

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

HENRI GUITTON

Articles. Statistiques, économétrie, analyse des données

Journal de la société statistique de Paris, tome 114 (1973), p. 107-112

http://www.numdam.org/item?id=JSFS_1973__114__107_0

© Société de statistique de Paris, 1973, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/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/>

II

ARTICLES

STATISTIQUES, ÉCONOMÉTRIE, ANALYSE DES DONNÉES

The word "Econometrics" was created in 1930. According to R. Frisch, the Econometrics is like a synthesis by mathematics of the theory and the facts statistically seized. In reading specialized publications, we see this word diverged its first tendency : too much place for theory, not enough for real facts. The synthesis could be done by analysis of the data or multidimensional analysis.

Der Ausdruck « Econometrie » wurde im Jahre 1930 geschaffen. Nach R. Frisch erschrint die Econometrie als eine Zusammenfassung durch die Mathematik der Theorie und der statistisch erfassten Tatsachen Wenn man die spezialisierten Zeitschriften verfolgt, hat man den Eindruck, dass das Wort von seiner ursprünglichen Richtung abgewichen ist : zuviel Platz für die Theorie, nicht genügend für die wirklichen Tatsachen. Die Zusammenfassung könnte verwirklicht werden durch eine Analysierung der Tatsachen oder eine Analysierung, die in mehreren Richtungen geht.

El vocable « econometría » lo crearon en 1930. Segun R. Frisch se presenta la econometría como una síntesis por las matemáticas de la teoría y de los hechos cogidos estadísticamente. Al leer las revistas especializadas se podría pensar que se ha desviado la palabra de su sentido natural : demasiado espacio se reserva a la teoría y demasiado poco a los hechos reales. La síntesis se podría realizar con el análisis de los datos o análisis multidimensional.

On parle depuis fort longtemps de statistique. Le mot date de la fin du xvii^e siècle. On aime beaucoup aussi employer le terme récent d'économétrie. Mais quand on demande à un étudiant quelle est la différence entre statistique et économétrie, il est la plupart du temps embarrassé. J'en ai fait récemment l'expérience auprès de ceux qui avaient achevé leurs études de licence et qui commençaient celles de doctorat. Les réponses m'ont surpris. On avait d'abord l'impression que les deux termes étaient synonymes et qu'on pouvait employer l'un à la place de l'autre. En insistant, on faisait préciser que la statistique avait une vocation beaucoup plus générale, puisqu'elle s'applique à tout ce qui est nombreux dans l'univers, capable de mesure et donc d'expression chiffrée : aussi bien dans le domaine des choses matérielles inanimées ou animées, géologiques ou biologiques, que dans celui des données humaines, psychologiques, démographiques, et donc aussi économiques. Comme l'économétrie contient explicitement le mot économie et que l'expression *metron* veut dire mesurer, on est amené à dire que l'économétrie n'est qu'une partie de la statistique.

Une autre précision est aussitôt apportée. Comme dans les cours de statistique et dans ceux d'économétrie, on fait usage de mathématiques, on pourrait se dire : les deux

disciplines emploient le même langage, les mêmes méthodes. Mais ici une autre difficulté se présente : elle tient à la manière d'exprimer les relations entre la statistique, l'économétrie et l'économie mathématique. Dans l'opinion la plus courante, on assimile économétrie et économie mathématique, la statistique faisant toujours bande à part, puisqu'elle est un simple outil, mathématique certes, mais applicable à toutes les données du monde, même non économique.

Je poursuis mes investigations. De quand datent les deux disciplines et quels en ont été ou en sont encore les représentants les plus éminents? J'entends dire : c'est walras (1871) le créateur de l'une et de l'autre. Et je rétorque : que pensez-vous de Cournot (1838)?

Me voici ainsi amené à rétablir la vérité historique. C'est en 1930 seulement qu'est né le vocable d'Économétrie — année où s'est fondée à Chicago la Société *Econometrica* à l'instigation des trois auteurs : Schumpeter, Divisia, R. Frisch. Ce dernier, prix Nobel de Science économique, 1971, professeur à l'Université d'Oslo est le seul survivant. Je me plais à rappeler d'année en année, le merveilleux discours qu'il prononça en 1931 à Lausanne, ville choisie à dessein pour introduire l'économétrie en Europe. Il a justement défini ce qu'il fallait désormais appeler *Économétrie*.

