

# STATISTIQUE ET ANALYSE DES DONNÉES

JEAN-JACQUES RIGAL

## **Classification par ajustements linéaires**

*Statistique et analyse des données*, tome 8, n° 1 (1983), p. 59-75.

[http://www.numdam.org/item?id=SAD\\_1983\\_\\_8\\_1\\_59\\_0](http://www.numdam.org/item?id=SAD_1983__8_1_59_0)

© Association pour la statistique et ses utilisations, 1983, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). 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/>

CLASSIFICATION PAR AJUSTEMENTS LINEAIRES

Un besoin et une méthode d'estimation  
de classes d'observations ajustées  
par des systèmes linéaires propres

Jean-Jacques RIGAL

Centre d'Analyse des Données  
U.E.R. de Droit et Sciences Economiques  
Université de Pau et des Pays de l'Adour  
64000 PAU UNIVERSITE

*Résumé : Cet article met en évidence, sur un exemple d'analyse des objectifs d'entreprise, un besoin de classification telle qu'une relation linéaire ajuste les observations au sein de chaque classe. A cet effet, une méthode itérative "estimation-classification" est proposée pour la détermination de tels groupes et de leurs relations. Les propriétés de convergence et de stabilité sont étudiées à partir d'observations simulées.*

*Abstract : This study brings out, from an example of analysis about firm objectives, a need for classifying observations such that a linear relation fit the observations into each class. An iterative method "estimation classification" is proposed to determine groups and their relations. The convergence stability properties are studied from simulated observations.*

*Mots-clés : Classification, Relations linéaires simultanées, Objectifs d'entreprise.*

0 - INTRODUCTION

Dans le cadre d'une étude statistique des modes de comportement des entreprises, nous avons dû prolonger les outils habituels d'analyse que

constituent les régressions, les analyses en composantes principales, les classifications, ... En effet, nous souhaitons faire apparaître des classes d'observations voisines, non pas sur la base des résultats observés, mais par rapport à des paramètres descriptifs d'un modèle de comportement liant ces résultats.

Cette description statistique pourrait être formulée de la façon suivante :

- soit un ensemble E d'éléments e, décrits par n + p variables,
  - on suppose l'existence de M modèles linéaires "expliquant" n variables Y par l'intermédiaire de la totalité des n + p variables :
- $$A.Y = B.X + u$$
- on suppose encore que chaque élément de E relève de l'un de ces M modèles inconnus,
  - on cherche à estimer simultanément, les modèles linéaires (M matrices A et B), et le partitionnement de E en M classes.

Ce problème de classification présente un caractère particulier. On recherche un partitionnement sur un ensemble dont on ne connaît pas les mesures des éléments, celles-ci étant les coefficients d'un modèle rattaché à la classe de l'élément. Nous sortons ainsi du cadre habituel des méthodes de classification et de modélisation linéaire, pour utiliser, dans un processus itératif, une combinaison de ces deux techniques.

Précisons d'abord l'intérêt pratique d'une telle méthode, par référence à une étude des comportements d'entreprise (11 - J.J. RIGAL Thèse). Nous proposerons ensuite le schéma d'une méthode itérative susceptible de mettre en évidence les groupes et les modèles. Enfin, dans un troisième temps, nous donnerons les principaux résultats d'une étude de cette méthode effectuée à partir de simulations d'observations, et de ces applications à une approche des comportements d'entreprise.

## 1 - UN BESOIN ET UNE METHODE DE CLASSIFICATION

Les études récentes du comportement d'entreprise se caractérisent, en particulier, par la prise en compte de variables privilégiées, spécifiées comme les objectifs de la firme. (Par variable, nous entendons une mesure

des résultats de l'entreprise. Ce sont généralement des résultats comptables annuels. Mais d'autres mesures peuvent être retenues, comme par exemple : l'effectif, un indicateur de croissance, ...). Au plan économétrique, elle cherche à mettre en évidence des relations "explicatives" des "variables objectifs" par des "variables instrumentales" et par des "variables externes" (cf. §2).

Diverses approches, théoriques ou statistiques, du comportement de l'entreprise admettent l'existence d'indicateurs, variés mais limités susceptibles de prendre en compte les objectifs des décideurs de l'entreprise. Ainsi en est-il du profit et du cash flow, du risque et de la croissance, de la survie et de la rentabilité, ... chacun de ces indicateurs pouvant prendre des formes multiples.

Or, pour de telles caractéristiques, l'analyse des résultats individuels des entreprises met en évidence une dispersion otant toute signification à d'éventuelles classes. Ainsi, dans l'étude citée (11), où nous avons retenu trois indicateurs d'objectif : la part de marché, le risque, et la rentabilité, nous avons pu montrer que la dispersion spatiale des entreprises ne révèle pas de classes homogènes (11 - 1ère partie). Une classification automatique de type visuel aurait permis, tout au plus, de fixer des types de réalisation des objectifs, types représentatifs de la variété individuelle, certes, mais non nécessairement porteurs de l'information suffisante à une analyse de la formation des ces résultats.

