



Mathématiques et sciences humaines

Mathematics and social sciences

191 | Automne 2010
Variabilité et inégalités

Quelques aspects de la statistique en sociologie, 1950-2010

Some statistical aspects in sociology, 1950-2010

Louis-André Vallet



Édition électronique

URL : <http://journals.openedition.org/msh/11870>

DOI : 10.4000/msh.11870

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 1 septembre 2010

Pagination : 65-80

ISSN : 0987-6936

Référence électronique

Louis-André Vallet, « Quelques aspects de la statistique en sociologie, 1950-2010 », *Mathématiques et sciences humaines* [En ligne], 191 | Automne 2010, mis en ligne le 16 février 2011, consulté le 23 avril 2019. URL : <http://journals.openedition.org/msh/11870> ; DOI : 10.4000/msh.11870

QUELQUES ASPECTS DE LA STATISTIQUE EN SOCIOLOGIE, 1950 – 2010

Louis-André VALLET¹

RÉSUMÉ – *L'article argumente que, dans la sociologie empirique contemporaine qui se fonde sur des enquêtes quantitatives à large échelle, l'analyse de la variabilité et de l'inégalité est à la fois conduite et révélée dans le cadre de l'usage d'un modèle statistique. Cet argument est illustré par la présentation de trois exemples importants issus de la littérature empirique sur la stratification sociale : l'usage de l'analyse en pistes causales pour le développement du modèle de l'acquisition du statut dans la sociologie américaine des années 1960 et 1970 ; l'usage du modèle multi-niveau pour l'analyse des effets contextuels ou de voisinage sur le devenir des individus dans la recherche en sciences sociales des années 1990 ; l'usage des modèles log-linéaire et log-multiplicatif d'un tableau de contingence pour mettre au jour et analyser les tendances temporelles de l'inégalité des chances sociales devant l'enseignement au sein d'une société.*

MOTS-CLÉS – Analyse en pistes causales, Effet contextuel, Modèle d'acquisition du statut, Modèle log-linéaire, Modèle multi-niveau, Modèle statistique

SUMMARY – Some Statistical Aspects in Sociology, 1950-2010

The paper argues that, within contemporaneous empirical sociology based on large-scale quantitative surveys, the analysis of variability and inequality is both embedded and revealed in the framework of a statistical model. This argument is illustrated with the presentation of three important examples in the empirical literature on social stratification: the use of path analysis for the development of the status attainment model in the 1960s and 1970s American sociology; the use of the multilevel model for the analysis of contextual and neighborhood effects on individual outcomes in the 1990s social science research; the use of the log-linear and log-multiplicative models for contingency tables to discover and analyze historical trends in inequality of educational opportunity within a country.

KEYWORDS – Contextual effect, Log-linear model, Multilevel model, Path analysis, Statistical model, Status attainment model

Pour traiter de la manière dont la sociologie empirique contemporaine étudie les phénomènes de dispersion, de variabilité et d'inégalités, il est utile de partir d'un peu plus loin en évoquant deux articles qui, à quelques années de distance, proposent des bilans contemporains de la relation entre statistique et sociologie.

Le premier est dû à Clifford Collier Clogg qui était professeur de sociologie et de statistique à Pennsylvania State University et est décédé brutalement en 1995 à l'âge de quarante-cinq ans [Sobel, 1996]. Publié en 1992 dans la revue *Statistical Science*, l'article de Clogg traite de « l'impact de la méthodologie sociologique sur la

¹ Laboratoire de Sociologie Quantitative, Centre de recherche en Économie et Statistique (CREST), CNRS, Timbre J350, 3 avenue Pierre Larousse, 92245 MALAKOFF Cedex, louis-andre.vallet@ensae.fr

méthodologie statistique » et s'accompagne aussi d'une discussion par les statisticiens David Bartholomew et Paul Holland, l'économiste Charles Manski et le psychométricien Ivo Molenaar. L'auteur y insiste sur le fait, patent à ses yeux, que les développements contemporains de la méthodologie sociologique et, plus largement, le contexte de la recherche empirique en sciences sociales ont assez largement contribué au développement de nouvelles méthodes, de portée générale, en statistique. À l'appui de cette thèse, Clogg détaille plusieurs exemples convaincants. Il s'agit d'une part de deux types particuliers de modèles statistiques. D'un côté, le modèle log-linéaire pour le traitement général des tableaux de contingence à plusieurs dimensions et des variables catégorielles (ou qualitatives) a été développé dans le cadre d'une interaction assez forte avec des problématiques de science sociale – tout spécialement dans le domaine de l'analyse de la mobilité intra ou intergénérationnelle. D'un autre côté, le modèle général d'analyse de « l'histoire des événements » (*event-history model*), également connu en France sous le nom d'analyse démographique des biographies et qui permet l'étude dynamique de données longitudinales, s'il est indubitablement lié à la bio-statistique, doit aussi nombre de ses développements à la recherche économique et sociologique, par exemple pour l'analyse des destinations multiples et celle des événements répétables dans le temps. Il s'agit d'autre part de la notion de variable latente et des modèles associés. La contribution de Lazarsfeld et Henry [1968] a été essentielle pour le développement du modèle des classes latentes – ou modèle de mélange fini de plusieurs sous-populations non directement observées. Par ailleurs, les modèles à variables latentes continues incluent à la fois les modèles à traits latents, spécialement développés dans le domaine de la psychométrie, et les modèles d'analyse factorielle et d'équations structurelles à variables latentes. Ces derniers sont très utilisés en éducation, psychologie et sociologie dans une optique d'analyse causale (cf. par exemple [Bollen, 1989]).

Publié en 2001 dans *Sociological Methodology*, généralement considérée comme la principale revue de méthodologie sociologique, le second article proposant un bilan contemporain de la relation entre statistique et sociologie est dû à Adrian Raftery, professeur des deux disciplines à l'Université de Washington et membre de la *National Academy of Sciences* américaine. Étudiant la période 1950-2000, l'auteur identifie trois générations successives – qui se recoupent en partie et demeurent aujourd'hui vivantes – de méthodes statistiques en sociologie, sur la base des types de données qu'elles manipulent. La première génération débute à la fin de la décennie 1940, traite de tableaux croisés et se concentre sur les mesures d'association entre variables et les modèles log-linéaires – sans doute la partie de la statistique à laquelle la sociologie a le plus contribué, selon Raftery. La deuxième génération apparaît dans les années 1960, traite de données individuelles d'enquêtes et se concentre sur les modèles causaux de type LISREL et l'analyse de « l'histoire des événements ». Enfin, une troisième génération émerge à la fin des années 1980 et aborde des données encore différentes, soit parce qu'elles ont une forme distincte – des textes ou des récits – soit parce que la dépendance y constitue un aspect crucial – comme les données spatiales ou les données de réseaux sociaux.

À la lecture, il apparaît clairement que les deux articles proposent des vues d'ensemble de la relation entre statistique et sociologie largement concordantes, mais on peut aussi remarquer que ni l'un, ni l'autre n'évoque explicitement un corps constitué de littérature méthodologique en sociologie qui concernerait la mesure de l'inégalité (ou des inégalités). Ainsi, il ne semble pas que l'on trouve, dans la littérature sociologique contemporaine, l'équivalent fonctionnel de ce que sont par exemple, en économie, les textes séminaux de Tony Atkinson, Amartya Sen ou Anthony Shorrocks. Un examen

systématique des sommaires de *Sociological Methodology*, depuis le premier volume paru en 1969, confirme nettement un tel diagnostic. Il existe en revanche en sociologie une littérature méthodologique foisonnante et en évolution constante qui traite de l'analyse de la ségrégation spatiale, du point de vue racial ou social, ou encore de l'analyse de la ségrégation professionnelle du point de vue du genre. Elle est très largement d'origine américaine et est aujourd'hui prise en compte dans la sociologie française².