R. Frisch constate deux choses :

1^o les recherches économiques théoriques ne se sont pas inspirées autant qu'on aurait pu le désirer de l'observation ininterrompue des institutions et des faits;

2^o d'un autre côté les investigations statistiques ont été conçues d'une façon trop empirique, on s'est laissé entraîner par une sorte de manie d'information statistique. Ainsi, c'est la collaboration intime et suivie entre l'observateur et le penseur qui a manqué en économie politique.

Ainsi l'économétrie aura un triple but :

- a) préciser la théorie économique elle-même;
- b) développer des procédés statistiques;
- c) utiliser les mathématiques pour établir la relation adéquate entre la théorie économique et les réalités statistiques.

L'économétrie se présente comme une synthèse par les mathématiques, de la théorie et des faits statistiquement saisis.

Comme on le voit : la statistique ne constitue pas par elle-même une construction, pas plus que la mathématique, mais c'est bien son intégration dans la recherche qui est indispensable en principe à la construction mathématique d'une théorie économique à qui elle doit donner une substance réaliste.

Tandis que l'économie mathématique, en tant que telle, ne s'intéresse pas aux faits mais travaille sur des concepts, l'économétrie est bien mathématique, mais elle ne devrait pas se comprendre sans procéder comme le dit son nom, à des *mesures* expérimentales.

Walras n'est pas un économètre : l'économie pure est par définition même une économie hors les faits; Debreu qui a fondamentalement amélioré le système de l'équilibre abstrait de Walras n'est pas un économètre. Ils sont des mathématiciens de l'économie. On pourrait alors se demander : mais où sont donc les vrais économètres?

Et voici qu'on est à nouveau gêné pour répondre à la question. Une manière tout à fait naturelle paraît être de dire : Mais consultez donc les livraisons successives de la Revue *Econometrica*.

Tous les auteurs qui y ont écrit, et y écrivent encore, constituent la grande famille

des économètres. Personne ne pourra le nier. Cependant, j'entends encore Divisia me confier quelques temps avant sa mort :

« Je ne reconnais plus l'enfant grandi dont nous étions le père. J'ai de la peine à suivre beaucoup des articles de la revue *Econometrica*. Les mathématiques, dont nous étions si partisans, ont pris une telle part qu'elles ont fini par tout envahir, par tout absorber. Les faits statistiques qui, selon nos conventions, devraient toujours être présents, sont souvent absents. Si des séries statistiques interviennent dans les constructions, ce ne sont pas des séries vécues apportées par telle ou telle réalité, de telle ou telle période de l'histoire, ce sont des séries conçues, imaginées par l'esprit. »

Et ainsi l'économétrie aurait, provisoirement au moins, dévié de sa tendance originale. Elle n'aurait pas respecté sa vocation.

J'exagère peut-être — je ne voudrais pas déformer la vérité : car je sais bien par ailleurs qu'avant de saisir les faits, avant d'incorporer des séries statistiques apportées par la vie dans un modèle destiné à les expliquer et aussi à les orienter dans un sens nouveau pour demain, il faut forger de bons outils, il faut être sûr de la valeur des liaisons que l'on va expliciter. Des faits tout seuls, assemblés sans principe directeur ne constitueront jamais un modèle digne de ce nom. Nous sommes dans une phase de l'histoire économétrique où nous ne sommes pas capables de travailler directement sur cette matière première. Plus tard dans quelques années, prenez patience, vous verrez se constituer la vraie économétrie conforme aux aspirations synthétiques que vous avez définies.

Qu'il me soit permis d'utiliser ici l'image que proposait en 1620 le philosophe Bacon dans son ouvrage fameux écrit alors en latin (*Novum Organum*). Après plus de trois cents ans elle garde toute sa valeur. Elle reste très actuelle : « Les sciences ont été très aidées par les empiriques. Les empiriques, semblables aux fourmis, ne savent qu'amasser et user. Les rationalistes semblables à des araignées, font des toiles qu'ils tirent d'eux-mêmes. Le procédé de l'abeille tient le milieu entre les deux ; elle recueille ses matériaux sur les fleurs des jardins et des champs, mais elle les transforme et les distille par une vertu qui lui est propre. C'est l'image du véritable travail de la philosophie, qui ne se fie pas aux seules forces de l'esprit humain, et n'y prend pas même son véritable appui, qui ne se contente pas non plus de déposer dans la mémoire, sans y rien changer, des matériaux recueillis dans l'histoire naturelle et les arts mécaniques, mais les porte jusqu'à l'esprit, modifiés et transformés. C'est pourquoi il y a tout à espérer d'une alliance intime et sacrée de ces deux facultés expérimentale et rationnelle. »

