

Avant-propos

L'essentiel du contenu de ce livre a été rédigé il y a une vingtaine d'années, lorsque j'ai commencé à dispenser mes premiers cours sur la modélisation par équations structurales à l'université François Rabelais de Tours. N'étant jamais resté en souffrance, le manuscrit a été constamment mis à jour pour les besoins de mes cours universitaires et des multiples ateliers d'initiation à la modélisation structurale que j'animais dans différentes universités françaises et étrangères (francophones). Ces cours et ces ateliers étaient à la fois le lieu d'une initiation à l'outil statistique et une prise en main d'un logiciel sans lequel cet outil demeure abscons, abstrait et désincarné. Autant le dire d'emblée et sans ambages, toute initiation à la modélisation structurale est obligatoirement assortie d'une prise en main d'un logiciel de modélisation. Entre LISREL, Amos, EQS, *Mplus* Sepath/Statistica, Calis/SAS, le choix ne manque pas. Ces logiciels payants ont permis, à n'en pas douter, une démystification indéniable de la modélisation par équations structurales et lui ont assuré ainsi une véritable popularité, qui n'en finit pas d'ailleurs.

Rédiger un livre, en l'occurrence un manuel pratique d'utilisation de la modélisation structurale, exige de présenter un ou plusieurs de ces logiciels payants qui sont, il faut l'avouer, assez onéreux. C'est là où le bât blesse. Non pas qu'ils ne le méritent pas, mais en choisir un, c'est forcément lui faire de la réclame. Je ne peux ni ne veux y consentir. De plus, l'accès à ces logiciels payants demeure pour beaucoup d'étudiants et jeunes chercheurs une contrainte et un obstacle souvent insurmontables. J'ai pu éprouver à plusieurs reprises la difficulté d'enseigner la modélisation structurale dans des universités francophones africaines où il était impossible d'avoir un logiciel payant de modélisation structurale. Il m'est arrivé d'utiliser une version estudiantine bridée d'un logiciel payant pour illustrer mon cours.

R, logiciel *open source* gratuit, est apparu et la donne a changé. R est composé de packages dédiés à toutes sortes de traitement et analyse de données. Le package lavaan proposé par Rosseel (2012) est dédié à la modélisation structurale. Son succès fut immédiat tant il dispose de toutes les fonctionnalités mises en avant par les logiciels payants, et tant il offre une facilité d'utilisation déconcertante. Comme pour tout outil statistique, la pratique demeure la voie royale pour maîtriser la modélisation structurale. Rien de tel pour ce faire que de disposer d'un logiciel à portée de la main. R et lavaan ont bouleversé notre manière d'enseigner les outils statistiques, ainsi que la manière dont les étudiants peuvent les apprivoiser, se les approprier et les utiliser. Notre niveau d'exigence à leur égard change, comme évolue le regard de ces étudiants sur ces outils. Comprendre, apprendre et surtout s'exercer sans contrainte aucune (hormis celle de disposer d'un ordinateur) : voilà ce qu'offrent aux étudiants R et son package lavaan. Ce livre y contribue assurément.

Sans l'impulsion et la participation décisive et substantielle de Guillaume Broc, auteur du livre *Stats faciles avec R* (chez De Boeck), ce manuel n'aurait pu voir le jour. Nous partageons le crédo suivant : l'accès à la science et à ses outils doit se populariser et se démocratiser. Ce manuel, premier livre en français consacré à la modélisation par équations structurales avec lavaan, y participe pleinement, pensons-nous.

C'est parce que cet ouvrage se propose d'être un manuel didactique et pratique d'initiation à la modélisation structurale destiné aux étudiants et aux utilisateurs qui n'ont pas nécessairement besoin de formules mathématiques complexes pour s'approprier cet outil et pouvoir s'en servir à bon escient, que nous l'avions au préalable soumis à un comité de lecture composé, par ordre alphabétique, de : Nicolas Calcagni, Benjamin Caumeil et Stéphane Faury.

