



---

## L'apport des modèles multiniveaux à la recherche en éducation

*What Multilevel Models Bring to Educational Research*

Pascal Bressoux

---



### Édition électronique

URL : <http://journals.openedition.org/educationdidactique/168>

DOI : 10.4000/educationdidactique.168

ISBN : 978-2-7535-1615-1

ISSN : 2111-4838

### Éditeur

Presses universitaires de Rennes

### Édition imprimée

Date de publication : 1 septembre 2007

Pagination : 73-88

ISBN : 978-2-7535-0534-6

ISSN : 1956-3485

### Référence électronique

Pascal Bressoux, « L'apport des modèles multiniveaux à la recherche en éducation », *Éducation et didactique* [En ligne], 1-2 | septembre 2007, mis en ligne le 01 septembre 2009, consulté le 09 décembre 2020. URL : <http://journals.openedition.org/educationdidactique/168> ; DOI : <https://doi.org/10.4000/educationdidactique.168>

---

# L'APPORT DES MODÈLES MULTINIVEAUX À LA RECHERCHE EN ÉDUCATION

Pascal Bressoux,  
Laboratoire des Sciences de l'Éducation Université  
Pierre Mendès France & IUFM de Grenoble

Résumé : La modélisation des effets de l'environnement sur les individus pose des problèmes statistiques qui sont tout à fait redoutables. A cet égard, les modèles classiques par les moindres carrés ordinaires souffrent d'évidentes limites. Les modèles multiniveaux, plus récents, offrent quant à eux des solutions adéquates à ces problèmes et ouvrent des possibilités d'analyse très importantes en ce domaine. Cet article fait le point sur ces questions de modélisation statistique des effets de l'environnement sur les individus et montre, sur la base d'un exemple empirique précis, les avantages que procure la modélisation statistique multiniveau.

Mots-clés : modèle MCO ; modèle multiniveau ; effet de l'environnement ; agrégation ; désagrégation

*Pascal Bressoux*

La recherche en sciences de l'éducation s'affronte souvent à la question des effets que produisent certains dispositifs de formation sur les individus qui les fréquentent. Ce type de recherche entre dans une problématique générale des sciences sociales qui est celle des relations qui s'opèrent entre les individus et leur environnement. Comment la vie dans certains quartiers peut-elle affecter les comportements délinquants ? Comment la fréquentation de certains groupes peut-elle conduire à des phénomènes d'addiction ? Comment les milieux professionnels opèrent-ils une socialisation secondaire des individus ? Comment la prison peut-elle prévenir ou au contraire favoriser les récidives ? Comment l'environnement familial inculque-t-il certaines normes et valeurs aux enfants ? etc. S'il est bien des questions de recherche en sciences sociales qui se rapportent à cette thématique générale, elles se posent sous des considérations théoriques fort diverses et sous des formes méthodologiques très variées. C'est ce dernier point qui nous intéresse dans cet article : les aspects méthodologiques du traitement des effets de l'environnement sur les individus. Mais il nous faut être encore plus précis. L'objet du présent article est d'exposer des questions de *modélisation* des effets de l'environnement sur les individus. L'analyse scientifique consiste en effet pour l'essentiel à construire des modèles de la réalité. Ces modèles, bien que pluriels et variés, possèdent plusieurs éléments distinctifs. Il

s'agit tout d'abord d'une représentation schématique de la réalité. Le modèle est en ce sens une épure qui s'efforce de reproduire les propriétés essentielles d'un phénomène complexe. Deuxième élément distinctif, le modèle ne se contente pas de fixer les connaissances relatives à un phénomène. Il a une fonction instrumentale et heuristique ; il devient l'objet même de l'analyse en tant qu'il a été réalisé sous une forme commode, apte à être manipulée. Enfin, le modèle s'exprime dans un langage aussi rigoureux que possible, qui exclut toute forme de métaphore et qui conduit souvent à une certaine formalisation.