La raison de l'absence mentionnée plus haut me paraît être la suivante. Fondée sur les données d'observation que fournissent des enquêtes statistiques sur grands échantillons, la sociologie quantitative contemporaine traite très généralement de corrélations ou d'associations statistiques entre variables. Une partie substantielle du travail scientifique consiste alors à mesurer la direction et la force de ces associations, à observer leur degré de robustesse une fois que d'autres variables sont prises en compte dans l'analyse, ou encore à examiner à quel degré ces associations varient selon une variable supplémentaire, par exemple le pays, le temps, le sexe, etc. – c'est-à-dire, en termes plus techniques, à étudier l'existence d'éventuels effets d'interaction.

Ainsi, le problème de l'analyse de la variabilité ou de l'inégalité se trouve d'emblée posé dans un cadre de modélisation statistique. L'objet de cet article est d'en fournir une illustration à partir de trois exemples marquants du développement de la recherche sociologique sur la stratification sociale : d'une part, le modèle en pistes causales (*path analysis*) de l'acquisition du statut dans la littérature américaine des années 1960 et 1970 ; d'autre part, les recherches des années 1990 relatives à la mesure d'un effet des caractéristiques de l'environnement social sur le devenir des individus et qui utilisent le modèle linéaire mixte ou modèle multi-niveau ; enfin, les développements récents de la recherche sur l'évolution historique de l'inégalité des chances devant l'enseignement à partir des modèles log-linéaire et log-multiplicatif d'un tableau de contingence.

1. PATH ANALYSIS ET MODÈLE D'ACQUISITION DU STATUT SOCIAL

Au cours des années 1950 et jusqu'au milieu de la décennie suivante, les analyses de la mobilité sociale entre générations portaient sur des tableaux croisés père-fils qui utilisaient le plus souvent des nomenclatures de positions sociales très rudimentaires – par exemple non manuel, manuel, agriculteur. Elles mesuraient le taux de mobilité observée entre générations pour s'attacher ensuite à décomposer celle-ci en deux parties : d'une part, la mobilité dite « structurelle », liée à la seule transformation de la structure sociale entre la génération des pères et celle des fils et évaluée par l'écart entre les deux marges du tableau de mobilité ; d'autre part, la mobilité dite « nette », censée traduire le degré d'ouverture ou de perméabilité de la structure sociale et mesurée en défalquant la mobilité structurelle de la mobilité totale observée. La critique de Duncan [1966] a été décisive : si l'enquête statistique est représentative de la génération des fils, elle ne peut simultanément l'être de la génération des pères. Dans le tableau de mobilité, la marge des origines sociales n'est donc, en un sens strict, représentative d'aucune génération réelle du passé. Il s'ensuit que l'évaluation de la mobilité structurelle est biaisée, ce qui invalide la décomposition précédente.

Une inflexion nette se produit donc dès l'article de Duncan et Hodge (1963) et, surtout, de l'ouvrage de Blau et Duncan (1967). L'accent est mis désormais sur

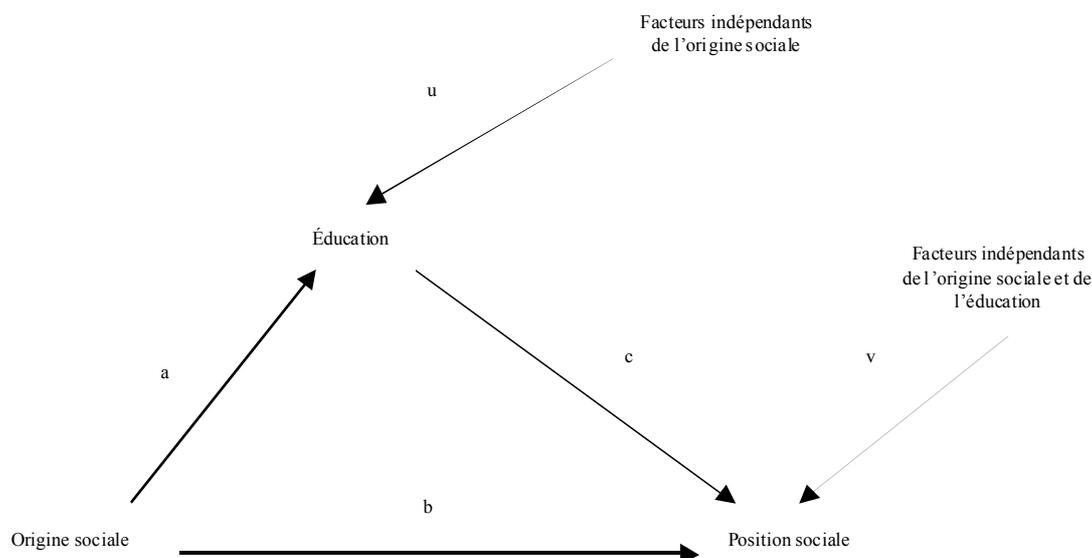
² Une livraison récente de la *Revue française de sociologie* l'illustre clairement (2009, volume 50, n° 3).

l'analyse du processus d'acquisition du statut social (*status attainment*), c'est-à-dire sur la mesure de l'influence des différents facteurs qui affectent la détermination de la position sociale des individus. Un tel déplacement du regard est permis par – et s'accompagne de – l'adoption de conventions et avancées méthodologiques.

D'une part, les sociologues de la stratification conviennent que les variables peuvent être mesurées sur des échelles numériques. Par exemple, le niveau d'éducation atteint est assimilé au nombre d'années d'études suivies. Ou encore, les position et origine sociales sont approchées par les scores caractéristiques de la profession exercée par un individu et de celle occupée par son père sur une échelle de statut socio-économique ou une échelle de prestige des professions préalablement construite. Par exemple, Duncan [1961] a régressé les scores de prestige disponibles pour quarante-cinq professions sur des mesures, tirées du recensement de 1950, de la proportion des titulaires de ces professions au moins diplômés de la fin des études secondaires et de la proportion de ceux gagnant 3500 dollars ou davantage. Établie *via* un modèle de régression linéaire multiple, la relation statistique entre les scores de prestige et ces deux mesures s'est avérée excellente. Cela a conduit Duncan à utiliser les coefficients de régression estimés pour construire une mesure de prestige pour toutes les professions présentes dans le recensement, mesure connue sous le nom d'indice de statut socio-économique de Duncan. On a pu mettre en évidence par ailleurs la forte stabilité de telles échelles, que celle-ci soit appréciée au cours du temps, entre plusieurs sociétés, ou en comparant les échelles établies en interrogeant des fractions de population différentes dans une même société [Treiman, 1977].

D'autre part, les sociologues adoptent aussi l'analyse statistique de régression et, plus généralement, l'analyse en pistes causales ou analyse de dépendance (*path analysis*) qu'ils reprennent du biométricien Sewall Wright [1921] afin de décomposer les mesures de corrélation entre variables en mesures d'influence, directe ou indirecte.

Dans leur article pionnier, Duncan et Hodge soulevaient trois questions principales. Quelle est la force de la corrélation entre le statut professionnel d'individus et celui de leurs pères ? Dans quelle mesure cette corrélation se constitue-t-elle *via* le niveau d'éducation obtenu ? Si l'on fait abstraction du fait qu'elle dépend du statut professionnel paternel, quel est « l'effet propre » de l'éducation obtenue sur le statut professionnel atteint ?



