA	17 335	0	6 610	5 370	7 411	8 039	180	1 022	45 968
ITA	10 786	6 500	0	3 160	4 242	3 256	472	673	29 088
P-B	18 225	8 354	4 999	0	6 337	6 178	482	804	45 379
BLX	10 523	8 408	4 895	4 622	0	4 011	429	898	33 785
R-U	10 245	8 285	3 413	4 093	3 644	0	4 453	1 640	35 772
IRL	557	141	377	243	297	3 390	0	231	5 236
DAN	2 015	982	657	496	760	1 528	283	0	6 722
Total	69 686	51 765	33 035	30 888	32 907	37 353	7 081	7 579	270 295

TABLEAU 4

Echanges commerciaux intra CEE obtenus par ajustement des échanges observés en 1980

	RFA	FRA	ITA	P-B	BLX	R-U	IRL	DAN	Total
RFA	0	18 533	12 534	13 091	10 171	10 544	578	2 894	68 345
FRA	15 957	0	10 419	3 923	7 428	7 159	468	613	45 967
ITA	10 987	10 003	0	1 728	1 726	4 080	174	391	29 088
P-B	18 994	6 583	3 336	0	8 981	5 859	360	1 266	45 379
BLX	10 470	9 117	2 555	6 536	0	4 402	182	524	33 785
R-U	10 124	6 195	3 348	4 818	4 181	0	5 258	1 848	35 772
IRL	634	616	192	291	185	3 275	0	43	5 236
DAN	2 521	719	651	502	235	2 033	61	0	6 722
Total	69 687	51 765	33 035	30 888	32 907	37 353	7 081	7 579	270 295

fixées étant celles de 1981. Le tableau 3 permet de comparer les estimations aux valeurs réelles données précédemment :

La prévision n'est pas très satisfaisante. On pourrait regretter de n'avoir considéré que les trois premiers axes, mais l'examen des quatrième et cinquième facteurs laisse penser que nous aurions rencontré quelques difficultés dans leur projection à l'horizon fixé. Les valeurs propres associées ont en effet tendance a se confondre, ce qui a pour conséquence :

i) une permutation éventuelle des facteurs de même interprétation

ii) une instabilité numérique dans le calcul des vecteurs propres (valeurs propres doubles)

Pour mesurer la précision des estimations, nous avons simplement ajusté le tableau observé en 1980 aux marges du tableau 1981 (Tabl. 4). Les marges des trois tableaux étant alors les mêmes, les distances du χ^2 sont comparables. L'ajustement du dernier tableau observé donne, suivant ce critère, un bien meilleur résultat (0,002 au lieu de 0,06). Cette conclusion est d'ailleurs visible sur la figure 4, sur laquelle les points codés 10 (R.A.S. sur le tableau de 1980) sont systématiquement plus près des points codés 9 (tableau réel) que les points codés 11 (R.A.S. après projection des facteurs).

CONCLUSIONS

Les méthodes de description développées dans cet article ont donné de bons résultats dans la description des tableaux de données que nous avons étudiés. Elles sont simples d'emploi, faciles à interpréter, et le calcul des termes diagonaux ne présente aucune difficulté. Une simple manipulation de fichiers suffit pour pouvoir utiliser un logiciel classique d'analyse des correspondances.

Si, dans le cas de données chronologiques, une tendance est mise en évidence, cela ne signifie pas que la projection des facteurs va permettre une bonne prévision des échanges à un horizon choisi, même si les marges sont fixées. Dans l'exemple que nous donnons comme dans les autres données que nous avons traitées, l'ajustement du dernier tableau observé aux marges fixées donne de meilleurs résultats que l'ajustement du tableau obtenu en appliquant la formule de reconstitution des données a partir des projections des facteurs. A ce titre, le problème reste ouvert : comment intégrer la nature temporelle des données, que l'on retrouve de façon évidente dans les analyses, à une méthode de prévision d'une suite de tableaux de contingence ?

C'est l'une des questions auxquelles réfléchit un groupe de travail du GRECO 59 Analyse des Données et Informatique, dont les discussions sont à l'origine de certaines des idées exposées dans cet article.

BIBLIOGRAPHIE

J.P. BENZECRI (1973). – *L'analyse des données tome 2 l'analyse des correspondances,* p. 231, 243 Dunod, Paris.

A.G. CONSTANTINE, J.C. GOWER (1978). – *Graphical representation of asymmetric matrices Applied Statistic,* 27, 3 p. 297, 304.

B. ESCOFIER (1981). – Quelques indices de comparaison de tableaux de contingence. *Statistique et analyse des données,* Vol. 6, n° 1, p. 39, 51.

Y. ESCOUFIER, A. GRORUD (1979). – Analyse factorielle des matrices carrées non symétriques, *2e Journées Internationales. Analyse des données et informatique.* Editeur North Holland.

T. FOUCART (1978). – Sur les suites de tableaux de contingence indexés par le temps. *Statistique et analyse des données,* n° 2, p. 67, 85.

J.B. KAZMIERCZAK (1978). – Migrations interurbaines dans la banlieue sud de Paris. *Cahiers de l'analyse des données,* Vol. 3 n° 2, p. 203. 218.

Y. Le FOLL, B. BURTSCHY (1983). – Représentation optimale des matrices imports-exports. Revue de Statistiques appliquées, Vol. 31, n° 3, p. 57, 72.

J.L. MADRE (1980). – Méthodes d'ajustement a des marges. *Cahiers de l'analyse des données,* Vol. 5, n° 1, p. 87, 100.

P.NIJKAMP, H. PAELINCK (1976). – Some methods for updating input-output tables. *Revue belge de statistique, d'informatique et de recherche opérationnelle,* Vol. 16, n° 4, p. 17, 41.

E. STEMMELEN (1977). Tableaux d'échanges : description et prévision. *Cahiers du BURO série Recherche,* n° 28.

E. YAGOLNITZER (1977). – Comparaison de deux correspondances entre les mêmes ensembles. *Cahiers de l'analyse des données,* Vol. 2, n° 3, p. 251, 264.