

J.-P. BENZÉCRI

## L'analyse des données en économie

*Les cahiers de l'analyse des données*, tome 5, n° 1 (1980),  
p. 9-16

[http://www.numdam.org/item?id=CAD\\_1980\\_\\_5\\_1\\_9\\_0](http://www.numdam.org/item?id=CAD_1980__5_1_9_0)

© Les cahiers de l'analyse des données, Dunod, 1980, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

## L'ANALYSE DES DONNÉES

### EN ÉCONOMIE

[A.D. ÉCO]

par J.-P. Benzécri <sup>(1)</sup>

Dès l'origine, statistique et économie ont grandi ensemble : selon A. Piattier, le terme de statistique a été créé au XVIII<sup>e</sup> siècle, la référence à l'état *status* étant claire.

Dans le traité sur l'A. des D. on trouve deux études relevant de l'économie : "La construction des nomenclatures d'activités économiques de l'industrie "TIC n° 12, par M. Volle et coll. ; et "Les budgets familiaux dans les régions de la C.E.E." par N. Tabard. Les *Cahiers* ont publié plusieurs analyses de données issues de ce domaine ; citons : "L'évolution de la production industrielle française de 1963 à 1975" Vol I n° 4, par P. Teillard (il s'agit d'une analyse de séries chronologiques, entreprise sous la direction de M. Volle) ; deux études de V. Gouvêa sur les importations brésiliennes de machines et outils mécaniques (Vol II n° 3 et Vol III n° 3) ; un article de P. Clapier et J.L. Madre sur les budgets des ménages français (Vol IV n° 4). Enfin, plusieurs thèses et de nombreux rapports non publiés, traitent de l'extraction, de l'élaboration et de l'utilisation des matières premières (Maignant ; Chawoushi ; Mudiayl..) ; du marché de l'énergie (A. Fredj) etc. . Tous ces travaux nous intéressent en ce que les cartes issues de l'analyse factorielle offrent, de vastes ensembles de données, une vision globale fidèle qu'il serait impossible d'acquérir par une lecture, fût-elle attentive, des tableaux initiaux : les spécialistes eux-mêmes découvrent aussi des relations dont ils n'étaient pas avertis. Cependant, l'analyse dépend de données : telles qu'on les trouve publiées, celles-ci ne satisfont guère le statisticien ; il nous faut conjuguer de nombreuses sources ; remonter autant que possible aux documents non publiés. Et même ainsi, on a peine à démêler dans les nomenclatures actuelles (conçues pour une économie dépourvue des moyens de synthèse que seul l'ordinateur met en oeuvre), les bilans détaillés exhaustifs et homogènes dont l'analyse devrait dégager les termes et les lois d'une théorie nouvelle affranchie de l'*a priori*. Nourrie des expériences des analyses déjà faites, la présente note, propose aux chercheurs une esquisse des retouches que nous suggérons d'effectuer.

Au § 1, cheminant d'exemple en exemple, nous suggérons de constituer puis d'analyser des tableaux détaillés de flux de biens et services ; en faisant à l'homme dans ces bilans une place rigoureusement définie (§ 1.4). Au § 2 (qui sera publié ultérieurement) nous considérerons les différentes fonctions de la monnaie ; la définition statistique d'indices de valeur étant selon nous un thème urgent de recherche (§ 2.4).

1.1 Macroéconomie et nomenclature : Dans sa thèse (3<sup>e</sup> cycle Paris 1973), Jean Robert consacre un chapitre à l'analyse des sources de crédit auxquelles ont recours les différentes branches de l'industrie. Il semblerait

(1) Professeur de statistique. Université Pierre et Marie Curie.