Pour ces raisons, nous avons retenu un schéma de liaison entre les variables objectifs et les variables de "stimuli" fondé sur un système d'équations simultanées et linéaires, où chaque relation "explique" une variable objectif à partir de certaines des variables (autres objectifs, instruments et variables externes. Une des premières applications significatives de ce schéma d'analyse a été consacrée à l'étude de comportements financiers d'entreprises (MUELLER 1967 (9)). On peut également citer quelques études du même type : (1), (12) et (13).

Dans la notation introductive  $A.Y = B.X + u$ , le vecteur  $Y$  est celui des objectifs, alors que le vecteur  $X$  est celui des variables instruments et des variables externes ; les matrices  $A$  et  $B$  sont alors celles des coefficients des variables spécifiées dans chaque relation.

Le choix d'une liaison linéaire est bien entendu sujet aux critiques usuelles, mais il a été fait en l'absence de connaissance sérieuse de relations concurrentes de comportement.

Mais l'estimation de relations explicatives de la formation de variables objectifs n'est pas aisée car, d'une part elle doit prendre en compte la variété spatiale des comportements (les entreprises n'ont pas des relations identiques), et d'autre part le comportement de l'entreprise -et donc ses relations propres- peut évoluer avec le temps.

On se trouve confronté à une "analyse des données individuelles" insoluble sous cette forme générale, et qui nécessite des hypothèses restrictives quant aux coefficients des relations à estimer. (On trouvera dans les actes du Colloque INSEE-CNRS (4) de multiples études qui, tant sur le plan méthodologique que sur le plan des applications, relèvent de cet axe de recherche). C'est ainsi que nous avons supposé (11), qu'à un instant donné, il existe des classes d'entreprises à l'intérieur desquelles les relations de comportement sont "voisines", ce qui autorise l'estimation des relations propres à chaque classe. La connaissance de ces relations permet d'indiquer des types de comportement, ou, au pire, elle donne une image de la variété des comportements individuels.

Nous montrerons au § 3-2 l'éclairage qu'une telle analyse des données peut apporter à la connaissance des comportements individuels d'entreprises.

## 2 - SCHEMA DE LA METHODE ITERATIVE "ESTIMATION-CLASSIFICATION"

Chaque élément  $e$  de l'ensemble  $E$  est décrit par  $n$  variables endogènes (notées  $Y$ ) et par  $p$  variables exogènes (notées  $X$ ). Par rapport à l'étude que nous venons d'esquisser à titre d'exemple, les variables  $X$  regroupent les variables instruments, dont le niveau est décidé par l'entreprise : investissements, ..., et des variables externes, dont le niveau est subi par l'entreprise : marché financier, résultats antérieurs, ...). On suppose l'existence d'un partitionnement de  $E$  en  $M$  classes  $C_m$  tel que les variables  $X_e$  et  $Y_e$  de chaque  $e$  de la classe  $C_m$  soient liées par un modèle  $(A_m, B_m)$  unique pour la classe :

$$A_m \cdot Y_e = B_m \cdot X_e + u_e$$

Il nous faut donc déterminer une méthode d'estimation qui, à partir des observations  $X_e$  et  $Y_e$  pour tout  $e$  de  $E$ , détermine les  $m$  modèles  $(A_m, B_m)$ , où  $A$  et  $B$  sont des matrices respectivement  $(n,n)$  et  $(n,p)$ , avec des restrictions possibles de spécification de certains coefficients. Certains coefficients ont une valeur spécifiée avec le modèle. Ainsi la diagonale de  $A$  est formée d'unités (la variable expliquée  $Y_i$  est dotée du coefficient 1). D'autres variables peuvent être exclues de la spécification d'une variable  $Y_i$  (relation  $i$ ), ou mieux, elles sont spécifiées avec un coefficient nul (cf. Annexe 1 où huit coefficients sont dans ce cas). D'autres restrictions pourraient être envisagées qui compliqueraient l'estimation : symétrie de  $A$ , égalité de certains coefficients de  $B$ ,...

Nous avons retenu et programmé une méthode itérative, succession d'estimations de modèles relatifs à des classes différentes de celles recherchées, et de classification "autour" des modèles ainsi mis en évidence. Cette approche se situe dans le prolongement des travaux de l'équipe de E.DIDAY où les principes de classification automatique sont étendus à des recherches locales de régression, de discriminations, d'analyses factorielles,..(3-4).

Deux problèmes sont alors posés. Le premier, usuel, est celui de l'estimation d'un système linéaire pour les éléments d'une classe fixée de  $E$ . Le second est de déterminer des classes "autour" des modèles estimés sur les classes de l'itération précédente.

## 2.1 - L'estimation d'un système linéaire

L'estimation d'un système linéaire est chose usuelle dès que les coefficients sont certains, constants et que la seule restriction de spécification est la nullité de certains d'entre eux (une variable  $X_j$  ou  $Y_j$  est supposée sans effet sur la variable  $Y_i$  de la relation  $i$  : le coefficient  $B_{ij}$  est supposé nul). Ce problème d'estimation est traité de façon courante (4),(6),(8),(14) et divers programmes informatiques permettent le calcul des estimations.