Le rôle de ce comité de lecture, composé de ces trois étudiants novices en modélisation structurale, était de relire le manuel afin d'en évaluer la clarté et la compréhensibilité. Leurs lectures minutieuses et leurs remarques pertinentes et judicieuses ont permis une amélioration substantielle de ce manuel. Qu'ils en soient très vivement remerciés. De fait, ils sont pleinement associés à ce projet. Toutefois, nous conservons et assumons l'entière responsabilité des erreurs, lacunes ou insuffisances pouvant subsister dans ce manuel.

Kamel GANA

Introduction

« Le temps des hypothèses décousues et mobiles est passé,
comme est passé le temps des expériences isolées et curieuses.
Désormais, l'hypothèse est synthèse. »

Gaston Bachelard
Le Nouvel Esprit scientifique, 1934

« Il n'y a de science que quand peut intervenir la mesure. »

Henri Piéron, 1975

Prédire et expliquer des phénomènes à partir d'observations non expérimentales est un enjeu méthodologique et épistémologique majeur pour les sciences humaines et sociales. Passer d'une démarche purement descriptive à une démarche explicative nécessite un corpus théorique solide assorti d'un outillage méthodo-statistique approprié. « C'est parce qu'ils s'inscrivent dans une perspective déjà tracée que les modèles structuraux constituent une étape importante dans l'évolution méthodologique et épistémologique de la psychologie, et non pas seulement l'un des engouements passagers trop fréquents dans l'histoire de notre discipline », écrivait Reuchlin (1995, p. 212). Le constat décrit avec justesse par Reuchlin vaut également pour plusieurs autres disciplines en sciences humaines et sociales.

La désignation « modèles structuraux » utilisée par Reuchlin (1995) est un raccourci renvoyant aux « modèles d'équations structurales » ou plus globalement à « la modélisation par équations structurales ». Nous traduisons *structural equation modeling* par « modélisation par équations structurales ». Certains auteurs l'ont rendu en français par « modélisation d'équations structurales » ou « modélisation en équations structurales ». Les trois propositions nous paraissent acceptables. Les traductions « modélisation structurelle » et « modélisation d'équations structurelles »

furent utilisées (Paulré, 1985 ; Roussel, Compoy et Durrieu, 2002). Nous ne les avons pas retenues. D'autres expressions, de moins en moins utilisées, désignent ces modèles et cette approche : « modèles de structures de covariances » (*covariance structure models*) et « analyse de structures de covariances » (*covariance structure analysis*). L'expression « modèles Lisrel » (*Lisrel Models*), du nom du premier logiciel de modélisation (voir Sörbom, 2001, pour l'historique de ce logiciel), a été souvent employée pour désigner les modèles structuraux. Elle est aujourd'hui tombée en désuétude. À l'instar de Reuchlin (1995), nous utiliserons ici parfois pour simplifier notre propos les raccourcis « modélisation structurale » et « modèles structuraux ».