possible de mener une telle étude en ne distinguant que quelques dizaines de branches : c'est-à-dire de se borner à la nomenclature à deux chiffres de l'INSEE (nomenclature en 27 branches, telles que : construction navale ; industrie chimique et caoutchouc...). Toutefois, disposant de statistiques détaillées donnant les sources de crédit (à court, moyen ou long terme) utilisées par plus de 100 branches (nomenclature à 3 chiffres de l'INSEE), J. Robert a constaté l'hétérogénéité des profils de financement des subdivisions de certaines grandes branches, particulièrement de l'industrie chimique. Au sein même de celle-ci un poste tel qu' "engrais et insecticides" qui apparaît couramment dans les statistiques, agrège, d'après leurs utilisations agricoles, une industrie lourde (les engrais) avec une industrie légère dont certaines productions confinent à la pharmacie (insecticides). Quant à l'interprétation des résultats de l'analyse, elle mettait le plus en évidence non la nécessité de la production, mais certaines décisions de politique générale accordant ou refusant à telle industrie le recours aux sources de crédit les plus favorables.

De cette expérience, nous avons retenu deux règles :

A) utiliser une nomenclature très fine : même si l'on recherche une vue d'ensemble de l'économie, considérée d'un certain point de vue (le crédit dans la thèse de J. Robert), on ne peut se fier aux agrégats usuels. Sans doute est-il vraisemblable qu'à tout problème particulier (e.g. incidence des coûts de l'énergie) une nomenclature en moins de 30 postes peut suffire pourvu que ces postes soient bien choisis. Mais d'une part toute étude débouche sur une autre ; le point de vue change, la nomenclature initiale devient inadéquate. D'autre part, en l'état d'ignorance où nous sommes des agrégats qui d'un certain point de vue sont homogènes, il faut partir d'une nomenclature aussi détaillée que possible ; l'un des buts de la recherche étant justement de faire une typologie des branches et de définir les agrégats qui conviennent à une étude particulière. Eventuellement dans une première étape, on ne considérera que les postes les plus importants dont le total représente 50% puis 80% de l'économie du pays.

B) considérer les mouvements de biens réels, avant les opérations financières : les décisions de politique intérieure et plus encore la conjoncture internationale perturbent violemment le crédit. Sur certains marchés mondiaux étroits, tel celui du blé, les cours sont instables ; tandis qu'à l'intérieur des pays ou des communautés, règne la taxation. Cependant les flux de biens réels (consommations privées et publiques ; besoins en énergie et matières premières ou en produits intermédiaires des industries...) ne varient que lentement.

1.2 Séries et matrices : La nécessité étant admise de se rapporter aux postes les plus fins de la nomenclature, et de considérer le flux de biens et services avant les capitaux et les crédits, il ne suffit pas d'analyser les variations de ces postes pris individuellement ; c'est-à-dire des séries ; même si celles-ci sont étudiées simultanément par l'analyse multidimensionnelle comme dans la thèse de P. Teilhard. Il importe de considérer des données relatives aux apports des séries deux à deux : c'est-à-dire de flux. Classiquement on construit à cet effet la matrice de Léontieff ou tableau des échanges industriels (TEI). De façon précise on notera :

I : un ensemble de produits,  $i$  ;

J : un ensemble de branches,  $j$  ;

$k(i,j)$  : valeur (dans une monnaie quelconque) de la quantité du produit  $i$  absorbée, au cours d'une année par la branche  $j$ .

Dans la pratique, les branches peuvent être confondues avec les

produits ; et par conséquent  $I = J$ , on a un TEI carré (mais non symétrique) : e.g. on utilise en France un tableau  $65 \times 65$ . (Pour une étude de ce tableau, cf D. Domengès ; [C.A.H. FLUX] in *Cahiers* Vol V n° 2).

Dans son principe la conception de Léontieff est très intéressante. Mais l'analyse détaillée des TEI, soulève des questions très embarrassantes. C'est ce que nous a montré un rapport de B. Ghermani et A. Langlat (DEA de stage ; 1974).