La figure ci-dessus permet de répondre en ces termes. Fondamentalement, les différences de niveau d'éducation observables sur un ensemble d'individus proviennent de deux sources. D'une part, ceux-ci peuvent être d'origines sociales diverses, les individus « mieux nés » bénéficiant d'une éducation plus poussée : la flèche a représente la force de cet effet qui contribuera à la facette « reproduction sociale » du rôle de l'éducation. D'autre part, les mêmes individus se différencient également sous un ensemble de facteurs qui sont *indépendants* de l'origine sociale, mais influent également sur le niveau d'instruction atteint : la flèche u représente la force de cet effet qui contribuera quant à lui à la facette « mobilité sociale » du rôle de l'éducation. Les mêmes individus peuvent ensuite être observés sous l'angle des différences qui les séparent en matière de position sociale atteinte. En suivant la même logique et compte tenu de l'ordonnancement temporel des variables, trois facteurs causaux sont ici envisageables : une influence de l'origine sociale qui s'exercerait « directement », c'est-à-dire indépendamment de l'éducation obtenue et se trouve représentée par la flèche b ; un effet du niveau d'éducation atteint, matérialisé par la flèche c ; enfin, l'action de tous les facteurs indépendants de l'origine sociale et de l'éducation (flèche v). À partir d'une mesure des corrélations bivariées entre origine sociale, éducation et position sociale, l'apport fondamental de la *path analysis* est d'estimer quantitativement la force des influences a , b , c , u et v .

Les deux résultats majeurs sont les suivants. D'une part, la valeur du produit ac est supérieure à b . L'influence de l'origine sociale sur la position sociale s'exerce donc davantage « indirectement », *via* le niveau d'éducation atteint, que « directement » et cela établit que, dans le cadre du schéma causal considéré, l'éducation est le facteur principal de la reproduction intergénérationnelle du statut social. D'autre part, l'influence u s'avère supérieure à l'influence a de sorte que la valeur du produit uc l'emporte sur celle du produit ac . Il s'ensuit que, des deux facettes du rôle de l'éducation, c'est l'aspect « promotion de la mobilité » qui l'emporte sur celui de « reproduction sociale ».

Au-delà de ces faits stylisés, le courant de recherche initié par Duncan et Hodge a aussi contribué à établir nombre de résultats empiriques en complexifiant progressivement le modèle d'analyse. À partir d'une enquête nationale, Blau et Duncan [1967] ont raisonné sur un ensemble de cinq variables : le niveau d'éducation et le statut professionnel du père, le niveau d'éducation du fils, son statut professionnel au premier emploi et son statut professionnel au moment de l'enquête. Aux États-Unis en 1962, la corrélation entre statut professionnel du père et statut professionnel du fils à l'enquête s'établit à 0,405. Elle s'avère donc bien inférieure à ce que serait la corrélation maximale (1) et nettement plus faible aussi que le lien entre l'éducation du fils et son statut à l'enquête (0,596). L'analyse de dépendance décompose la première corrélation en un effet indirect *via* l'éducation de 0,227 (56 %) et un effet indépendant de l'éducation de 0,178 (44 %). Ce dernier peut lui-même être décomposé. Il correspond d'une part, pour 0,063 (16 % du total), à la combinaison de l'effet du statut du père sur le statut du fils au premier emploi et de l'effet de ce premier statut sur le statut ultérieur. Il reste d'autre part, pour 0,115 (28 % du total), l'influence directe du statut paternel sur le statut professionnel du fils lors de l'enquête. Enfin, l'effet, direct ou *via* le premier emploi, de l'éducation du fils sur son statut final s'avère 2,9 fois plus fort que l'influence, directe ou *via* le premier emploi, du statut paternel sur la même variable. Blau et Duncan y voyaient le signe que, dans la société américaine, le statut professionnel est beaucoup plus acquis (*achievement*) qu'il n'est hérité (*ascription*).

Ces résultats et interprétations ont été largement commentés et discutés. Par exemple, dans une livraison de 1972 du *Journal of Political Economy*, l'économiste Samuel Bowles a argumenté qu'il était vraisemblable que Blau et Duncan surestimaient le rôle de l'éducation. D'une part, leur modèle d'analyse décrit de façon incomplète le milieu d'origine et, de ce fait, sous-estimerait son influence : en particulier, le revenu familial, la richesse des parents et leur position dans la hiérarchie des relations de travail sont absents. D'autre part, les variables présentes intègrent probablement des erreurs de mesure, ce qui aurait pour effet d'atténuer les corrélations observées. Aussi l'analyse du processus d'acquisition du statut s'est-elle enrichie progressivement et a-t-elle culminé avec les travaux réalisés sur une cohorte de diplômés de fin d'études secondaires du Wisconsin suivis au cours du temps [Sewell *et al.*, 2004]. Dans ce qu'ils nomment le « modèle socio-psychologique de l'acquisition du statut professionnel », les auteurs mobilisent près de vingt variables dont la quasi-totalité sont temporellement et causalement antérieures au niveau d'éducation atteint. Y figurent notamment, outre les variables classiques, le revenu parental, l'éducation et l'activité professionnelle de la mère, les aptitudes intellectuelles, des mesures de l'influence de l'entourage ainsi que les aspirations scolaires et professionnelles. Aucun de ces travaux, même dans ses versions les plus raffinées, n'a conduit à remettre sérieusement en cause la place centrale et le double rôle de l'éducation dans la détermination du statut professionnel, au sein de la société américaine.

Durant les décennies 1970 et 1980, le courant de recherche initié par Duncan et Hodge a connu un retentissement international considérable et de nombreuses études ont visé à reproduire les modèles les plus simples afin de comparer, entre plusieurs sociétés, les composantes du processus d'acquisition du statut. L'article de Treiman et Yip publié en 1989 en représente sans doute le point culminant. Il estime le modèle élémentaire (3 variables) d'acquisition du statut sur des données préalablement standardisées de 21 pays. La corrélation médiane entre les statuts professionnels du père et du fils s'établit à 0,345 avec un maximum de 0,547 en Inde. La proportion médiane de transmission « directe »³ du statut du père à celui du fils vaut 45 % avec un maximum de 93 % en Inde et un minimum de 13 % en Suède. Enfin, le rapport de l'effet de l'éducation du fils à l'effet direct du statut du père atteint un maximum de 16,0 en Suède et un minimum de 0,3 en Inde ; sa médiane vaut 3,1 : le rapport des deux quantités excède donc cette valeur dans la moitié des pays étudiés et se situe en deçà dans l'autre moitié. En dépit de cette centralité internationale, le modèle de l'acquisition du statut est resté assez méconnu en France. Raymond Boudon l'a évoqué brièvement dans le premier chapitre de *L'inégalité des chances* [1973] et la seule estimation sur données françaises, publiée en 1977 dans les *Annales de l'INSEE*, a été effectuée par un chercheur japonais dans le cadre d'un séjour à l'Institut de statistique français [Tachibanaki, 1977] : un examen attentif conduit à souligner une certaine proximité avec les estimations américaines de Blau et Duncan.

La recherche d'aujourd'hui sur la stratification sociale n'utilise plus guère l'analyse de dépendance pour étudier le processus d'acquisition du statut social. D'une part, l'usage d'indices composites de statut socio-économique a fait l'objet de critiques, à la fois conceptuelles et empiriques [Hauser, Warren, 1997]. D'autre part, dans le cadre de l'approche contrefactuelle pour l'estimation d'effets causaux, on a pu mettre en question l'interprétation strictement causale des mesures d'influence que fournit le modèle d'acquisition du statut [Sobel, 1998]. Il reste que ce dernier, intimement lié à la

³ C'est-à-dire indépendamment de l'éducation, ce qui correspond à la flèche *b* de la figure présentée plus haut.

path analysis, a eu un impact de long terme sur le champ de recherche, en particulier dans la décomposition fondamentale qu'il propose d'un effet total en ses composantes directe et indirecte.