Nous avons utilisé un programme mis au point au Laboratoire d'Analyse et de Recherche Économique de l'Université de BORDEAUX I, adapté de (7). Ce programme constitue un des modules de notre logiciel d'estimation.

Diverses méthodes d'estimation sont utilisables : les moindres carrés ordinaires (biaisés), les moindres carrés indirects, les doubles moindres carrés, le maximum de vraisemblance, et les triples moindres carrés. Notons que les estimateurs de la "classe-k" de THEIL regroupent les MCO, DMC et M3. Notons également que les triples moindres carrés prennent en compte une corrélation éventuelle des résidus des relations du système. Nous avons retenu celle des doubles moindres carrés, qui présente l'avantage d'être la moins volumineuse en mémoire centrale d'ordinateur, et ce bien que dans le cas d'une liaison stochastique entre les résidus des n relations du système, la méthode des triples moindres carrés soit la seule valide.

## 2.2 - La classification "autour" de M modèles.

Notre problème sort du cadre usuel des méthodes de classification, où l'on forme des groupes d'éléments e de E sur la base de leurs distances réciproques. Dans notre cas, deux éléments e de E ne sont pas proches quand leur distance dans  $R^{n+p}$  est faible, mais quand leurs écarts à un même modèle sont faibles.

Il nous faut donc préciser la mesure de la proximité entre un élément e de E (caractérisé par n+p variables) et un modèle (caractérisé par les coefficients de deux matrices A(n,n) et B(n,p)). Le résidu observé pour e est :  $u_e = A.Y_e - B.X_e$ , c'est-à-dire pour la relation i :

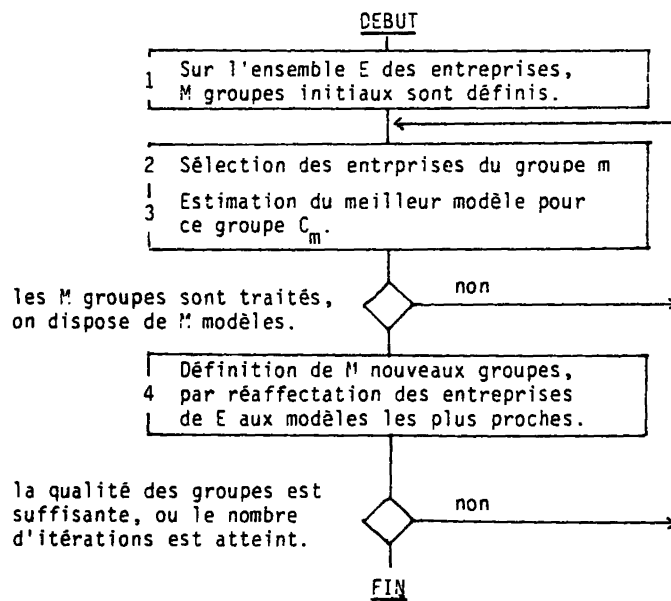
$$u_{ei} = \sum_{j=1,n} A_{ij} \cdot Y_{ej} - \sum_{j=1,p} B_{ij} \cdot X_{ej}$$

Une norme de  $u_e$  pourrait donc caractériser cet écart au modèle (A,B). En pratique, les résidus présentant, selon les variables objectifs, des variances très dissemblables, nous avons retenu, non pas le vecteur résidu  $u_e$ , mais le vecteur  $v_e$  obtenu en rapportant chaque résidu  $u_{ei}$  à la valeur de la variable objectif  $y_{ei}$  :  $v_e' = (\dots, u_{ei}/y_{ei}, \dots)$ . Ainsi, alors que la norme euclidienne de  $u_e$  aurait introduit un écart de la forme :  $u_e' \cdot u_e$ , la norme retenue considère l'écart :  $u_e' \cdot (D_e^{-1})^2 \cdot u_e$ , où  $D_e$  est la matrice diagonale des  $Y_{ei}$  ( $v_e = D_e^{-1} \cdot u_e$ ). Ce choix permet de pondérer les résidus selon les relations du modèle. Mais d'autres techniques pourraient être envisagées : pondération par la moyenne ou par la dispersion de la variable endogène sur la population ou sur une classe (associée au modèle par exemple), pondération par la dispersion des résidus pour la classe du modèle,...

Ainsi la mesure de l'écart entre une entreprise et un modèle nous a parue s'imposer pour que les classes et les modèles estimés aient, tout au long du processus itératif une signification en référence à l'ajustement des objectifs par des variables exogènes. On doit toutefois objecter que d'autres mesures de cet écart pourraient peut-être autoriser une convergence plus rapide vers des groupes et des modèles limites tout aussi significatifs. Ce sont là des questions qui restent posées et auxquelles nous donnerons des éléments de réponse au § 3.

2.3 - Le schéma de l'estimation itérative : "estimation-classification"

Sur les bases de ces deux principes, d'estimation par les doubles moindres carrés d'un modèle ajustant un groupe donné d'observations, et de mesure de la proximité d'une observation et d'un modèle, on peut entreprendre le processus itératif décrit ci-après.