La matière de ce rapport est originale. Les données fournies aux auteurs par la DATAR et provenant de l'Institut Battelle, sont des TEI relatifs à six pays de l'Europe Occidentale : Allemagne ; Belgique (avec Luxembourg) ; France ; Grande-Bretagne ; Hollande ; Italie. Pour harmoniser les statistiques de tous ces pays, il a fallu se borner à une nomenclature en 22 branches. Les tableaux de base sont relatifs à 1970 ; à quoi s'ajoutent des prévisions pour 1980 et 1990 lesquelles, élaborées avant la crise pétrolière, ne sont qu'un mirage !

Avant toute analyse, on s'interroge sur la diagonale de ces tableaux. Les éléments  $k(i,i)$  sont dits *intraconsommation*, c'est-à-dire consommation de la branche  $i$  en son propre produit. La comptabilité nationale française les suppose nulles, faute de pouvoir les mesurer exactement : des biens sont produits par un atelier ou un établissement, et transférés à un autre atelier ou établissement de la même entreprise : il n'y a pas de transaction visible. Mais en est-il bien toujours ainsi ? Supposons que l'on passe d'une nomenclature détaillée des produits à une nomenclature plus agrégée : e.g. deux postes  $i_1$  et  $i_2$  sont agrégés en un seul  $i$  ; la consommation interbranche  $k(i_1, i_2)$  devient une intraconsommation et rentre dans  $k(i,i)$ . En fait l'intraconsommation, définie comme la consommation par une branche de produits rentrant dans la catégorie à laquelle elle se consacre elle-même, peut porter non seulement sur des produits que la branche de l'industrie nationale de tel pays a effectivement fabriqués, mais également sur des produits importés. C'est sur quoi l'analyse a appelé l'attention des auteurs du rapport.

Dans cette analyse, on a traité un tableau  $132 \times 44$ , construit comme suit :

$132 = 6 \times 22$  : chaque ligne représente une branche considérée dans un pays : e.g. A15 désigne la branche 15 (automobile) considérée en Allemagne ; B10 = acier en Belgique ; etc.

$44 = 2 \times 22$  : chaque colonne désigne les consommations  $C$  ou les ventes  $V$  : e.g.  $k(A15, C10)$  désigne la consommation en acier de l'industrie automobile allemande ; et le même nombre apparaît aussi en  $k(A10, V15)$  comme une vente de la branche 10 à la branche 15.

Autrement dit on a construit pour chaque pays un tableau rectangulaire  $22 \times 44$ , en juxtaposant le TEI à son transposé ; puis empilé les 6 tableaux relatifs aux divers pays.

L'analyse a montré dans l'ensemble une grande similitude des profils des lignes (i.e. de l'ensemble des achats et des ventes) d'une même branche, quel que soit le pays considéré, mais le point B15 (automobile en Belgique) se détache nettement des profils de la branche 15 dans les autres pays. Ceci s'explique parce qu'il n'y a pas en Belgique la totalité des fabrications de la branche 15 : pourtant une usine de carburateurs, une fonderie etc. ne consomment pas en Belgique d'autres produits qu'ailleurs, et elles fournissent les mêmes utilisateurs. Des anomalies tout aussi importantes que celles relatives au point 15 apparaîtraient si au lieu du TEI de la France, on considérait les TEI de chacune de ses 21 régions ; avec l'embarrassant problème des échanges de région à région etc. .

Ainsi partant de la diagonale du TEI et du cas de la Belgique , nous avons découvert qu'il ne suffit pas de tabuler même suivant une nomenclature assez fine les échanges de branche à branche : il faut analyser un graphe d'échanges.