2. MODÈLE MULTI-NIVEAU ET MISE AU JOUR D'UN EFFET CONTEXTUEL

Une part croissante de la recherche empirique en sciences sociales étudie et tente d'isoler « l'effet du quartier ou du voisinage » sur les comportements et les situations des individus, en particulier des jeunes. Elle se fonde, pour ce faire, sur des modélisations statistiques qui correspondent souvent à des formes sophistiquées de l'analyse de régression, connues sous le nom de « modèle multi-niveau ». L'une des premières revues de littérature est rédigée par les sociologues américains Christopher Jencks et Susan E. Mayer en 1990 : elle traite des conséquences sociales du fait d'avoir grandi dans un quartier pauvre. Dans la décennie 1990, les études empiriques se sont multipliées et des revues de travaux sont disponibles qui dressent un point actuel de l'avancée de la connaissance en ce domaine [Small, Newman, 2001 ; Dietz, 2002 ; Sampson, Morenoff, Gannon-Rowley, 2002]. Elles mettent l'accent sur les processus sociaux qui rendent compte des mécanismes par lesquels le contexte écologique pourrait affecter le comportement individuel – on distingue notamment les *théories de la contagion* qui insistent sur l'influence du groupe des pairs comme mécanisme générateur des effets de voisinage, et les *théories de la socialisation collective* pour lesquelles les modèles de rôle que les adultes de l'environnement proposent jouent un rôle central. Les mêmes revues de travaux s'appesantissent aussi sur les obstacles méthodologiques – à bien des égards « formidables », comme on le verra plus loin – qu'il faut s'efforcer de lever pour isoler et mesurer l'ampleur d'un effet de voisinage, obstacles dont la recherche empirique a pris conscience progressivement.

Le volume de cette littérature aux États-Unis doit sans doute être mis en relation avec l'importance qu'a revêtu la perspective d'écologie sociale dès les premiers écrits de Park et Burgess, et plus généralement dans toute l'École de Chicago. Dès les années 1960, des recherches sociologiques quantitatives étudient ainsi la question de l'influence du contexte sur la réussite ou les choix scolaires. Par exemple, Sewell et Armer [1966] ont examiné la thèse communément répandue selon laquelle le statut socio-économique du quartier influencerait fortement les aspirations scolaires des jeunes. Pour un échantillon scolarisé en 1957 en fin d'études secondaires dans les établissements publics de Milwaukee, il s'avère que les fortes différences initiales en matière de projet d'entrée à l'université sont grandement réduites lorsque le statut socio-économique de la famille, le sexe de l'élève et ses capacités intellectuelles sont simultanément contrôlés. Aussi les auteurs concluent-ils que le sens commun tend à surévaluer l'influence du contexte de voisinage sur le développement des aspirations scolaires. Plus généralement, Hauser [1970] a attiré l'attention sur la difficulté fondamentale à laquelle se heurte la mise en évidence d'effets contextuels : « En général, leur ampleur sera inversement proportionnelle au degré d'adéquation et de complétude du modèle sous-jacent de relation parmi les variables individuelles » (p. 662). En d'autres termes, le risque est réel d'interpréter comme un effet contextuel un artefact qui ne serait créé que comme la conséquence d'un modèle mal spécifié au niveau individuel.

Prenant en compte l'objection de Hauser, les études empiriques contemporaines – notamment celles relatives à la réussite scolaire – s'appuient donc typiquement sur des modèles de régression qui, outre des mesures caractérisant le voisinage (statut socio-

économique, taux de chômage, diversité ethnique, densité de l'habitat, etc.), incluent un nombre aussi élevé que possible de variables explicatives de niveau individuel ou familial. Les résultats obtenus s'appuient parfois sur l'analyse de données longitudinales et leur tonalité générale peut être résumée comme suit. Même si les variables qui décrivent le voisinage apparaissent souvent comme des prédicteurs statistiquement significatifs du développement des enfants, l'ampleur de leur effet est généralement beaucoup plus faible que celle associée aux variables de niveau familial, par exemple l'éducation de la mère ou le revenu de la famille. Par ailleurs, plus que la concentration de la pauvreté en elle-même, ce serait le fait d'avoir ou de ne pas avoir dans son voisinage des adultes de statut socio-économique élevé qui affecterait le devenir des enfants – un résultat que Brooks-Gunn *et al.* [1997] interprètent comme étant en faveur des théories de la socialisation collective.

L'étude sur données écossaises de Garner et Raudenbush [1991] fournit une illustration utile : elle examine s'il existe un effet spécifique du voisinage sur la réussite scolaire mesurée au niveau individuel. Sous l'angle méthodologique, elle a l'intérêt de mettre en œuvre un modèle de régression de type multi-niveau, encore appelé « modèle linéaire hiérarchique » (*hierarchical linear model*) ou « modèle à coefficients aléatoires » (*random-coefficient model*). En prenant explicitement en compte la structure hiérarchique des données – les jeunes sont inclus ou « emboîtés » dans les voisinages – ainsi que l'effet de grappe commun à tous les individus observés dans une même zone géographique, un tel modèle autorise des conclusions plus précises que le modèle classique de régression estimé par les moindres carrés ordinaires. L'analyse porte ainsi sur environ 2500 jeunes dont le niveau de réussite est mesuré à la fin de la scolarité obligatoire et qui appartiennent à 437 aires ou « voisinages », chacune pouvant être décrite par un score composite de pauvreté à partir du recensement de population. Les résultats obtenus sont les suivants. D'une part, la variance totale du niveau de réussite scolaire se décompose en 20 % entre les aires et 80 % entre les individus à l'intérieur des aires, ce qui confirme que « l'effet » de l'environnement doit être considéré comme de second ordre. D'autre part, la prise en compte d'une batterie très complète de variables explicatives au niveau individuel et familial⁴ rend compte de 85 % de la variance de réussite scolaire entre les aires : cela reflète le fait que les écarts observés entre celles-ci tiennent avant tout aux différences de performance initiale et d'appartenance socio-démographique des jeunes qui y résident. Enfin, l'ajout d'indicatrices pour les écoles fréquentées et du score composite de pauvreté explique 5 % supplémentaires de la variance de réussite scolaire entre les aires ; le coefficient de régression partiel qui lie score de pauvreté et réussite scolaire moyenne au niveau de l'aire (à variables individuelles et écoles contrôlées) est négatif et hautement significatif.

En raison du nombre et de la précision des « contrôles » statistiques qu'elle inclut, l'étude de Garner et Raudenbush tend donc à convaincre de l'existence d'un effet – relativement modeste, mais mesuré sur une période de seulement quatre ans – de la zone de résidence et, plus spécifiquement, du degré de pauvreté de celle-ci, sur la réussite scolaire. Elle ne lève pas pour autant l'ensemble des difficultés méthodologiques que la recherche d'un effet causal proprement contextuel doit affronter. Duncan *et al.* [1997] en ont dressé une liste relativement complète. Premièrement, il est possible que la structure additive particulièrement simple des modèles estimés induise en erreur. En effet, on ne peut *a priori* exclure l'existence d'interactions importantes entre conditions familiales et conditions environnementales.

⁴ Performances scolaires en fin d'école élémentaire, sexe, catégorie sociale du père, éducation du père, chômage éventuel du père, éducation de la mère, nombre de frères et sœurs, appartenance éventuelle à une famille monoparentale.

Deuxièmement, le fait que les familles aient une latitude de choix quant au voisinage dans lequel elles vivent implique qu'un biais d'endogénéité reste possible : si des caractéristiques non mesurées des familles les conduisent à *la fois* à sélectionner certains types de voisinage et à avoir des enfants qui connaissent tel ou tel développement, alors l'effet apparent du voisinage sur celui-ci, tel qu'il est appréhendé dans les modèles habituels, est susceptible de surévaluer ou de sous-évaluer l'effet « vrai » et il est impossible de prédire *a priori* la direction de ce biais. Troisièmement, les modèles estimés supposent que l'environnement de voisinage influence les enfants. Cependant, s'il influence *aussi* leurs parents, alors les mêmes modèles sous-estiment l'effet total du voisinage puisqu'ils ignorent le fait que celui-ci affecte les caractéristiques parentales. En réalité, le fait d'introduire dans les modèles des contrôles statistiques pour les caractéristiques familiales a pour conséquence que l'on estime *les seuls effets* du voisinage sur le développement de l'enfant qui sont *non corrélés* avec ses effets sur le développement familial. Quatrièmement, il est possible que, dans les modèles habituels, l'influence du voisinage soit sous-estimée en raison des effets suppressifs d'autres variables non mesurées. Cinquièmement, les modèles estiment habituellement un effet linéaire des caractéristiques du voisinage. Pourtant, on ne peut *a priori* exclure que cet effet ne soit en réalité pas linéaire, mais s'exprime plutôt à partir d'un certain seuil (*threshold effect*). Sixièmement, les modèles de régression classique ont l'inconvénient d'ignorer la structure typiquement multi-niveau des données analysées. Alors que la variable dépendante – une mesure du développement de l'enfant – est située au niveau individuel, certaines variables explicatives sont de niveau familial et d'autres de niveau quartier ou voisinage. Ne pas en tenir compte peut induire en erreur quant à la significativité statistique des effets de voisinage. En dépit de son intérêt, l'étude de Garner et Raudenbush ne prend explicitement en compte que cette dernière difficulté et laisse entière les cinq autres.