Le vœu que formulait Bacon en 1620 reste toujours actuel. On ne peut pas dire que cette alliance intime et sacrée des facultés expérimentale et rationnelle soit réalisée. Les fourmis et les araignées continuent de travailler. Il y a encore bien peu d'abeilles. Et cependant il faudrait rendre hommage aux travaux de l'école américaine, de l'école hollandaise, de l'école française dont E. Malinvaud est le chef ⁽¹⁾, et à laquelle se rattache G. Rottier qui dirige l'enseignement et les recherches de l'Université de Paris I, sans oublier celui qui a le premier rendu opérationnelles les prévisions économétriques de la consommation française, Vangrevelinghe, en relation avec le Centre de recherches du CREDOC.

Je vais maintenant me demander si une forme d'analyse qui prend en ces présentes années de plus en plus d'importance ne serait pas apte à s'approcher de l'idéal de synthèse qui nous préoccupe. Cette analyse est dénommée sous un terme d'ensemble, « l'analyse des données », et d'une manière plus précise l'analyse multicritère et l'analyse multidimensionnelle. Je tenterais de dire en quoi ces nouveaux types d'analyse sont susceptibles de réaliser

1. E. MALINVAUD, *Méthodes statistiques de l'Économétrie*, 2^e éd., 1969.

l'objectif dont les premiers essais économétriques se sont, consciemment ou non, éloignés.

Il s'agit de rechercher une explication des réalités économiques par des liaisons révélées par ces réalités elles-mêmes, permettant ainsi de mettre sur pied une théorie économique, la mathématique, ne l'oublions pas, demeurant l'instrument capable de révéler ce qui d'abord n'apparaît pas à l'observateur, ce qui est *caché* derrière le visible. On appliquerait ainsi la proposition que Bachelard mettait en exergue de son œuvre : il n'y a de science que de ce *qui est caché*, caché dans l'existence, et non pas tellement caché dans notre esprit ⁽¹⁾.

Les travaux consacrés à pareille recherche se multiplient dans les années présentes. Plusieurs thèses de doctorat s'y adonnent (parues ou à paraître). Citons seulement à titre indicatif : A. BONNAFOUS, *La logique de l'investigation économétrique*, thèse, Lyon, 1971; J. L. GOUGOU, *Théorie économique et transformation de l'espace agricole*, Gauthier-Villars, 1971.

Il faut ajouter que l'avènement des ordinateurs a contribué à permettre ces nouveaux types d'analyse qui n'auraient pas pu prendre corps sans eux. On pourrait même dire que l'analyse *d'avant* les ordinateurs n'est plus de la même nature que l'analyse *d'après* les ordinateurs ⁽²⁾. C'est ainsi que l'on pourrait peut être expliquer la tendance trop conceptuelle qu'a prise la première économétrie qui n'était pas encore capable de saisir les faits en trop grand nombre.

Essayons de dégager l'idée directrice de cette nouvelle ligne de recherches en opposant l'analyse unidimensionnelle à l'analyse multidimensionnelle. Est unidimensionnelle une analyse qui cherche à exprimer *une* dimension par un *seul* nombre. Si l'objet a 3 dimensions, trois mesures seront nécessaires. Si les 3 mesures sont agrégeables, un seul point suffira à le représenter. Si les dimensions sont des critères, par exemple les qualités : confort, sécurité, prix, et si la monnaie sert de commune mesure pour exprimer ces critères, alors l'unidimensionnalité est assurée; un seul chiffre permet l'expression d'une réalité complexe.

Le problème est singulièrement simplifié, tellement peut-être qu'il fait disparaître ce qu'il faut expliquer, mais en revanche il faut assurer une *mesure* correcte de chaque élément et une *agrégation* convenable en un chiffre unique, la monnaie étant ici évidemment d'un secours merveilleux.