On retrouve ici les préoccupations naturelles de nombreux chercheurs qui ont anticipé les besoins d'estimation de classes sur la base de critères autres que celui du voisinage des observations. Ainsi, dans un ordre d'idée comparable au nôtre Y. OK-SAKUN (10) propose l'estimation de classes qui minimisent l'inertie par rapport à des composantes principales propres aux classes. Quant aux régressions typologiques, les travaux portent, à notre connaissance, exclusivement sur des régressions temporelles, ne répondant pas ainsi à notre besoin (cf. par exemple (2)).

### 3 - APPLICATION ET TESTS

Il est bien évident que la possibilité de recherche de classes et de leurs modèles associés est liée à l'hypothèse de leur existence. En pratique, si les éléments de l'ensemble à partitionner ont chacun un modèle liant leurs caractères observés, la méthode proposée devrait tendre à indiquer des classes regroupant des éléments dont les modèles seraient voisins.

Nous proposons ici deux tests par simulations, qui permettent de porter un jugement sur les propriétés de stabilité et de convergence de la méthode ci-dessus esquissée. Nous proposons également un résumé de résultats obtenus dans l'étude (11) d'une typologie des comportements d'entreprise et montrant que, malgré l'imperfection de la convergence de la méthode, on peut mettre en évidence certaines propriétés comportementales utiles à l'orientation de recherches ultérieures.

#### 3.1 - Deux tests par simulation d'observations

Deux tests sur des données simulées peuvent permettre d'approcher les propriétés de stabilité et de convergence de la méthode proposée. Dans un premier test, nous avons engendré des observations relevant exactement de modèles fixés. Son principal intérêt est de montrer la convergence sur un cas limite. Dans un second test, nous avons engendré des observations en introduisant un terme aléatoire additionnel sur les variables endogènes (objectifs). Nous pourrions ainsi prendre la mesure de l'imperfection de la convergence.

#### Simulation 1.

Nous avons retenu cinq modèles qui décrivent trois variables endogènes à partir de cinq variables exogènes, sous la forme :  $A.Y = B.X$  (cf. annexe 1). A partir de ces cinq modèles, nous avons simulé 250 observations (50 par modèle), sans terme aléatoire pour les trois variables endogènes : ayant engendré une observation exogène  $X_e$ , l'observation  $Y_e$  est calculée par  $Y_e = A^{-1}.B.X_e$ .

Après avoir constitué des groupes arbitraires, et en utilisant l'estimation des doubles moindres carrés, le processus itératif a fait apparaître deux des groupes "vrais" à la 23ième itération. (On trouvera en annexe 2 le graphique d'évolution de la troisième variable tout au long du processus itératif).

On observe, sur cet exemple, la convergence de la méthode d'estimation vers les modèles et les groupes originels. La stabilité est donc obtenue pour ces observations qui répondent exactement à des modèles de référence.

#### Simulation 2.

Toujours sur la base de cinq modèles initiaux, nous avons engendré mille observations (200 par modèle) en introduisant un terme additif aléatoire pour chacune des trois variables endogènes :  $Y = A^{-1}.B.X + u$ . Le terme aléatoire additionnel  $u$  porte sur les variables endogènes  $Y$ . Chacune des composantes de  $u$  est distribuée normalement, autour de zéro, avec un écart type égal au dixième de la moyenne de la variable endogène correspondante.

Ici, et contrairement à la précédente simulation, on ne retrouve pas totalement les groupes et les modèles vrais qui ont servi à l'élaboration des observations.

- Toutefois, on peut noter que les groupes formés après 80 itérations sont constitués de façon dominante par les groupes vrais (cf. Annexe 3).
- Quant aux modèles mis en évidence sur ces classes, l'analyse de l'évolution de chaque coefficient confirme l'imperfection de la convergence (cf. Annexe 4).

Au total, on constate donc que la convergence et la stabilité dépendent de conditions qui restent à établir. La proximité des modèles générateurs des observations, et surtout le caractère direct de la déformation des groupes par simple réaffectation, ne permettent pas une estimation

nécessairement convenable. Cette insuffisance devrait conduire à définir un mode plus fin de déformation des groupes, qui tienne compte de la permanence des groupes et des modèles lors des itérations successives. Le tracé de l'évolution des coefficients au long du processus itératif montre, par exemple, la nécessité d'éliminer les écarts importants lors d'une itération, par exemple par un procédé de moyenne mobile, ou par l'abandon de valeurs extrêmes.

### 3.2 - Une application à l'étude des comportements d'entreprise

Nous avons indiqué au § 1.1 l'impossibilité de révéler des classes d'entreprises homogènes par rapport à des variables de comportement. Lors de cette étude (11 - lière partie), on constatait également le peu de liaisons entre les objectifs et les "stimuli" (constitués par des variables instruments, sur lesquelles l'entreprise peut agir, et par des variables externes, prenant en compte son contexte économique). Nous avons alors spécifié une structure de formation des objectifs sur la base de trois relations linéaires simultanées (11 - Ch.IV).

Il était bien évident, dès l'abord, qu'un schéma explicatif universel ne serait pas significatif, et qu'il faudrait estimer des classes d'entreprises homogènes par rapport à des modèles propres. Quels sont les principaux résultats de cette investigation, tout au moins ceux éclairant l'intérêt et les limites de la méthode ici proposée ?