L'identification et la mesure précise d'un effet des caractéristiques de l'environnement social sur tel ou tel aspect du développement de l'enfant soulève donc des difficultés méthodologiques redoutables. Les prolongements récents de ce champ de recherche sont principalement intervenus dans deux directions. D'une part, alors que les premières études visaient la mise au jour d'un effet *contextuel* – au sens où la propension à adopter un comportement varierait en fonction des caractéristiques socio-démographiques du voisinage – on s'est mis à rechercher aussi un effet *endogène* – effet d'imitation, de renforcement ou du « groupe de référence », au sens où la propension à adopter un comportement varierait en fonction de la prévalence du même comportement dans le voisinage. D'autre part, l'effort a porté sur les aspects d'identification des effets et d'estimation des modèles, notamment sur la question du biais d'endogénéité – la deuxième difficulté. Par exemple, le chapitre de Duncan *et al.* [1997] inclut, sur les données longitudinales du *Panel Study of Income Dynamics*, une analyse qui prend appui sur une variable instrumentale assez convaincante : les caractéristiques du voisinage de la résidence de la mère une fois que tous les enfants – et non pas seulement celui dont on analyse la réussite – ont quitté le foyer. L'argument est ici que ce choix maternel de résidence ne sera plus influencé par une éventuelle préoccupation relative au devenir des enfants. Il peut donc être utilisé comme variable instrumentale, i.e. variable non ou peu corrélée avec une caractéristique parentale non mesurée et susceptible d'affecter la variable dépendante, mais corrélée avec la variable initiale d'environnement, en vue d'identifier un effet causal de ce dernier⁵. De façon plus

⁵ Aaronson [1997] présente une autre stratégie en vue de lever le biais lié au fait que les familles ne sont pas affectées aléatoirement à un voisinage, mais choisissent au moins en partie celui-ci pour des raisons

simple et pour lever la même question de l'endogénéité, une recherche suédoise a étudié l'effet, sur la réussite scolaire individuelle, de la composition socio-économique de l'établissement scolaire dans des zones géographiques ne comportant chacune qu'une seule école. Au total, même lorsqu'elles traitent sérieusement la question du choix du voisinage, les études récentes mettent en évidence un effet spécifique – relativement modeste – de l'environnement sur la réussite scolaire [Breen, Jonsson, 2005].

3. MODELE LOG-LINEAIRE ET VARIATION DE L'INEGALITE DES CHANCES DEVANT L'ENSEIGNEMENT

Au sein des sociétés développées et depuis les premières décennies du XX^e siècle, les cohortes de naissance successives ont été confrontées à une offre d'éducation fortement croissante – ces sociétés se caractérisent par une *distribution* considérablement élargie de l'éducation. Dans nombre d'entre elles, des réformes ont été aussi mises en œuvre pour ouvrir l'accès à celle-ci aux enfants de tous les milieux sociaux. Les sociologues se sont donc efforcés d'évaluer si le niveau d'éducation obtenu était progressivement devenu moins dépendant des caractéristiques héritées des individus, notamment leur origine sociale, et si une *allocation* moins inégale de l'éducation avait émergé dans nos sociétés. Un tel programme de recherche a pu être poursuivi en retenant un cadre analytique à la fois très simple et très général : celui de l'étude de tableaux de contingence, c'est-à-dire de tableaux croisés. Si, dans une enquête statistique, les individus interrogés sont invités à décrire à la fois le diplôme le plus élevé qu'ils ont obtenu et la catégorie socioprofessionnelle de leur père au cours de leur jeunesse, l'évolution éventuelle de l'inégalité des chances scolaires pourra être décelée dans la comparaison des tableaux de contingence bidimensionnels croisant origine sociale et diplôme le plus élevé obtenu, dressés pour des individus nés à différentes époques et regroupés, par exemple, en cohortes décennales de naissance⁶.

Une difficulté majeure doit cependant être résolue : la perception d'une évolution éventuelle du lien entre milieu d'origine et diplôme obtenu ne saurait être immédiate car de tels tableaux de contingence nous apparaissent comme littéralement bouleversés en quelques décennies. En effet, dans une société donnée, la distribution des origines sociales se transforme au fil des générations du fait de l'évolution de la structure des emplois ; de même et en raison de l'expansion scolaire, la distribution des niveaux d'éducation change considérablement au fil des cohortes. Bref, les marges des tableaux de contingence croisant origine sociale et diplôme obtenu sont fort variables d'une génération à une autre. C'est donc à la question suivante que les sociologues devaient répondre : quelle a été l'évolution du lien entre origine sociale et diplôme obtenu, *abstraction faite des transformations intervenues dans les distributions marginales* ? On voudrait ici retracer les grandes étapes des avancées de la statistique qui, de l'année 1900 à la fin du XX^e siècle, ont permis de répondre à une telle question.

L'histoire de la statistique retient généralement de l'année 1900 qu'elle a été celle de la découverte – ou de l'invention – de la statistique du chi-deux (χ^2) par Karl

éventuellement liées à l'importance qu'elles accordent au bien-être de leurs enfants. Il s'agit d'un modèle à effets fixes estimé sur des fratries.

⁶ À titre d'exemple, l'article de Thélot et Vallet [2000] étudie, pour la France, l'évolution de l'association statistique entre huit origines sociales et sept niveaux de diplôme au fil de treize cohortes quinquennales de naissance, de la plus ancienne (1908-1912) à la plus récente (1968-1972). De même, l'article de Vallet et Selz [2007] traite de l'évolution du même lien entre onze origines sociales et huit niveaux de diplôme au fil de dix-neuf cohortes triennales, de la plus ancienne (1920-1922) à la plus récente (1974-1976).

Pearson, chef du département de statistique de *University College* à Londres. Inspirée par ses travaux en biologie, la vision qu'avait Pearson de l'association statistique entre deux variables catégorielles consistait à supposer qu'une distribution continue bivariable sous-tend un tableau croisé. Il s'efforçait donc d'approcher la mesure que fournirait le coefficient de corrélation sur le continuum sous-jacent et, en suivant cette voie, il dérivait le coefficient de contingence et le coefficient de corrélation tétrachorique. George Udny Yule, qui était assistant de Pearson et réfléchissait sur des problèmes de science sociale, était convaincu au contraire que beaucoup de variables catégorielles sont intrinsèquement discrètes, par exemple vacciné ou non vacciné, vivant ou décédé, etc. Il s'employait donc à construire des indices d'association qui utilisaient directement les effectifs figurant dans les cases du tableau croisé et ne supposaient en rien l'existence d'un continuum sous-jacent. Et c'est aussi en 1900 que Yule propose la statistique de l'*odds ratio* – ou bi-rapport – de même qu'une mesure d'association qui lui est liée et qu'il nomme Q, en l'honneur du statisticien belge Quetelet.