L'engouement pour la modélisation structurale dont parlait Reuchlin (1995) demeure largement ancré, est-il besoin de le préciser, outre-Atlantique, comme en attestent plusieurs bibliographies annotées, le nombre d'ouvrages en anglais qui lui sont dédiés et la quasi-absence d'ouvrages en français. Ceci n'est point étonnant, surtout lorsque l'on sait que toutes les tentatives d'abstraction en sciences humaines et sociales, et plus particulièrement en psychologie et en sociologie, ont été accueillies en France, plus qu'ailleurs, nous semble-t-il, sans beaucoup d'enthousiasme. L'une des raisons principales est à rechercher certainement dans la méfiance de la recherche française à l'égard des instruments de mesure (c'est-à-dire la psychométrie, la sociométrie, l'économétrie). Méfiance qui trouve sa justification dans, entre autres, la crainte de voir les construits (*constructs*) mesurés se réduire à leurs instruments de mesure, et dans l'inquiétude de voir ceux-ci devenir une simplification de ceux-là. Or, sans mesure, on ne le sait bien, point de méthodes quantitatives, et donc point de science. Mais nul n'ignore ici que l'utilité des méthodes quantitatives en sciences humaines et sociales est encore marquée au coin de la suspicion. Suspicion due, entre autres, à l'assimilation de ces méthodes au fait de vouloir concevoir le monde social ou les phénomènes sociopsychologiques ou encore socio-économiques comme des systèmes entièrement modélisables et mathématiques. Les données issues de ces sciences sont, pense-t-on, trop variables et incertaines, trop complexes et subtiles pour pouvoir se plier à la rigueur de la formalisation et du raisonnement mathématique. Le recours à l'abstraction et à la modélisation concentré (et concentre encore) à la fois l'hostilité de nombreux spécialistes en sciences humaines et sociales qui y voyaient une manière de ravalier l'humain au rang de simples « choses quantifiables » et le scepticisme des mathématiciens. Le mathématicien et épistémologue Bertrand Russell (1947) (prix Nobel de littérature, 1950), cité par Allais (prix Nobel d'économie 1988), a parfaitement décrit cette suspicion en écrivant que « beaucoup de gens éprouvent une haine passionnée pour l'abstraction... Ils disent que la réalité est concrète et qu'en faisant des abstractions on laisse échapper tout l'essentiel. Ils disent encore que toute abstraction est falsification et qu'en omettant un des aspects du réel vous courez le risque de vous enfoncer dans le faux, en jugeant seulement d'après les aspects qui restent. Ceux qui raisonnent ainsi restent, eux, en dehors de la science » (Allais, 1954, p. 59).

Alors que l'utilisation de plus en plus fréquente de la modélisation structurale a démarré aux États-Unis dans le courant des années 1970, leur absence en France jusqu'à récemment est d'autant plus curieuse que la contribution du sociologue français Raymond Boudon (1965a) à son avènement a été substantielle et reconnue. Jeune enseignant de sociologie à l'université de Bordeaux, Boudon a fait en 1965 dans la *Revue française de sociologie* un exposé mathématique de sa nouvelle méthode d'analyses causales appelée « analyse de dépendance ».

Dans cet article intitulé « Méthodes d'analyses causales », Boudon (1965b) présente sa méthode non seulement comme « une extension de l'analyse de régression », mais surtout comme étant plus puissante que cette dernière (voir à la fin de ce livre la « Brève histoire de la modélisation structurale »). Et l'on s'étonnera encore davantage de cette absence lorsque l'on apprendra que Boudon avait fourni ultérieurement (Boudon, 1967) une familiarisation didactique de ces analyses causales dans un livre intitulé *L'Analyse mathématique des faits sociaux*, tiré de sa thèse de doctorat rédigée sous la direction de Jean Stoetzel, professeur de psychologie sociale à la Sorbonne et fondateur en 1938 du célèbre institut de sondage qu'est l'Institut français d'opinion publique (IFOP).

« Ce livre, comme la plupart de ceux de Raymond Boudon, n'a pas touché un large public », écrivait Bastin dans la rubrique nécrologique « Disparition » du journal *Le Monde* du 12 avril 2013 consacrée au décès de Boudon, survenu le 10 avril 2013. Ainsi, en France, la modélisation structurale n'a commencé à sortir de sa discrétion qu'au tout début des années 2000. En psychologie, par exemple, les travaux précurseurs de vulgarisation de Françoise Bacher (1984, 1987, 1988) y sont sans doute pour quelque chose. Elle a attiré l'attention sur l'importance de la modélisation structurale en psychologie dès 1984, et elle lui a même consacré en 1987/1988 un article (en deux parties) sous forme d'un exposé d'ensemble, émaillé d'exemples d'utilisation tout à la fois clairs et didactiques. Sans oublier son article de vulgarisation très complet, publié en 1999 dans la revue *L'Année psychologique*.