#### Modèles et groupes homogènes

Sur un échantillon de plus de 1200 entreprises, nous avons appliqué la méthode itérative successivement à 26, 7 et 4 secteurs. La déformation des 26 secteurs a conduit à fixer 9 groupes "homogènes", alors que dans les deux autres cas, 7 et 4 groupes pouvaient être estimés. Le passage de 26 à 9 groupes est dû à l'abandon d'un groupe dès que son effectif se révèle insuffisant à l'estimation d'un modèle, les entreprises sont alors affectées à d'autres groupes, c'est-à-dire à d'autres modèles.

Remarquons, à cet égard, que les estimations sectorielles montrent des ajustements qui nient toute homogénéité aux secteurs : ici encore le produit ne détermine pas le comportement (11 - Annexe 7.2).

Comme lors du second test, on a pu constater que le mode de déformation des groupes ne favorisait pas la convergence. Ainsi, pour les neuf groupes homogènes estimés à partir des 26 secteurs, à la 20ième itération les entrées et les sorties des groupes sont encore de l'ordre de la moitié de l'effectif du groupe considéré (11 - Annexe 7.1, p.22).

Toutefois, sur la base de critères de distance entre modèles interprétables en termes de comportement (11 - Ch. IV-A1), on a pu considérer les modèles estimés comme une représentation satisfaisante de la variété des comportements individuels.

L'absence d'unicité du comportement est confirmée par l'estimation du modèle sur l'ensemble de la population : aucun objectif n'est convenablement ajusté, peu de coefficients sont significativement différents de zéro.

Par opposition, l'analyse des modèles types, obtenus pour les groupes homogènes et pour divers partitionnements, montre la permanence du sens de l'influence de certaines variables. Ainsi, un comportement assez constant peut être retenu pour les deux objectifs de part de marché (PM) et de risque (RI), comportement que l'on pourrait symboliser :

$$PM = +RI +DT - CA -S/VA$$
$$RI = +FP +S/VA$$

- La part de marché croît quand le risque diminue (RI croît) et quand les dettes de trésorerie (DT) croissent, elle décroît avec l'appartenance à un secteur à fort volume d'affaires (ZCA) ou à forte part des salaires dans la valeur ajoutée (S/VA).
- Le risque baisse (RI croît) quand croissent les fonds propres (FP), et pour les secteurs où S/VA est élevé.

#### La variété des comportements

Au-delà de ces traits de comportement, limités mais universels, l'analyse de la variété des modèles types associés aux groupes "homogènes" des divers partitionnements estimés, a fait ressortir des traits de comportement plus précis mais variés, ainsi que des interdépendances entre objectifs.

On a pu constater l'importance des variables externes dans la formation des objectifs. Ainsi, le ratio sectoriel médian des salaires sur valeur ajoutée (S/VA) intervient de façon significative, importante et variée dans chacun des trois objectifs ou encore, le coût sectoriel médian de la dette (frais financiers/dettes) est prépondérant dans la formation du risque.

Pour l'objectif de part de marché, on a montré l'importance des variables instruments alors qu'elles sont peu influentes pour le risque, sensible aux variables externes.

Les interdépendances entre objectifs sont l'objet de considérations théoriques en termes de substitution, d'utilité, d'optimisation conjointe,... De ce point de vue, nous avons interprété la dissymétrie de la matrice A des coefficients des objectifs en considérant trois formes de comportement (a et b positifs) (11 - V8 p. 408) :

- Des objectifs  $O_1$  et  $O_2$  sont conjoints quand :  $O_1 = aO_2 + \dots$  et  $O_2 = bO_1 + \dots$
- Des objectifs sont substituables, quand :  $O_1 = aO_2 + \dots$  et  $O_2 = -bO_1 + \dots$
- Un objectif est privilégié par rapport à l'autre, quand :  $O_1 = +aO_2 + \dots$  et  $O_2 = -bO_1 + \dots$   
:  $O_1 = -aO_2$  et  $O_2 = +bO_1 \dots$

$O_1$  est privilégié par rapport à  $O_2$  quand l'amélioration de  $O_2$  ne contraint pas celle de  $O_1$  ( $O_1 = + aO_2 + \dots$ ) mais que l'amélioration de  $O_2$  contraint celle de  $O_1$  ( $O_2 = - bO_1 + \dots$ ).

Selon cette interprétation, on a constaté :

- Pour la part de marché et risque :
  - 4 fois sur 10 : ce sont des objectifs conjoints,
  - 1 fois sur 2 : la part de marché est privilégiée par rapport au risque,
  - 1 fois sur 10 : le risque est privilégié par rapport à la part de marché.
- Pour la part de marché et rentabilité :
  - 1 fois sur 3 : ce sont des objectifs conjoints,
  - 1 fois sur 4 : la part de marché est privilégiée,
  - 1 fois sur 3 : la rentabilité est privilégiée.
- Pour la part de risque et rentabilité :
  - 1 fois sur 2 : ce sont des objectifs conjoints,
  - 1 fois sur 4 : la rentabilité est privilégiée,
  - 1 fois sur 4 : ce sont des objectifs substituables.
- Notons que rarement :
  - . le risque est privilégié par rapport à la rentabilité,
  - . part de marché et risque sont substituables,
  - . part de marché et rentabilité sont substituables.