Par exemple, dans une société fictive où n'existeraient que deux origines sociales (supérieure S ou populaire P) et un seul diplôme (que l'on peut donc obtenir ou ne pas obtenir), l'*odds ratio* exprime le rapport des chances d'être diplômé plutôt que non diplômé, pour les individus d'origine supérieure relativement à ceux d'origine populaire. Il s'écrit donc $(n_S/\tilde{n}_S)/(n_P/\tilde{n}_P)$, soit $n_S\tilde{n}_P/\tilde{n}_S n_P$ (Tableau a). Dans son article de 1912, Yule attirait l'attention sur une qualité du coefficient Q et de l'*odds ratio* qui lui semblait hautement désirable s'agissant d'une mesure d'association entre deux variables : le fait qu'elle n'est pas affectée et demeure invariante en cas de multiplication par une constante à l'intérieur des lignes et/ou des colonnes du tableau de contingence.

(a)	(b)	(c)	(d)																																	
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">n_S</td> <td style="padding: 2px 10px;">\tilde{n}_S</td> <td style="padding: 2px 10px;">N_S</td> </tr> <tr> <td style="padding: 2px 10px;">n_P</td> <td style="padding: 2px 10px;">\tilde{n}_P</td> <td style="padding: 2px 10px;">N_P</td> </tr> <tr> <td style="padding: 2px 10px;">N</td> <td style="padding: 2px 10px;">\tilde{N}</td> <td></td> </tr> </table>	n_S	\tilde{n}_S	N_S	n_P	\tilde{n}_P	N_P	N	\tilde{N}		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">αn_S</td> <td style="padding: 2px 10px;">$\alpha \tilde{n}_S$</td> <td style="padding: 2px 10px;">αN_S</td> </tr> <tr> <td style="padding: 2px 10px;">βn_P</td> <td style="padding: 2px 10px;">$\beta \tilde{n}_P$</td> <td style="padding: 2px 10px;">βN_P</td> </tr> </table>	αn_S	$\alpha \tilde{n}_S$	αN_S	βn_P	$\beta \tilde{n}_P$	βN_P	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">γn_S</td> <td style="padding: 2px 10px;">$\delta \tilde{n}_S$</td> <td></td> </tr> <tr> <td style="padding: 2px 10px;">γn_P</td> <td style="padding: 2px 10px;">$\delta \tilde{n}_P$</td> <td></td> </tr> <tr> <td style="padding: 2px 10px;">γN</td> <td style="padding: 2px 10px;">$\delta \tilde{N}$</td> <td></td> </tr> </table>	γn_S	$\delta \tilde{n}_S$		γn_P	$\delta \tilde{n}_P$		γN	$\delta \tilde{N}$		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">$\alpha \gamma n_S$</td> <td style="padding: 2px 10px;">$\alpha \delta \tilde{n}_S$</td> <td style="padding: 2px 10px;">αN_S</td> </tr> <tr> <td style="padding: 2px 10px;">$\beta \gamma n_P$</td> <td style="padding: 2px 10px;">$\beta \delta \tilde{n}_P$</td> <td style="padding: 2px 10px;">βN_P</td> </tr> <tr> <td style="padding: 2px 10px;">γN</td> <td style="padding: 2px 10px;">$\delta \tilde{N}$</td> <td></td> </tr> </table>	$\alpha \gamma n_S$	$\alpha \delta \tilde{n}_S$	αN_S	$\beta \gamma n_P$	$\beta \delta \tilde{n}_P$	βN_P	γN	$\delta \tilde{N}$	
n_S	\tilde{n}_S	N_S																																		
n_P	\tilde{n}_P	N_P																																		
N	\tilde{N}																																			
αn_S	$\alpha \tilde{n}_S$	αN_S																																		
βn_P	$\beta \tilde{n}_P$	βN_P																																		
γn_S	$\delta \tilde{n}_S$																																			
γn_P	$\delta \tilde{n}_P$																																			
γN	$\delta \tilde{N}$																																			
$\alpha \gamma n_S$	$\alpha \delta \tilde{n}_S$	αN_S																																		
$\beta \gamma n_P$	$\beta \delta \tilde{n}_P$	βN_P																																		
γN	$\delta \tilde{N}$																																			

Supposons en effet qu'à une date ultérieure, *seule* ait changé l'importance relative des catégories sociales. Dans le Tableau b, cela est représenté par le fait que les effectifs figurant sur la première ligne (origine supérieure) sont tous multipliés par une même constante α alors que ceux figurant sur la deuxième ligne (origine populaire) sont tous multipliés par une même constante β . L'*odds ratio* vaut donc désormais $(\alpha n_S)(\beta \tilde{n}_P)/(\alpha \tilde{n}_S)(\beta n_P)$, soit de nouveau et après simplification $n_S\tilde{n}_P/\tilde{n}_S n_P$.

De même et en repartant du tableau initial, on peut supposer qu'en raison d'une transformation du système scolaire, *seul* ait été modifié le taux d'obtention du diplôme. Ce sont donc ici les effectifs des deux colonnes qui sont multipliés par des constantes arbitraires, respectivement γ et δ (Tableau c). L'*odds ratio* vaut alors $(\gamma n_S)(\delta \tilde{n}_P)/(\delta \tilde{n}_S)(\gamma n_P)$, soit de nouveau et après simplification $n_S\tilde{n}_P/\tilde{n}_S n_P$.

Enfin, il est possible que les deux transformations – celle de la structure sociale et celle du système scolaire – interviennent simultanément (Tableau d). Il s'ensuit que l'*odds ratio* prend la valeur $(\alpha \gamma n_S)(\beta \delta \tilde{n}_P)/(\alpha \delta \tilde{n}_S)(\beta \gamma n_P)$ qui se réduit encore à $n_S\tilde{n}_P/\tilde{n}_S n_P$. Bref, l'*odds ratio* a la propriété remarquable d'être invariant sous des transformations des distributions marginales qui ne se traduiraient que par des accroissements ou des réductions proportionnels à l'intérieur des lignes et/ou des colonnes du tableau de

contingence. En ce sens, il s'agit bien d'une statistique qui mesure l'association au *cœur* du tableau croisé. Cette propriété confère donc à l'*odds ratio* une grande utilité dès lors que l'on souhaite apprécier des évolutions en faisant abstraction des variations intervenues dans les distributions marginales, ce qui est précisément le cas des études de l'inégalité des chances scolaires. Il permettra en effet de répondre à la question suivante : en tenant compte du fait que, de la cohorte c1 à la cohorte c2, la structure des origines sociales et celle des niveaux d'éducation ont évolué, quelque chose a-t-il changé dans le lien intrinsèque qui unit l'origine sociale des individus et le diplôme qu'ils ont obtenu ?

On sait que la proposition de Yule fut très mal reçue par Pearson, au point que la carrière du premier en a durablement souffert. Mais la fécondité de la statistique de Yule peut être appréciée en retraçant le développement de ce qui est connu aujourd'hui sous le nom de modélisation log-linéaire et log-multiplicative des tableaux de contingence. En 1935, le statisticien anglais Maurice Bartlett définit formellement l'absence d'interaction du second ordre dans un tableau de contingence croisant trois variables dichotomiques. Il s'agit de l'égalité stricte des *odds ratios* mesurant l'association entre deux variables à chacun des niveaux de la troisième variable – c'est-à-dire, dans l'exemple précédent, le fait que l'avantage des enfants d'origine supérieure sur ceux d'origine populaire pour les chances d'être diplômé plutôt que non diplômé demeure rigoureusement constant d'une cohorte à l'autre. En 1940, aux États-Unis, le statisticien William Deming et le démographe Frederick Stephan proposent un algorithme très simple pour ajuster un tableau de contingence bidimensionnel à des marges différentes tout en préservant l'association – c'est-à-dire l'ensemble des *odds ratios* – qui le caractérise⁷. Étudiant la mobilité sociale à Indianapolis en 1910 et 1940, Duncan [1966] utilise cet algorithme pour simuler, à partir de la table observée en 1910, la table hypothétique de 1940 pour le cas où toute l'évolution de la mobilité résulterait de la seule transformation des distributions marginales. Enfin, sous l'impulsion notamment des travaux de M. W. Birch et Leo A. Goodman, la modélisation log-linéaire des tableaux de contingence d'une dimension quelconque est proposée dans les années 1960 et 1970 [Bishop, Fienberg, Holland, 1975 ; Fienberg, 1977]. Elle constitue, selon les termes mêmes de Stephen E. Fienberg [1977, p. 5], le développement le plus extensif de l'école d'analyse des tableaux de contingence fondée par Yule.