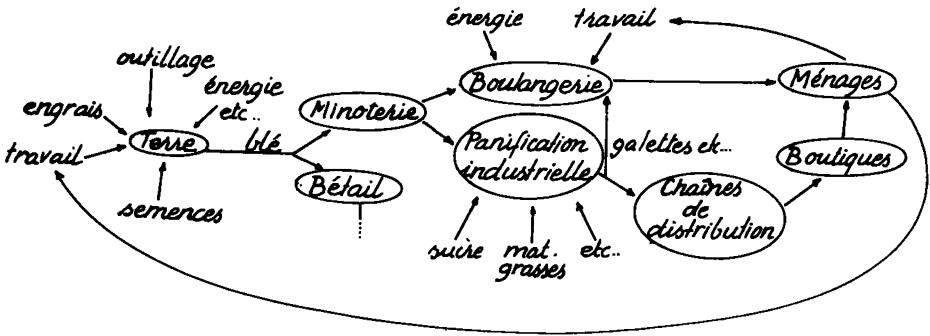
1.3 Le graphe des productions et consommations : Quelques images de la chaîne de fabrication d'une usine d'automobiles sont familières à tous : ici un bloc nu, venu de la fonderie, devient un moteur par usinages et apports de pièces successifs ; là une carrosserie de tôle blanche pénètre lentement sous un tunnel, d'où elle sort revêtue d'une couche de peinture déjà durcie sous les rayons de lampes à lumière infra-rouge ; ailleurs s'ajoutent les roues ou des fauteuils . De même l'ensemble des productions et des consommations de l'agriculture et de l'industrie est une chaîne, ou plutôt un enchevêtrement de cycles ; un graphe dont il faudrait distinguer les sommets et les arêtes ; dont le nombre est élevé mais non infini ; et, croyons-nous , à la portée de l'appareil statistique contemporain , surtout si l'on ne prétend pas achever la description de toute l'économie ; mais seulement reconnaître l'existence des principaux sommets (ou pôles) et des arêtes qui les relie. Dès maintenant on peut de ce point de vue écrire des monographies ; compiler et analyser des tableaux de données.

Deux figures dessinées sans étude de détail (!) sur les thèmes du blé et de l'acier suffiront à montrer au lecteur quel graphe nous voudrions voir élaborer. Et voici quelques suggestions.

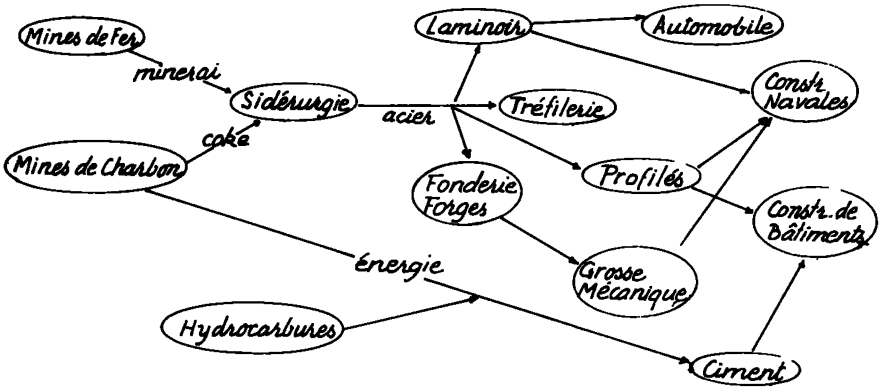
*A Etude d'un pôle* : cas d'un monopole ou oligopole : l'aluminium ; l'acier : on dénombre dans un pays les producteurs indépendants. Tout à l'opposé, le blé est produit dans des centaines de milliers d'exploitations agricoles. Celles-ci font l'objet de statistiques très complètes : on fera une typologie des entreprises céréalières d'après la taille ; la structure du personnel ; le mode de culture (engrais, semences) ; le climat ; l'association avec d'autres activités (élevage notamment) ; finalement on aboutit à une vue d'ensemble assez précise pour pouvoir e.g. estimer le rendement à l'hectare en fonction des autres variables décrivant l'exploitation.

On notera l'originalité de ce pôle qu'est la terre, dont les productions peuvent être dans une certaine mesure modifiées en une année . Cependant, comme l'a étudié G. Guigou, la terre est aujourd'hui disputée à l'agriculture par des usages qui en quelques décennies menacent de compromettre notre alimentation : zones industrielles, zones d'habitat étendu ("HLM horizontal"), espaces de loisirs. Seule en Europe la Hollande a su préserver sa terre. D'autres pôles naguère ignorés parce que jugés inépuisables : e.g. l'atmosphère en air, la rivière en eau, la forêt en fagots sont aujourd'hui l'objet de nos inquiétudes !