Cet article se focalise sur un type particulier de modélisation : la modélisation statistique. Outre le fait qu'un modèle statistique possède les éléments distinctifs énumérés ci-dessus, il a aussi pour caractéristique de permettre un raisonnement « toutes choses égales par ailleurs ». Cette clause est fondamentale en modélisation statistique et dans le raisonnement expérimental en général. On vise en effet par là à estimer l'effet spécifiquement attribuable à un facteur, après avoir contrôlé l'effet d'autres facteurs. De cette manière, on s'assure que la relation enregistrée entre un facteur et la variable-réponse n'est pas due à l'action exercée par une autre variable, elle-même liée (partiellement confondue) au facteur dont veut étudier l'effet sur la variable-réponse. Les modèles statistiques permettent d'estimer l'impact propre à chaque facteur, indépendamment de

celui des autres facteurs inclus dans le modèle. Bien évidemment, la clause « toutes choses égales par ailleurs » est un raccourci pour dire, dans le cas de la modélisation statistique, « toutes variables incluses dans le modèle égales par ailleurs ». Bien que restrictive par rapport à ce que serait un « toutes choses égales par ailleurs dans l'absolu », elle ne l'est cependant pas tant qu'il pourrait paraître au premier abord pour deux raisons essentielles. D'une part parce que le modèle est issu de considérations théoriques et qu'il propose toujours un certain niveau d'explication de la réalité sans prétendre l'épuiser de manière exhaustive, sa parcimonie étant même une condition de son efficacité et de sa robustesse. D'autre part, parce que les facteurs contrôlés dans le modèle entretiennent souvent des relations avec d'autres facteurs non inclus dans le modèle, ces derniers étant ainsi partiellement contrôlés. En sciences sociales, on sait par exemple que le contrôle de la profession des parents permet aussi de contrôler en filigrane et en partie le niveau de diplôme, le niveau de revenus, les activités culturelles, l'éducation familiale, etc.

Nous nous concentrerons ici plus précisément sur la question de la modélisation statistique des effets de l'environnement sur les individus. Cette modélisation pose en effet des problèmes très spécifiques qu'il a paru opportun de porter au jour ici. Après avoir pointé ces problèmes, nous montrerons que les modèles multiniveaux constituent une réponse adéquate et que ces modèles ont une capacité heuristique tout à fait remarquable pour la recherche en éducation. Nous prendrons nos exemples dans le cadre des recherches sur les effets-maîtres et les effets-écoles. Nous tenons toutefois à insister sur la généralité des modèles multiniveaux et leur application à d'autres thématiques traitant des relations entre les individus et leur environnement comme nous essaierons de le montrer.

### **Modéliser les effets de l'environnement sur les individus**

Modéliser les effets de l'environnement sur les individus revient à traiter des données qui relèvent de deux niveaux différents : certaines concernent l'environnement et sont donc des caractéristiques globales ou agrégées, tandis que les autres sont des données individuelles et se situent à un niveau inférieur à celui de l'environnement dans la mesure où ce dernier englobe plusieurs individus (une famille comporte

plusieurs membres, une classe plusieurs élèves, un quartier de nombreux habitants, une prison plusieurs détenus, une entreprise plusieurs employés, etc.). Dans cette conception, nous avons affaire à une structure hiérarchisée composée de différents niveaux, les unités de niveau inférieur étant « groupées » dans les unités de niveau supérieur. Revenons sur les caractéristiques de l'environnement ; nous avons dit qu'elles étaient globales ou agrégées. Cela revient à dire que l'environnement peut être caractérisé par des variables globales ou agrégées. Il convient de les définir.

Les variables agrégées caractérisent le groupe dans lequel sont insérés les individus ; elles procèdent de l'agrégation des caractéristiques individuelles des membres du groupe. Il peut s'agir par exemple de la tonalité sociale du groupe exprimée par le pourcentage de certaines catégories sociales, ou bien du pourcentage d'étrangers, etc. Il peut s'agir de caractériser un niveau moyen du groupe : salaire moyen (ou médian) au sein d'une entreprise, niveau d'acquisition moyen des élèves d'une classe, etc. Il peut s'agir aussi de caractériser l'hétérogénéité du groupe : on se sert souvent alors de l'écart-type, ou de l'écart interquartiles pour indiquer par exemple l'hétérogénéité des élèves d'une classe, l'hétérogénéité des salaires au sein d'une entreprise, l'hétérogénéité de l'expérience au combat au sein de groupes de soldats, etc. En résumé, les variables agrégées résultent d'opérations faites sur les caractéristiques des membres du groupe de manière à en faire des indicateurs du groupe lui-même. Les variables agrégées montrent ainsi le caractère non substituable des individus qui composent le groupe : changer un seul individu du groupe revient à modifier la caractéristique agrégée de ce groupe, même si cette modification se révèle parfois très mineure. Les variables agrégées indiquent des effets de composition que l'on nomme aussi souvent des effets de contexte, à savoir l'impact de la composition du groupe sur les comportements des individus, au-delà de leurs propres caractéristiques individuelles : « Il y a effet de contexte quand les propriétés collectives d'un public d'élèves ont un effet sur les résultats des élèves, au-delà de l'effet de leurs caractéristiques individuelles » (Willms, 1985).