Ce modèle statistique général a donc fourni le cadre analytique au sein duquel les sociologues de la stratification ont pu développer leur travail. Par exemple, pour étudier l'évolution de l'inégalité des chances devant l'enseignement en France depuis le début du XX^e siècle, Garnier et Raffalovich [1984] utilisent les données de l'enquête « Formation et Qualification Professionnelle » conduite par l'INSEE en 1970. Ils en extraient un tableau de contingence tridimensionnel croisant cinq origines sociales, sept niveaux de diplôme et six cohortes décennales de naissance, puis ils estiment, sur celui-ci, trois modèles log-linéaires qui ont pour propriété de reproduire fidèlement les marges d'origine sociale et de niveau d'éducation caractéristiques de chaque cohorte. Le premier modèle suppose l'indépendance statistique entre origine sociale et diplôme obtenu, au sein de chaque cohorte : l'ensemble des *odds ratios* valant 1, il traduit donc l'absence complète d'inégalité des chances devant l'enseignement. Le second modèle exprime l'hypothèse formulée par Bartlett : un lien existe entre origine sociale et

⁷ L'algorithme consiste en une succession d'ajustements proportionnels (ou « règles de trois »), opérés alternativement sur les lignes, puis sur les colonnes, puis de nouveau sur les lignes, etc. Il est également connu en France sous le nom d'algorithme RAS et sa généralisation (*Iterative Proportional Fitting*) permet d'obtenir les effectifs estimés sous un modèle log-linéaire particulier.

diplôme, mais il est *constant* au fil des cohortes⁸. Le troisième modèle, dit saturé, reproduit fidèlement les données observées et relaxe donc la dernière hypothèse⁹. Constatant que les estimations obtenues sous le modèle de constance réduisaient considérablement la distance du modèle d'indépendance conditionnelle aux données observées et qu'elles étaient même parfois acceptables au sens d'un test statistique, Garnier et Raffalovich concluaient à la quasi-absence de changement historique dans l'inégalité des chances devant l'enseignement au sein de la société française. Plus généralement, le même diagnostic a été établi dans un programme comparatif dirigé par Shavit et Blossfeld [1993]. Avec des méthodes similaires, il a étudié l'évolution, au cours du XX^e siècle, de l'inégalité sociale devant l'éducation dans six pays d'Europe de l'Ouest, trois d'Europe de l'Est et quatre d'autres continents dont les États-Unis. Sa conclusion générale – *Persistent Inequality* – a retenu l'attention : parmi les treize pays, ce n'est en effet qu'aux Pays-Bas et en Suède que les analyses statistiques décelaient une diminution historique dans l'association statistique entre origine sociale et niveau d'éducation obtenu.

Les conclusions précédentes ont cependant été infléchies dans des recherches plus récentes qui ont mobilisé des échantillons plus vastes et des modèles statistiques plus puissants. Car le cadre log-linéaire a l'inconvénient de fournir un test statistique peu puissant de l'interaction entre origine sociale, niveau d'éducation et cohorte : si ces variables comptent respectivement NO, NE et NC modalités, un tel test utilise (NO-1)(NE-1)(NC-1) degrés de liberté – soit un nombre souvent très grand – et est donc susceptible d'échouer à détecter une variation d'ampleur modeste. En recourant à un cadre plus complexe – le modèle log-multiplicatif d'un tableau de contingence – il a été possible de mettre en évidence une telle variation¹⁰. Cela a notamment été le cas pour la France [Smith, Garnier, 1986 ; Thélot, Vallet, 2000 ; Vallet, Selz, 2007].

À ce jour, la recherche comparative la plus récente – et aussi méthodologiquement la plus aboutie – sur la variation historique de l'inégalité sociale devant l'enseignement correspond à l'article de Breen, Luijkx, Müller et Pollak [2009]. Pour huit pays européens – Allemagne, France, Grande-Bretagne, Irlande, Italie, Pays-Bas, Pologne, Suède – et après un aperçu des raisons qui conduisent à anticiper une variation à la baisse de l'inégalité¹¹, les auteurs ré-analysent des données d'enquête nombreuses, initialement recodées dans les mêmes nomenclatures de classe sociale et d'éducation par des chercheurs des pays concernés, et qui permettent de retracer l'évolution du lien entre origine sociale et éducation pour les générations nées dans les deux premiers tiers du XX^e siècle. L'analyse utilise un modèle logistique qui considère le niveau d'éducation atteint comme une variable ordinaire. Elle conclut à un déclin de l'inégalité (*Nonpersistent Inequality*) pour l'ensemble des pays, sous la forme surtout d'un

⁸ Au sens où, dans les estimations d'un tel modèle, les *odds ratios* homologues, obtenus à partir de deux origines sociales et deux niveaux d'éducation, diffèrent de 1, mais sont identiques d'une cohorte à une autre.

⁹ Au sens où les *odds ratios* homologues sont susceptibles de varier librement d'une cohorte à une autre.

¹⁰ Par exemple, le modèle log-multiplicatif de différence uniforme suppose que la structure de l'association statistique entre origine sociale et niveau d'éducation est constante, mais que son intensité varie au fil des cohortes [Xie, 1992]. Il exprime cette variation entre deux cohortes au moyen d'un seul paramètre, fournissant donc une « lunette » particulièrement puissante pour détecter un éventuel changement.

¹¹ Par exemple, il est vraisemblable que la réduction des écarts sociaux en matière de santé et d'alimentation conduise à une réduction des différences de performances moyennes entre enfants de différentes classes sociales. De même, l'introduction de la gratuité dans l'enseignement secondaire et une offre scolaire moins rare et plus proche des familles sont susceptibles de réduire les écarts sociaux d'orientation, à niveau de performance donné.

moindre handicap des enfants des catégories agricoles et ouvrières. Dans chaque pays, la tendance déclinante est spécialement visible pour les générations nées entre 1935 et 1954 : elles ont parcouru les points de bifurcation scolaire majeurs au cours des trente années qui ont suivi la fin de la Seconde Guerre mondiale, période de forte croissance économique et d'amélioration des conditions de vie en Europe. Les auteurs prennent soin de vérifier, d'une part que leur conclusion générale est inchangée lorsque l'analyse incorpore en outre le niveau d'éducation du père ; d'autre part qu'elle est compatible avec les résultats d'une approche conditionnelle qui révèle, sur les mêmes données et dans la plupart des pays, un déclin de l'inégalité dans la transition initiale vers l'enseignement secondaire ; enfin que la conclusion de *Persistent Inequality* à laquelle Shavit et Blossfeld parvenaient résulte largement d'un manque de puissance statistique lié à l'usage d'échantillons nationaux d'une taille souvent trop faible. Sous l'angle de la comparaison internationale, les auteurs soulignent que le déclin observé n'a pas fait disparaître l'écart entre deux groupes de pays : d'un côté, la Grande-Bretagne, les Pays-Bas et la Suède où l'inégalité sociale devant l'école est d'intensité plus faible ; de l'autre, l'Allemagne, la France, l'Irlande, l'Italie et la Pologne où elle est plus forte. Les résultats précédents, relatifs à la population masculine, sont enfin très largement reproduits sur la population féminine.