Depuis le début, il y a deux types d'acteurs majeurs qui ont œuvré, parfois parallèlement, à l'essor de la modélisation structurale : ceux qui étaient/sont dans une démarche d'application exigeante, approfondie et novatrice de la méthode à leur domaine d'étude, et ceux qui étaient/sont dans une démarche de développement et de perfectionnement de la méthode elle-même. Ces derniers sont généralement des statisticiens (mathématiciens, psychostatisticiens...) alors que les premiers sont généralement des analystes de données. Tout en se classant volontiers dans la catégorie des analystes de données, les auteurs du présent ouvrage reconnaissent l'importance des prérequis statistiques indispensables à un usage exigeant et efficace de tout outil d'analyse de données.

Le présent manuel, et ce vocable désigne selon *Le Petit Robert* un ouvrage didactique présentant les notions essentielles d'une technique, s'adresse à un public débutant souhaitant s'initier progressivement à la modélisation par équations structurales et exploiter sa souplesse, ses ouvertures et extensions. Et c'est en se plaçant du point de vue d'un utilisateur ayant un bagage statistique limité que nous avons entrepris cette tâche. Nous avons aussi pensé à ceux qui sont fâchés avec la statistique, et qui pourraient, plus que d'autres, se laisser impressionner par les beaux diagrammes et autres indices d'adéquation qui pullulent dans l'univers de la modélisation structurale. Nous serions fiers si ceux-là jugeaient insuffisante l'initiation, sans doute partielle, que nous leur proposons ici. Quant à ceux que l'utilisation des mathématiques en sciences humaines et sociales rebuterait, ceux qui n'ont jamais été convaincus par l'utilité des méthodes quantitatives dans ces sciences, il est fort probable que, quoi qu'on fasse, ils le demeureront à jamais. Ce manuel ne les concernera pas... Il n'est guère utile de s'arrêter sur le fait que l'utilisation de ces méthodes ne signifie nullement concevoir le monde social ou les phénomènes psychologiques comme des systèmes obligatoirement modélisables et formalisables mathématiquement. Un tel débat a fait perdre beaucoup de temps à l'évolution épistémologique et méthodologique des sciences humaines et sociales en France. Ce débat demeure vain...

En réalité, présenter une méthode quantitative de façon raisonnablement allégée en formules et détails mathématiques a été rendu possible grâce aux logiciels informatiques. Rien ne justifie désormais la présentation comme jadis d'une analyse statistique poussée jusqu'à son mode concret de calcul, comme lorsque ces calculs s'effectuaient, au pire, à la main, ou au mieux, à l'aide d'une simple calculatrice. Mais le risque d'une utilisation quasi mécanique de tels programmes, donnant souvent à l'utilisateur l'impression d'être dispensé de connaître les bases techniques des méthodes et tests qu'il manipule, est bien réel.

Nous avons tenté de limiter ce risque en évitant de nous substituer à un simple manuel d'utilisation d'un logiciel de modélisation. Tout en reconnaissant l'importance des prérequis statistiques indispensables à un usage exigeant et efficient de tout outil d'analyse de données, nous rassurons les lecteurs : point trop n'en faut. Précisons d'emblée que notre point de vue dans cet ouvrage est à la fois méthodologique et pratique et que nous n'avons nullement la prétention d'offrir un recueil de procédures de calculs détaillées de la modélisation structurale. Nous nous sommes placés sous l'angle de l'utilisateur désirant y trouver, sans peine, aussi bien une initiation technique qu'une initiation pratique, orientées vers l'utilisation de la modélisation structurale. Il ne s'agit pas d'un livre de « recettes » relatives à l'utilisation de la modélisation structurale conduisant à des résultats dépourvus d'une signification suffisamment précise et étayée. L'exercice est difficile, car il s'agit en même temps d'une prise en main d'un logiciel qui, elle, s'inscrit dans une logique de « mode d'emploi ».

Après cette introduction, un premier chapitre où les concepts fondateurs et fondamentaux sont d'abord introduits, le principe et les conventions de base présentés et illustrés par des exemples simples. Et la nature de la démarche clairement explicitée. Il s'agit d'une démarche confirmatoire : le modèle est d'abord spécifié, ensuite éprouvé. Une prise en main du logiciel lavaan constitue le contenu du deuxième chapitre. Développé par Rosseel (2012), le package *open source* lavaan dispose de toutes les fonctionnalités mises en avant par les logiciels payants de modélisation structurale, et ce malgré son jeune âge, puisqu'il est toujours en version BETA, c'est-à-dire encore en phase test et construction. Il offre une facilité d'utilisation déconcertante.