*B Etude d'un produit à la sortie d'un pôle* : Où va l'acier : laminier, tréfilerie, partage de ces demi-produits entre leurs principaux utilisateurs. Ou encore : étude simultanée de l'évolution temporelle du partage des métaux concurrents (cuivre, aluminium...) entre plusieurs usages ; de la concurrence entre plusieurs sources d'énergie pour diverses industries. Il importe ici de surveiller l'apparition d'arêtes nouvelles du graphe ; c'est-à-dire de fabrications nouvelles, qu'un statisticien négligent pourra ranger dans les postes d'une nomenclature figée ; mais que le producteur doit adopter à temps. On nous a cité l'exemple des tubes d'aciers de très gros diamètre utilisés notamment pour les plates-formes de forage au large ; et qui sont un marché récemment ouvert à la sidérurgie, et déjà occupé par les fournisseurs les plus vigilants.



Figures : Deux exemples sommaires de graphes des production et consommation : ci-dessus le blé : ci-dessous l'acier.



C Répartition (géographique notamment) des flux sur une arête du graphe Sur une même arête, e.g., approvisionnement de la sidérurgie en minerai de fer, se superposent des flux spatialement très étalés : e.g. le minerai utilisé à Dunkerque peut venir du Québec ou de Mauritanie. D'où des tableaux de correspondance, évoluant au cours du temps: certaines analyses en ont déjà été faites.

Finalement, en multipliant les études de détail (dont certaines sont en cours ; et beaucoup sont vitales pour l'une ou l'autre des branches de l'économie française, en quête de marchés) on parviendrait à l'intelligence de l'ensemble.

1.4 La place de l'homme dans les bilans : François Perroux (in : Annuaire du Collège de France ; 74<sup>e</sup> année, 1974) déclare avoir cherché en vain une de nos comptabilités, privée, publique ou nationale -, qui relève correctement l'amortissement de l'agent humain, ni sous la forme rudimentaire de l'amortissement-usure, ni sous la forme moderne de l'amortissement-obsolescence qu'on réserve aux choses et qui, justement généralisé, inclurait le recyclage et l'élévation du niveau professionnel et culturel : "C'est là, selon F.P. une "Remarquable illustration du fétichisme de la marchandise (Karl Marx)". Et de conclure : "les comptes des marchands laissent percer la mauvaise foi objective de l'*homos computans*"

Nous croyons nous aussi qu'en faisant à l'homme une place à part, en le mettant hors concours avec tous les honneurs, on peut bien le spolier le plus complètement. Qu'on reconnaisse d'abord l'homme comme un pôle : consommateur et producteur, au même plan que les autres, autant qu'il se peut ; pour corriger seulement ensuite par la considération que mérite l'âme.

Matériellement, la place de l'homme est-elle unique ? L'analyse de la matrice des flux impose-t-elle de lui attribuer le premier rôle ? Est-il bien le consommateur universel vers qui tout converge ? Le producteur universel qui contribue à tout ? N'acceptons pas sans examen les polarités convenues : un "cartoon" américain montre deux rats en conversation dans un labyrinthe : "Enfin dit l'un d'eux ! j'ai réussi à conditionner mon psychologue : chaque fois que j'appuie sur le levier, il me donne une dragée de nourriture". On ne sait pas toujours en effet, de deux lequel est le maître, et lequel est l'esclave.

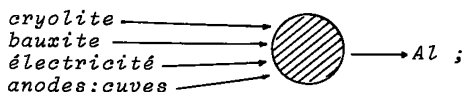
Imaginons une représentation de la société centrée sur la voiture : tout sert à la concevoir, à l'enfanter, à l'alimenter ; à lui donner des logements, des voies ; un serviteur qui la conduit. Mais objectera-t-on, la voiture sans chauffeur est oisive ; tandis que celui-ci produit encore et consomme. Quoi ! le valet aussi travaille pendant que son maître dort ; etc. .