4. CONCLUSION

Cet article a argumenté que, dans la sociologie empirique contemporaine qui se fonde sur des enquêtes quantitatives à large échelle, l'analyse de la variabilité et de l'inégalité est à la fois conduite et révélée dans le cadre de l'usage d'un modèle statistique qui fournit donc une base essentielle pour l'établissement des faits scientifiques. Cet argument a été illustré à partir de la présentation de trois exemples importants issus de la littérature empirique sur la stratification sociale, principalement, mais non exclusivement, de langue anglaise : l'usage de l'analyse en pistes causales pour le développement du modèle de l'acquisition du statut, dans la sociologie américaine des années 1960 et 1970 ; l'usage du modèle multi-niveau pour l'analyse des effets contextuels (ou de voisinage) sur le devenir des individus, dans la recherche en sciences sociales des années 1990 ; l'usage des modèles log-linéaire et log-multiplicatif d'un tableau de contingence pour mettre au jour et analyser les tendances temporelles de l'inégalité des chances sociales devant l'enseignement au sein d'une société.

BIBLIOGRAPHIE

AARONSON D. (1997), "Sibling Estimates of Neighborhood Effects", Brooks-Gunn J., Duncan G.J., Aber J.L. (eds), *Neighborhood Poverty. Volume II: Policy Implications in Studying Neighborhoods*, New York, Russell Sage Foundation, p. 80-93.

BARTLETT M.S. (1935), "Contingency table interactions", *Supplement to the Journal of the Royal Statistical Society* 2(2), p. 248-252.

BISHOP Y.M.M., FIENBERG S.E., HOLLAND P.W. (1975), *Discrete Multivariate Analysis. Theory and Practice*, Cambridge, MIT Press.

BLAU P.M., DUNCAN O.D. (1967), *The American Occupational Structure*, New York, Wiley.

BOLLEN K.A. (1989), *Structural Equations with Latent Variables*, New York, Wiley.

BOUDON R. (1973), *L'inégalité des chances. La mobilité sociale dans les sociétés industrielles*, Paris, Colin.

- BREEN R., JONSSON J.O. (2005), "Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility", *Annual Review of Sociology* 31, p. 223-243.
- BREEN R., LUIJKX R., MÜLLER W., POLLAK R. (2009), "Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries", *American Journal of Sociology* 114(5), p. 1475-1521.
- BROOKS-GUNN J., DUNCAN G.J., ABER J.L. (eds) (1997), *Neighborhood Poverty. Volume I: Context and Consequences for Children, Volume II: Policy Implications in Studying Neighborhoods*, New York, Russell Sage Foundation.
- CLOGG C.C. (1992), "The Impact of Sociological Methodology on Statistical Methodology", *Statistical Science* 7(2), p. 183-207 [with discussion by David J. Bartholomew, Paul W. Holland, Charles F. Manski, Ivo W. Molenaar].
- DEMING W.E., STEPHAN F.F. (1940), "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known", *Annals of Mathematical Statistics* 11(4), p. 427-444.
- DIETZ R.D. (2002), "The Estimation of Neighborhood Effects in the Social Sciences: An Interdisciplinary Approach", *Social Science Research*, 31(4), p. 539-575.
- DUNCAN G.J., CONNELL J.P., KLEBANOV P.K. (1997), "Conceptual and Methodological Issues in Estimating Causal Effects of Neighborhoods and Family Conditions on Individual Development", Brooks-Gunn J., Duncan G.J., Aber J.L. (eds), *Neighborhood Poverty. Volume I: Context and Consequences for Children*, New York, Russell Sage Foundation, p. 219-250.
- DUNCAN O.D. (1961), "A Socioeconomic Index for All Occupations" in REISS A.J. (ed.), *Occupations and Social Status*, New York, Free Press, p. 109-138.
- DUNCAN O.D. (1966), "Methodological Issues in the Analysis of Social Mobility", Smelser N.J., Lipset S.M. (eds), *Social Structure and Mobility in Economic Development*, Chicago, Aldine, p. 51-97.
- DUNCAN O.D., HODGE R.W. (1963), "Education and Occupational Mobility: A Regression Analysis", *American Journal of Sociology* 68(6), p. 629-644.
- FIENBERG S.E. (1977), *The Analysis of Cross-Classified Categorical Data*, Cambridge, MIT Press.
- GARNER C.L., RAUDENBUSH S.W. (1991), "Neighborhood Effects on Educational Attainment: A Multilevel Analysis", *Sociology of Education* 64(4), p. 251-262.
- GARNIER M.A., RAFFALOVICH L.E. (1984), "The Evolution of Equality of Educational Opportunities in France", *Sociology of Education* 57(1), p. 1-11.
- HAUSER R.M. (1970), "Context and Consex: A Cautionary Tale", *American Journal of Sociology* 75(4-2), p. 645-664.
- HAUSER R.M., WARREN J.R. (1997), "Socioeconomic Indexes for Occupations: A Review, Update, and Critique", *Sociological Methodology* 27, p. 177-298.
- JENCKS C., MAYER S.E. (1990), "The Social Consequences of Growing Up in a Poor Neighborhood", Lynn Jr. L.E., McGeary M.G.H. (eds), *Inner-City Poverty in the United States*, Washington, National Academy Press, p. 111-186.
- LAZARUSFELD P.F., HENRY N.W. (1968), *Latent Structure Analysis*, Boston, Houghton Mifflin.
- RAFTERY A.E. (2001), "Statistics in Sociology, 1950-2000: A Selective Review", *Sociological Methodology* 31, p. 1-45.
- SAMPSON R.J., MORENOFF J.D., GANNON-ROWLEY T. (2002), "Assessing 'Neighborhood Effects': Social Processes and New Directions in Research", *Annual Review of Sociology* 28, p. 443-478.
- SEWELL W.H., ARMER J.M. (1966), "Neighborhood Context and College Plans", *American Sociological Review* 31(2), p. 159-168.
- SEWELL W.H., HAUSER R.M., SPRINGER K.W., HAUSER T.S. (2004), "As We Age: A Review of the Wisconsin Longitudinal Study, 1957-2001", *Research in Social Stratification and Mobility* 20, p. 3-111.
- SHAVIT Y., BLOSSFELD H.-P. (eds) (1993), *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*, Boulder, Westview Press.
- SMALL M.L., NEWMAN K. (2001), "Urban Poverty after *The Truly Disadvantaged*: The Rediscovery of the Family, the Neighborhood, and Culture", *Annual Review of Sociology* 27, pp. 23-45.

- SMITH H.L., GARNIER M.A. (1986), "Association between Background and Educational Attainment in France", *Sociological Methods and Research* 14(3), p. 317-344.
- SOBEL M.E. (1996), "Clifford Collier Clogg, 1949-1995: A Tribute to His Life and Work", *Sociological Methodology* 26, p. 1-38.
- SOBEL M.E. (1998), "Causal Inference in Statistical Models of the Process of Socioeconomic Achievement: A Case Study", *Sociological Methods and Research* 27(2), p. 318-348.
- TACHIBANAKI T. (1977) « Mobilité sociale et formation en France », *Annales de l'Insee* 28, p. 129-143.
- THELOT C., VALLET L.-A. (2000), « La réduction des inégalités sociales devant l'école depuis le début du siècle », *Économie et Statistique* 334, p. 3-32.
- TREIMAN D.J. (1977), *Occupational Prestige in Comparative Perspective*, New York, Academic Press.
- TREIMAN D.J., YIP K.B. (1989), "Educational and Occupational Attainment in 21 Countries" in Kohn M.L. (ed.), *Cross-National Research in Sociology*, Beverly Hills, Sage, p. 373-394.
- VALLET L.-A., SELZ M. (2007), « Évolution historique de l'inégalité des chances devant l'école : des méthodes et des résultats revisités », *Éducation et Formations* 74, p. 65-74.
- WRIGHT S. (1921), "Correlation and Causation", *Journal of Agricultural Research* 20, p. 557-585.
- XIE Y. (1992), "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables", *American Sociological Review* 57(3), p. 380-395.
- YULE G.U. (1900), "On the association of attributes in statistics", *Philosophical Transactions of the Royal Society of London*, Series A 194, p. 257-319.
- YULE G.U. (1912), "On the methods of measuring association between two attributes", *Journal of the Royal Statistical Society* 75(6), p. 579-642.