Le troisième chapitre de ce manuel présente les principales étapes de la mise à l'épreuve des faits d'un modèle structural. La modélisation structurale y est abordée tant du point de vue de son processus, c'est-à-dire les différentes étapes de son utilisation, que du point de vue de son produit, c'est-à-dire les résultats qu'elle génère et leur lecture. Aussi, les différents modèles structuraux y sont présentés et illustrés avec la syntaxe lavaan et l'examen de la solution (*output*) : modèles en pistes causales, modèle de l'interdépendance acteur-partenaire (*Actor-Partner Interdependence Model*, APIM). De même, les deux parties constitutives d'un modèle structural général y sont détaillées : le modèle de mesure et le modèle structural. Là aussi des illustrations en utilisant la syntaxe lavaan et l'examen des solutions permettent au lecteur de s'appropriier aussi bien le modèle que le logiciel.

Tout modèle est un mensonge tant que sa convergence avec les données n'a pas été confirmée. Mais un modèle ajustant bien les données ne signifie nullement qu'il représente la vérité (ou qu'il soit le seul modèle correct ; voir les modèles équivalents), il constitue simplement une bonne approximation de la réalité, et partant, une explication raisonnable des tendances affichées par nos données. Allais (1954) a raison d'écrire qu'à « *un niveau d'approximation donné, le modèle scientifiquement le meilleur est celui le plus commode* [l'italique est de l'auteur]. En ce sens il y a autant de théories vraies que de degrés d'approximation donnés » (p. 59). Quoi qu'il en soit, et plus que jamais, la réplication d'un modèle et sa contre-validation demeurent de mise.

Quant au quatrième chapitre, il est consacré à ce que l'on a appelé « les extensions plus ou moins récentes de la modélisation structurale ». Le vocable « extensions » est pris ici dans le sens de prolongement et progrès, car la démarche reste la même, quels que soient le degré de complexité des modèles spécifiés et le degré d'élaboration théorique qui les sous-tend. L'objectif est ici de montrer à travers quelques exemples d'utilisation la puissance et la souplesse de la modélisation structurale. Immenses sont ses potentialités, et multiples ses ouvertures. Riches et excitantes sont ses promesses. Il n'a cependant pas été possible de les passer toutes en revue. Il nous a semblé judicieux d'insister sur celles en passe de devenir incontournables. D'ailleurs, certaines analyses

sont devenues tellement courantes qu'elles pourraient cesser d'être considérées comme une simple extension des modèles structuraux de base. On peut penser aux analyses multigroupes qui offrent la possibilité de tester l'invariance d'un modèle à travers les populations, permettant ainsi d'établir la validité, voire l'universalité, de la construction théorique dont il est la représentation. Plus récents sont les modèles trait-état latent qui désignent un ensemble de modèles conçus et destinés à examiner la stabilité dans le temps (temporelle) d'un construit, et auxquels nous avons dédié un chapitre aussi bien technique que pratique. Enfin, les modèles de développement latent y trouvent leur place tant les données longitudinales, rares et précieuses n'en finissent pas d'intéresser les chercheurs. Leur exploitation à l'aide de modèles combinant l'analyse de structures de covariance (*covariance structure analysis*) et celle de structures de moyenne (*mean structure modeling*) constitue l'une des avancées récentes en modélisation structurale.

Nous suggérons au lecteur désireux d'acquérir sans perte de temps une initiation technique progressive de commencer par installer le logiciel gratuit lavaan. Une aide à la prise en main de ce logiciel est fournie par la lecture du deuxième chapitre de ce manuel. Le lecteur aura intérêt à s'exercer pas à pas en reprenant les données traitées dans le livre afin de répliquer les modèles présentés, et de ne passer à l'étape suivante que s'il obtient les mêmes résultats. Ces données seront disponibles sur un site dédié à ce manuel.