Mais, soit dans la mesure, soit dans l'agrégation, on sait que les erreurs sont inévitables. L'ordre de grandeur des erreurs commande la valeur des résultats. On se convainc facilement que l'unidimensionnel est appauvrissant par essence; il exclut tout ce qui n'est pas capable de mesure comparable, en particulier l'hétérogène et le qualitatif, qui font cependant essentiellement partie de la réalité.

Il faut bien reconnaître évidemment que la structure de notre esprit est de type unidimensionnel. Nous avons toujours l'ambition d'expliquer les choses par un seul principe. Regardez les théories que vous connaissez : elles sont unicausales ou monistes. Chaque grand auteur a la prétention de trouver *la* cause de toutes les autres causes emmêlées. Par exemple c'est *la* monnaie qui explique tout le reste, et l'on est quantitativiste. C'est le travail qui est le principe unique et exclusif de la valeur, et l'on est ricardien et marxien. On pourrait facilement continuer l'énumération.

Voilà la contradiction. La pensée est par nature unidimensionnelle. La réalité est multidimensionnelle. Jusqu'alors, c'est la réalité qui a dû céder et se transformer par les stratagèmes que nous venons de dire en expressions unidimensionnelles. La théorie ne pouvait se construire qu'à ce prix.

1. BACHELARD, *Le rationalisme appliqué*, 1949.

2. Cf. H. GUITTON, *L'ordinateur au service de la pensée économique* in *Essays in honour of Giuseppe Ugo Papi*, Cedam, Padova, 1973.

Les travaux statistiques jusqu'à maintenant traditionnels ne nous en ont-ils pas donné la preuve? L'analyse de la variance, à partir d'une équation qui exprime une variable expliquée en fonction de variables explicatives, en est, comme méthode de recherche inductive de liaisons, le témoignage.

J'avais beaucoup apprécié pour ma part la méthode des faisceaux de R. Frisch, qui généralisait la recherche des meilleures corrélations. J'avais bien l'intuition que l'introduction successive de nouvelles variables dont on se demandait si elles amélioreraient ou si elles détérioraient les corrélations antérieures, nous mettait sur la bonne voie. Et c'était le procédé de l'ouverture ou de la fermeture des faisceaux (*Bunch Maps*). Mais les spécialistes m'ont appris que la méthode était particulièrement lourde dès qu'il y avait plus de 3 variables.

Et c'est pourquoi, d'intuition encore, j'ai senti que l'analyse multidimensionnelle *directe* était la voie de l'avenir. Nous voici en présence d'une multitude de données saisies durant un certain temps au moyen de « chroniques ». Nous n'avons aucune raison *a priori* de privilégier l'une d'entre elles, que toutes les autres expliqueront. La question se pose alors de procéder à une réduction. À partir de données statistiques qui ne peuvent être représentées que dans un espace de dimensions élevées difficilement assimilables pour le statisticien, on se propose de les reproduire dans un espace de faibles dimensions avec le minimum de perte d'informations. Ainsi, si l'on considère, sans dénaturer le phénomène, deux ensembles I et J , on dira que ces deux ensembles sont en correspondance, s'il existe des relations entre les éléments de ces ensembles et si les éléments s'associent par couples $(i \cdot j)$, $i \in I$ et $j \in J$. I sera l'ensemble des *observations* : $i \in I$, (i) $(1, 2, \dots, n)$; J l'ensemble des *variables* : $j \in J$, (j) $(1, 2, \dots, m)$; n sera par exemple les 90 départements et m 15 activités. On aurait ainsi au départ deux espaces : un nuage de 15 points variables dans un espace à 90 dimensions; ou un nuage de 90 points observation dans un espace à 15 dimensions.

La réduction en question revient alors à ramener les m variables pour l'ensemble de n observations à quelques variables cachées k appelées « axes factoriels ». Ce terme est significatif. Évidemment les axes résultent d'une recherche d'un sous-espace selon une procédure appropriée qui représente la « proximité » systématique des éléments de I , des éléments de J et des éléments du produit $I \times J$. Il n'est pas question dans ces simples réflexions d'exposer des procédures ⁽¹⁾.

Il faudrait au surplus bien comprendre les caractères propres de l'« analyse des correspondances » et de l'« analyse en composantes principales ⁽²⁾ » qui forment deux aspects de l'analyse factorielle.