#### 4 - CONCLUSION

La méthode que nous venons de proposer paraît donc autoriser une classification sur un ensemble d'observations par l'estimation de systèmes linéaires descriptifs de ces classes.

Au plan méthodologique, elle ouvre une voie d'étude des hypothèses et des propriétés d'un tel processus itératif, principalement quant à sa convergence et à sa stabilité.

Au plan pratique, ces deux derniers impératifs se traduisent par la qualité des ajustements obtenus et par la permanence des groupes formés. Il sera donc nécessaire de développer divers outils portant sur la mesure de la qualité des résultats obtenus. On peut penser à : une pondération des résidus dans le cas de systèmes linéaires ou d'hétéroscédasticité, une indication d'un nombre de groupes a priori optimal, des indications sur les effectifs des groupes, des seuils de qualité des ajustements, des analyses de la permanence des groupes au long des itérations,...

Toutefois, et malgré des limites théoriques et pratiques, la détermination de modèles et de classes n'est pas sans intérêt. En effet, et sous réserve d'un examen attentif de la qualité des ajustements dans le contexte de l'analyse entreprise, la variété des modèles estimés donne une image de la diversité des modèles propres à chaque observation, autorisant ainsi, dans le domaine économique, des recherches spatiales sur des critères d'agrégation, et des recherches temporelles sur des problèmes de "modélisation dynamique".

#### 5 - REFERENCES BIBLIOGRAPHIQUES

- /1/ ~~BEEDLES~~ W. L. A micro econometric investigation of multi-objective firm - Journal of Finance - Sep. 1977
- /2/ 1- CHARLES C. Regression typologique - INRIA - Rapport Laboria N° 237 - 1977
- /2/ 2- CHARLES C. Pattern recognition by a piecewise polynomial approximation with variable joint - INRIA - Rapport Laboria N° 338 - 1979

- /3/ 1- DIDAY E. Introduction à l'analyse factorielle typologique - INRIA - Rapport LABOPIA N° 27
- /3/ 2- DIDAY E. Classification automatique séquentielle - MAIRO 1975
- /3/ 3- DIDAY E. Optimisation en classification automatique - INRIA - Collaborateurs 1978
- /3/ 4- DIDAY E. Problems of clustering and recent advances - INRIA - 1979 - N° 337
- /4/ GOLDBERGER A.S. Econometric Theory - WILEY - 1963
- /5/ INSEE-CNRS The econometrics of the panel data. Cf.: OUDIZ, EISNER, HARVEY, MUNDLAK, TROGNON. Annales INSEE N° 30-31 1980
- /6/ JOHNSTON J. Econometrics methods - Mc GRAWHILL - 1972
- /7/ LANGASKENS Y. Introduction à l'économétrie - DROZ - 1975
- /8/ MALINVAUD E. Methodes statistiques de l'économétrie - DUNOD
- /9/ MUELLER D.C. The firm decision process - Quaterly Journal of Economics - 1967
- /10/ OK-SAKUN Y. Analyse factorielle typologique et lissage typologique - Thèse 3ème cycle - PARIS VI - 1975
- /11/ RIGAL J. J. Analyse multidimensionnelle des entreprises - Recherche sur une typologie des comportements - Thèse d'Etat - BORDEAUX I - 1981
- /12/ SALTZMAN S. An econometric model of the firm - Review of Economics and statistics - 1967
- /13/ SIEKOWITZ J. A note of the simultaneous nature of finance methodology - Journal of Finance - Mars 1972
- /14/ THEIL Principles of econometrics - North Holland - 1971