En guise de conclusion, un chapitre dédié à une brève histoire de la modélisation structurale, histoire traversée de part en part par la tumultueuse notion de causalité (inférence causale) en sciences humaines et sociales, sciences davantage observationnelles qu'expérimentales, vous est proposé. Nous avons envisagé de présenter ce chapitre en préambule à cet ouvrage. Il nous a cependant été suggéré de l'offrir en fin de l'ouvrage tant l'histoire de la modélisation ne semble pas être indispensable à une initiation à la modélisation structurale qui est l'objectif principal de cet ouvrage. Toutefois, la lecture de ce chapitre, qu'elle soit au début ou à la fin, permettra aux lecteurs de comprendre pourquoi nous avons pris le parti de parler de modèle (ou analyse) en pistes causales pour désigner le *path model* et le *path analysis*. Nous avons un instant pensé rendre *path model* en français par « modèle en pistes prédictives ». Cette traduction ne jure pas avec la nature même de la méthode. Mais, fidèles à l'esprit des initiateurs de la modélisation tels que Wright (1920), Simon (1953), Boudon (1965a) et Duncan (1966), nous continuerons de faire usage des vocables d'« inférence causale » et d'« effet causal », tout en nous entourant de toutes les précautions et justifications nécessaires à un tel usage. Il aurait été facile de bannir tout vocabulaire causal en modélisation structurale et de remplacer « modèle/analyse en pistes causales » par « modèle/analyse en pistes prédictives ». Un tel bannissement, quand bien même

défendable et légitime s'agissant de données observationnelles, clora à jamais tout débat sur l'inférence causale en sciences humaines et sociales et la place de la modélisation structurale dans ce débat. Et c'est regrettable... Pour l'avenir de ces sciences surtout.

Nous avons commencé cette introduction en paraphrasant Reuchlin dont la contribution à la diffusion en France des méthodes quantitatives d'analyse en psychologie est magistrale. Rappelons ici que sa thèse de doctorat soutenue en 1961 s'intitulait « Méthodes d'analyse factorielle à l'usage des psychologues ». En France, Boudon (1967) en sociologie et Reuchlin (1975, 1^{re} édition 1962) en psychologie furent, contre vents et marées, les initiateurs et les promoteurs de la démarche quantitative en sciences humaines et sociales. Qu'ils en soient ici vivement remerciés. Nous souhaiterions conclure notre propos introductif en citant encore une fois Reuchlin (1995) lorsqu'il précise avec justesse que les modèles structuraux « sont des instruments de travail dont l'usage n'est possible, il est vrai, que si l'on dispose de certaines connaissances et de certaines hypothèses psychologiques sur le fonctionnement des conduites que l'on étudie. Il serait paradoxal que les psychologues considèrent cette contrainte comme un inconvénient ». On pourrait même dire qu'ils auraient tort de l'envisager de la sorte. Et ce ne sont pas Hair, Babin et Krey (2017), spécialistes en marketing et sciences de la publicité, qui contrediraient Reuchlin. En effet, dans une récente revue de littérature examinant l'utilisation de la modélisation structurale dans les articles parus dans le *Journal of Advertising* depuis son premier numéro de 1972, ces auteurs reconnaissent d'abord que l'attrait de la modélisation chez les chercheurs et les praticiens de la publicité peut être attribué au fait que la méthode s'avère être un excellent outil pour soumettre à l'épreuve des faits les théories de la publicité, et admettent ensuite, sans ambages, que l'utilisation croissante de la modélisation dans la recherche scientifique en publicité a contribué substantiellement aux progrès conceptuels, empiriques et méthodologiques de la science de la publicité.

Passer du descriptif à l'explicatif n'est-il pas une évolution épistémologique nécessaire à toute science digne de ce nom ? Et ceci est bien évidemment valable pour une multitude de champs disciplinaires où la modélisation structurale est déjà employée : agronomie, écologie, économie, gestion, psycho-épidémiologie, sciences de l'éducation, sociologie...