Puissions-nous en retenir l'esprit. Les axes factoriels permettent de résumer les aspects communs à l'ensemble des variables, et pour les observations faites, sans pour autant perdre de l'information relative à la position des observations dans l'espace R^m . Notons que, les axes étant orthogonaux, les facteurs sont indépendants les uns des autres. C'est ainsi une manière de résoudre une difficulté jusqu'alors préoccupante : celle de la colinéarité entre les variables explicatives.

La classification et le regroupement sur le premier axe factoriel constituent la structure *majeure*. Si on ne devait retenir qu'un seul axe, ce serait le plus représentatif, celui qui conserve le plus d'information. Mais les autres axes, qui ne résument que relativement peu

1. L'ouvrage de L. LEBART et J.-P. FÉNELON, *Statistique et informatique appliquées*, Dunod, 1974, nous en donne la substance.

2. B. ROY, *Décisions avec critères multiples — problèmes et méthodes*, Métra, vol. XI, 1972.

J.-P. BENZECRI, *Distance distributionnelle et métrique du X^2 en analyse factorielle des correspondances*, Publication ISUP, Paris, 1970.



d'information sur la proximité des points-observations, ne sont pas à négliger : ils constituent la structure *mineure*.

Il n'est pas inutile de rappeler les origines de l'analyse factorielle. Elle date de 1904. C'est en psychologie qu'elle est apparue à l'instigation de Ch. Spearman. Il s'agissait de rechercher une grandeur *g non observable*, telle que les diverses aptitudes intellectuelles d'un individu évaluées par des expériences puissent être exprimées par un « facteur général d'intelligence ». Le mot de facteur se comprend facilement, c'est ce qui *fait*, ce qui engendre ce qu'on observe. Il y a longtemps qu'il était employé en économique. Mais au lieu de se donner par avance les facteurs, on va à leur recherche. Les unités observées sont désormais des états successifs d'un phénomène économique, saisi à travers de multiples dimensions.

Mais le problème qui se pose désormais est d'interpréter les axes factoriels qui ont permis la réduction et le regroupement des données. Cette interprétation est la tâche de l'heure. On considère ces axes pour le moment, comme des variables d'un nouveau type, dites variables structurelles ou variables latentes, ou variables cachées. Chacune de ces épithètes porte son sens et évoque des idées voisines. C'est la première qui nous paraît la plus significative. C'est par la structure des choses qu'il est possible de les expliquer : mais précisément les structures sont latentes ou cachées. La mission de l'esprit est de pouvoir les mettre à jour.

Du point de vue statistique, les axes factoriels correspondent aux vecteurs propres associés aux valeurs propres de la matrice des variances-covariances. On pourra dire encore : ils sont le résultat d'une opération de l'esprit, d'un calcul, d'un artifice. J'ai l'impression que cette opération où l'opérateur n'a pas d'idée préconçue, pas d'hypothèse préalable, est donc véritablement neutre, nous aide à découvrir cette structure cachée révélée par la réalité elle-même.

Nous n'avons sans doute pas encore compris toute la portée de cette découverte. Ces axes factoriels nous mettent peut être en présence de concepts nouveaux, à partir desquels de nouvelles théories pourront s'édifier. C'est le problème de la correspondance entre les espaces mathématiques abstraits et l'espace des réalités économiques, qui est ainsi posé.

*
* *

C'est Ragnar Frisch lui-même qui, à l'issue d'un congrès d'Économétrie, avait écrit : « All that is playometrics not econometrics ». Lui seul, le fondateur, pouvait se permettre un tel jugement. Du reste, d'autres grands noms peuvent être évoqués. O. Morgenstern réfléchissant sur les difficultés de tout traitement statistique en économie expliquait : (L'« économétrie est-elle une science exacte », *La Recherche*, n° 18, décembre 1971) que « face à ces difficultés il n'est pas étonnant que les économètres trouvent plus commode d'élaborer des méthodes abstraites et raffinées que de les mettre en pratique »; et Leontieff (« Theoretical Assumptions and non observed Facts », *American Economic Review*, mars 1971) de son côté exprimait la même idée : on est facilement satisfait « de construire des méthodes statistiques déductives de plus en plus raffinées sans jamais se lancer dans la recherche empirique ».

L'analyse des données, que nous venons seulement d'esquisser, ouvre la route vers des recherches nouvelles. N'est-ce pas par elle que se fera la synthèse non encore accomplie, dont je me suis permis de dire qu'elle était la véritable vocation de l'économétrie?

Henri GUITTON
de l'Institut