6 - ANNEXES

Annexe 1 Les cinq modèles ayant servi à la simulation 1:  $Y = A^{-1}.B.X$

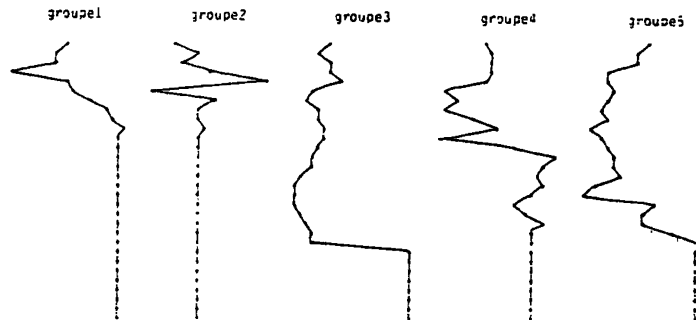
1	2	3
-1	1	-6
5	4	1

3.5	0.	0.8	0.	0.3
-0.35	0.	0.	-4.	0.
0.	0.5	0.	0.	0.85

modèle 1

<table border="1"> <tr><td>1</td><td>2</td><td>3</td></tr> <tr><td>-2</td><td>1</td><td>-4</td></tr> <tr><td>-7</td><td>1</td><td>1</td></tr> </table>	1	2	3	-2	1	-4	-7	1	1	<table border="1"> <tr><td>3.5</td><td>0.</td><td>0.8</td><td>0.</td><td>0.3</td></tr> <tr><td>-0.35</td><td>0.</td><td>0.</td><td>-4</td><td>0.</td></tr> <tr><td>0.</td><td>0.5</td><td>0.</td><td>0.</td><td>0.85</td></tr> </table>	3.5	0.	0.8	0.	0.3	-0.35	0.	0.	-4	0.	0.	0.5	0.	0.	0.85	modèle 2
1	2	3																								
-2	1	-4																								
-7	1	1																								
3.5	0.	0.8	0.	0.3																						
-0.35	0.	0.	-4	0.																						
0.	0.5	0.	0.	0.85																						
<table border="1"> <tr><td>1</td><td>2</td><td>-6</td></tr> <tr><td>8</td><td>1</td><td>2</td></tr> <tr><td>-7</td><td>-4</td><td>1</td></tr> </table>	1	2	-6	8	1	2	-7	-4	1	<table border="1"> <tr><td>-1.</td><td>0.</td><td>-2.</td><td>0.</td><td>-2.5</td></tr> <tr><td>0.2</td><td>0.</td><td>0.</td><td>0.45</td><td>0.</td></tr> <tr><td>0.</td><td>-0.25</td><td>0.</td><td>0.</td><td>1.5</td></tr> </table>	-1.	0.	-2.	0.	-2.5	0.2	0.	0.	0.45	0.	0.	-0.25	0.	0.	1.5	modèle 3
1	2	-6																								
8	1	2																								
-7	-4	1																								
-1.	0.	-2.	0.	-2.5																						
0.2	0.	0.	0.45	0.																						
0.	-0.25	0.	0.	1.5																						
<table border="1"> <tr><td>1</td><td>2</td><td>-6</td></tr> <tr><td>8</td><td>1</td><td>2</td></tr> <tr><td>-7</td><td>-4</td><td>1</td></tr> </table>	1	2	-6	8	1	2	-7	-4	1	<table border="1"> <tr><td>3.5</td><td>0.</td><td>0.8</td><td>0.</td><td>0.3</td></tr> <tr><td>-0.35</td><td>0.</td><td>0.</td><td>-4.</td><td>0.</td></tr> <tr><td>0.</td><td>0.5</td><td>0.</td><td>0.</td><td>0.85</td></tr> </table>	3.5	0.	0.8	0.	0.3	-0.35	0.	0.	-4.	0.	0.	0.5	0.	0.	0.85	modèle 4
1	2	-6																								
8	1	2																								
-7	-4	1																								
3.5	0.	0.8	0.	0.3																						
-0.35	0.	0.	-4.	0.																						
0.	0.5	0.	0.	0.85																						
<table border="1"> <tr><td>1</td><td>-3</td><td>-1</td></tr> <tr><td>3</td><td>1</td><td>2</td></tr> <tr><td>6</td><td>8</td><td>1</td></tr> </table>	1	-3	-1	3	1	2	6	8	1	<table border="1"> <tr><td>3.5</td><td>0.</td><td>0.8</td><td>0.</td><td>0.3</td></tr> <tr><td>-0.35</td><td>0.</td><td>0.</td><td>-4.</td><td>0.</td></tr> <tr><td>0.</td><td>-0.5</td><td>0.</td><td>0.</td><td>0.85</td></tr> </table>	3.5	0.	0.8	0.	0.3	-0.35	0.	0.	-4.	0.	0.	-0.5	0.	0.	0.85	modèle 5
1	-3	-1																								
3	1	2																								
6	8	1																								
3.5	0.	0.8	0.	0.3																						
-0.35	0.	0.	-4.	0.																						
0.	-0.5	0.	0.	0.85																						

**Annexe 2** Un exemple d'évolution d'un coefficient des modèles, pour cinq groupes initiaux et pour 40 itérations (cas de la variable 13, 2ième exogène de la 3ième équation).



**Annexe 3** Effectifs des croisements entre groupes "estimés" (verticaux) et groupes "vrais" (horizontaux). Itérations 75 à 80.

3.a - Résultats à l'itération 80, avec arrangement des colonnes.

groupes estimés		2	5	1	4	3
groupes vrais	1	70	51	31	2	46
	2	15	64	6	14	1
	3	27	82	30	51	10
	4	71	14	16	85	14
	5	56	92	7	8	37



## 3.b - Résultats des itérations 75 à 80

itération 75						it. 76					
groupes estimés						groupes estimés					
	1	2	3	4	5		1	2	3	4	5
1	1	13	89	15	82	1	5	32	84	15	84
2	5	40	1	3	151	2	7	30	0	2	151
3	94	19	1	35	51	3	91	36	0	25	48
4	16	14	17	134	10	4	13	39	19	104	21
5	4	27	87	4	74	5	8	8	92	1	91