Les variables globales ne proviennent pas quant à elles de l'agrégation de données individuelles ; elles relèvent strictement du niveau supérieur, celui de l'environnement, et elles n'identifient pas le groupe en tant que tel mais le « traitement » auquel il est soumis. Par exemple, tous les élèves d'une même

classe sont exposés au même enseignant ; tous les prévenus d'une même prison sont exposés à la même vétusté des locaux, etc. Les variables globales traduisent alors ce qu'Erbring et Young (1979) nomment un « destin commun » des individus qui appartiennent au même environnement. Ceux-ci sont exposés au même « traitement » (même maître, même chef d'atelier, même dotation budgétaire, mêmes bâtiments délabrés, etc.) mais, en ce cas, l'effet traduit un destin partagé qui, paradoxalement, n'implique pas fondamentalement la présence des autres membres du groupe. Changer un ou plusieurs membres du groupe ne modifie pas la caractéristique globale de l'environnement. Ce ne sont plus tout à fait les mêmes personnes qui y sont exposées, c'est tout.

Ce qui doit apparaître clair maintenant, c'est que le fait de traiter des effets de l'environnement sur les individus conduit inévitablement à mettre en relation des variables qui relèvent de niveaux distincts, qui ne caractérisent pas les mêmes unités statistiques : certaines variables relèvent du niveau des individus tandis que d'autres relèvent de l'environnement dans lequel ils sont insérés.

De plus, il faut préciser que les structures hiérarchisées peuvent comporter davantage de niveaux, certains environnements étant parfois eux-mêmes englobés dans des environnements plus vastes : par exemple, des détenus font partie de différentes cellules au sein d'une même prison. Ainsi, la structure est-elle composée de trois niveaux hiérarchisés distincts : des détenus (niveau 1), des cellules (niveau 2) et des prisons (niveau 3). En ce cas, on serait donc conduit à traiter des variables qui relèvent de trois niveaux distincts. On peut même envisager davantage de niveaux dans la structure. Ainsi, les élèves (niveau 1) font partie de classes (niveau 2) qui elles-mêmes appartiennent à des écoles (niveau 3), qui appartiennent elles-mêmes à des circonscriptions (niveau 4) comprises dans des académies (niveau 5)... On conçoit<sup>1</sup> ainsi l'existence de structures très complexes. Or, le traitement de données qui se situent à des niveaux différents de la hiérarchie pose des problèmes d'analyse redoutables comme nous allons le montrer maintenant.

### Les insuffisances de la régression par les moindres carrés ordinaires

La modélisation statistique des effets de l'environnement sur les individus se fait en régressant une variable-réponse sur une ou plusieurs variables

supposées explicatives. Les méthodes classiques pour traiter cette question ont longtemps consisté à utiliser l'analyse de régression par les moindres carrés ordinaires (MCO). Étant donné qu'on suppose en général qu'une variable-réponse est influencée par plusieurs variables explicatives  $X_1, X_2, \dots, X_k$ , on utilise la régression multiple qu'on formalise de la manière suivante

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (1)$$

où

l'indice  $i$  indique que l'équation s'applique à tous les individus  $i$  concernés

$\beta_0$  est la constante du modèle (i.e. la valeur sur la variable-réponse quand les valeurs des variables  $X_1, X_2, \dots, X_k$  sont nulles)

$\beta_0, \beta_1, \beta_k$  sont les coefficients directeurs de la droite de régression (i.e. les « effets » des variables explicatives  $X_1, X_2, \dots, X_k$  sur la variable-réponse  $Y_i$ )

$\varepsilon_i$  est le terme d'erreur du modèle (l'écart entre les valeurs estimées par le modèle et les valeurs observées).

Le terme d'erreur représente, pour chaque individu  $i$  de la population, l'écart entre l'estimation fournie par le modèle et l'observation effectivement réalisée. Évidemment, plus le terme d'erreur est faible, meilleur est le modèle dans sa capacité à ajuster les observations. Le terme d'erreur  $\varepsilon_i$  est fondamental dans la mesure où c'est lui qui définit le caractère « statistique » du modèle. Il en est ainsi parce que sa distribution fait l'objet d'hypothèses. Le terme d'erreur est supposé suivre une loi normale de moyenne nulle et de variance constante (i.e. hypothèse d'homoscédasticité : la variance des erreurs est supposée constante en fonction des valeurs des variables explicatives). Les erreurs sont supposées indépendantes les unes des autres. Or, l'analyse de données hiérarchisées par des modèles de régression MCO comporte de sérieuses lacunes et limites.