Trêve de fables cependant. Voici une notion économique classique, la *valeur ajoutée* qui met arbitrairement l'homme au centre. A prendre les mots avec leur sens la v. a. renvoie au schéma artisanal suivant : d'un fragment de bronze qui vaut cinq francs, le ciseleur fait un bougeoir, qui vaut 100 F : v.a. =  $100 - 5 = 95$  : c'est clair. Or considérons l'image d'un atelier d'ajustage sur une gravure du XVIII<sup>e</sup> siècle : on reconnaît un tour, une fraiseuse... ; mais on découvre aussi une grande roue qu'entraîne un apprenti, et dont l'arbre par des courroies met en mouvement toutes les machines. Dans le calcul d'un économiste contemporain, l'entretien de cet apprenti est compté dans la valeur ajoutée ; en ce sens que substituer à la grande roue un petit moteur équivaut à augmenter les coûts et diminuer d'autant la v. a. . Pourtant cette substitution ne modifierait pas immédiatement la fabrication. Quelques exemples feront voir combien la v. a., interprétable comme un agrégat hétéroclite de salaires, de bénéfices etc. change de

sens selon les branches (\*).

A Valeur ajoutée par le commerce de détail : les biens matériels n'ont pas changé de valeur, dans la mesure où ils n'ont pas changé de forme : il n'y a pas eu une transformation mais seulement une mobilisation ; où l'homme joue le premier rôle.

B L'industrie de la production de l'aluminium : faisons un schéma :



Ici peu de main-d'oeuvre, mais une profonde transformation de la matière, avec une grande consommation d'énergie.

C Il n'est même pas possible de se borner à considérer la valeur des flux (indépendamment de la part distincte de l'homme, cf A ; ou des transformations de la matière, cf B) ; car il y a les entrepôts : là, ni transformation, ni distribution ; pourtant, sans énergie ni main-d'oeuvre, les entrepôts remplissent un rôle : l'actualisation, vaincre le temps.

A la vérité, du point de vue même des coûts, c'est l'homme qui pose le problème principal. Même si la comptabilité que réclame François Perroux n'est pas faite, il est clair que la moitié des budgets nationaux est consacrée à la production, à l'entretien, au conditionnement de l'homme : Santé publique 20% ; Education nationale 20% ; à quoi on ajoutera l'information, l'armée, ... Dans ce bilan, les faits moraux ou sociaux majeurs : avortement, migrations des travailleurs, dont le poids est si lourd, apparaissent aussi avec un poids matériel formidable : deviennent disponibles en main-d'oeuvre des millions de femmes ; servent ici des hommes qui ont été produits là avec d'autres coûts...

Et que dire des robots dont l'informatique devrait dit-on faire bientôt d'authentiques concurrents des travailleurs humains ?

Sans doute ne peut-on être généreux sans commencer par être cynique : un modèle de l'économie existe, selon lequel les salariés sont des agents qui produisent et consomment ; et les travailleurs indépendants sont des inadaptés ou des parasites : il ne reste plus que des dirigeants et des hauts-fournaux. Le calcul statistique se doit d'établir la comptabilité qu'implique ce modèle.

---

(\*) Dans l'analyse des bilans d'entreprise effectué par C. Desroussilles (cf [BILANS], Cahiers Vol V n° 1 § 4) apparaissent trois pôles de structure : Premier pôle : Le commerce (cf A) associé à des activités industrielles ayant un taux rapide de rotation de capitaux (minoterie). Deuxième pôle : Les industries lourdes (cf B) associées aux entrepôts (cf C) et aussi aux forêts et location d'immeubles, parce qu'il s'agit dans toutes ces branches



*d'immobilisation à long terme. Enfin les industries où le travail humain contribue de façon décisive à la valeur ajoutée (e. g. le vêtement) constituent le troisième pôle.*