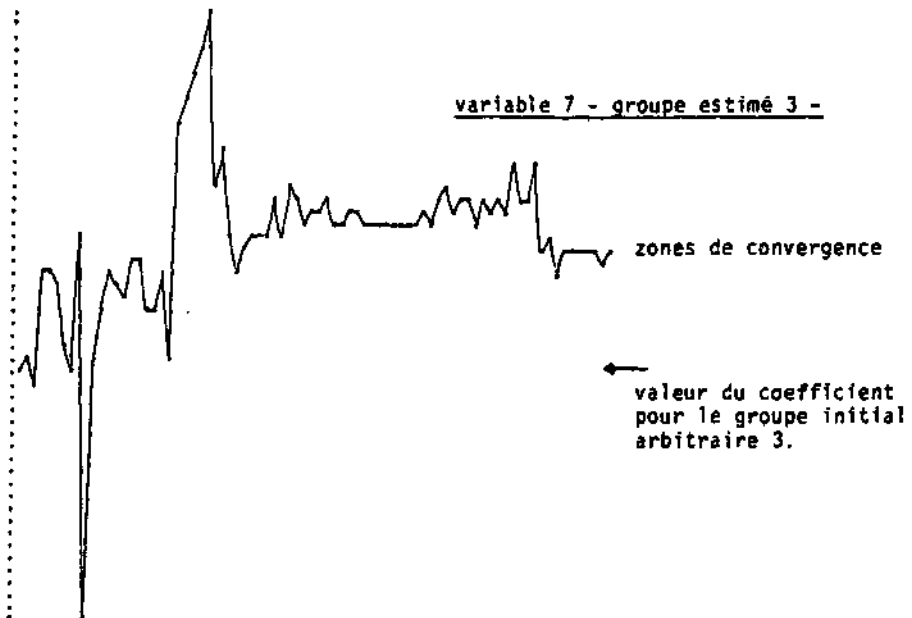
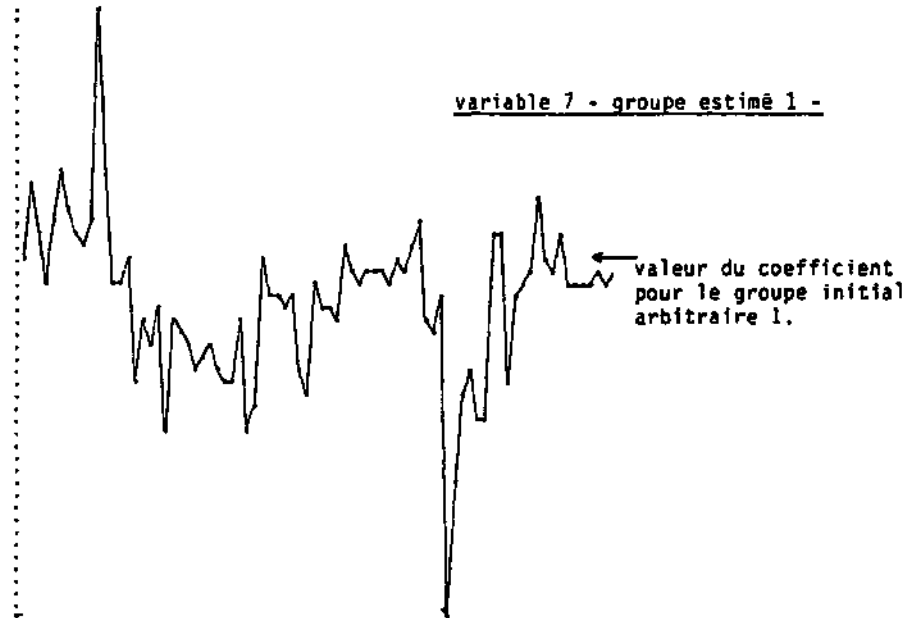
it. 77						it. 78					
groupes estimés						groupes estimés					
	1	2	3	4	5		1	2	3	4	5
1	1	40	62	14	78	1	4	24	71	2	97
2	4	25	2	1	168	2	4	17	0	7	170
3	47	64	1	11	24	3	70	46	1	51	32
4	20	27	7	121	21	4	17	15	4	149	15
5	4	15	84	7	93	5	8	29	72	0	91

it. 79						it. 80					
groupes estimés						groupes estimés					
	1	2	3	4	5		1	2	3	4	5
1	21	22	58	0	99	1	31	70	46	2	51
2	5	14	0	9	172	2	6	15	1	14	164
3	56	60	14	52	18	3	30	27	10	51	42
4	18	20	5	129	28	4	16	71	14	85	14
5	5	27	58	7	103	5	7	54	37	4	92

Annexe 4 Exemple d'évolution d'un coefficient (variable 7), pour deux groupes estimés (groupes 1 et 3).

- . Pour le groupe estimé 1, on ne constate pas de convergence, et ce bien que sur les dernières itérations une stabilité puisse s'amorcer (liée à la dominance du groupe vrai 3).
- . Pour le groupe estimé 3, une zone de convergence paraît se révéler, dont il faudrait mesurer l'amplitude, et ce bien que ce groupe estimé 3 soit principalement formé par les deux groupes vrais 1 et 5 (cf. annexe 3).



\*\*\*\*\*