### La non-indépendance des erreurs

Comme nous venons de le voir, le modèle de régression MCO suppose une indépendance des erreurs. Cela signifie que la connaissance de l'erreur liée à un individu  $i$  ne peut pas permettre de prédire l'erreur liée à un autre individu  $i'$ . Voilà qui est pourtant strictement contraire aux effets supposés de l'environnement sur les individus : on suppose au

contraire que l'environnement va « modifier » les individus de telle manière qu'une non-indépendance va s'établir entre eux du point de vue de la variable-réponse étudiée. Un cas typique est celui de l'effet-maître : si certains enseignants sont plus « efficaces » que d'autres, on suppose que la plupart des élèves de leur classe vont en profiter, rendant ainsi leurs acquisitions plus proches entre elles (non indépendantes donc) qu'elles ne le sont entre élèves qui appartiennent à différentes classes. C'est au fondement même de l'analyse des effets de l'environnement sur les individus que de supposer une non-indépendance des erreurs.

On distingue en général plusieurs sources de non-indépendance. Tout d'abord, les individus qui appartiennent à un même groupe tendent souvent à se ressembler dès avant leur fréquentation dudit environnement ; par exemple, les élèves d'une école de centre-ville se « ressemblent » plus entre eux, du point de vue du milieu économique, culturel, etc., qu'ils ne « ressemblent » à des élèves qui fréquentent une école de banlieue défavorisée en ZEP. Par ailleurs, les individus d'un même environnement interagissent, se comparent entre eux et peuvent ainsi s'influencer mutuellement. De plus, les élèves partagent des caractéristiques communes de leur environnement (le « destin commun »), le même enseignant par exemple, qui peut exercer des effets sur leurs acquisitions. Tous ces effets conduisent au fait que, en règle générale il n'y a pas indépendance des erreurs. Or, le fait de violer cette hypothèse du modèle MCO conduit à sous-estimer les erreurs-types des coefficients<sup>2</sup>, rendant ainsi plus élevé le risque de trouver des relations significatives alors qu'elles n'existent pas dans la réalité (autrement dit, on augmente le risque d'erreur de première espèce, qui consiste à rejeter à tort l'hypothèse nulle). Rappelons à ce propos qu'il existe en statistique deux types d'erreurs : celle qui consiste à déclarer significative une relation qui ne l'est pas, c'est-à-dire à rejeter l'hypothèse nulle alors qu'elle devrait être conservée. On la nomme erreur de première espèce. L'autre erreur consiste à ne pas déclarer significative une relation qui existe dans les faits, c'est-à-dire à ne pas rejeter l'hypothèse nulle alors qu'elle devrait l'être. On l'appelle erreur de seconde espèce. Elle est opposée à la première : choisir de diminuer le risque d'erreur de première espèce (en abaissant le seuil de significativité de 5 % à 1 % par exemple) conduit à augmenter le risque d'erreur de seconde espèce. La seule manière

de réduire le risque d'erreur de première espèce sans augmenter le risque d'erreur de seconde espèce consiste à augmenter la taille de l'échantillon.

### ***L'hypothèse d'homoscédasticité mise en question***

Le modèle de régression MCO repose aussi sur l'hypothèse que les erreurs doivent avoir une variance constante. Or, il est bien souvent difficile de dire *a priori* si cette hypothèse est réaliste ou non. Aussi faudrait-il pouvoir la tester et traiter l'hétéroscédasticité dans les cas où elle apparaît car, là encore, la violation de cette hypothèse conduit à sous-estimer les erreurs-types des coefficients de régression.

Prenons l'exemple des recherches sur l'effet-maître. Elles consistent à examiner dans quelle mesure, et pourquoi, les acquis scolaires sont plus élevés dans certaines classes que dans d'autres. Autrement dit, on se demande si, à niveau initial donné, les acquis finaux diffèrent selon la classe. On traite là d'une efficacité globale de la classe, en faisant comme si cette efficacité était la même pour tous les élèves de la classe. Or, on peut aussi se demander si les classes réservent un « traitement » identique à tous leurs élèves. Existe-t-il des classes plus efficaces avec certaines catégories d'élèves que d'autres ? Cela revient à se demander si la relation entre le score initial et le score final est la même d'une classe à l'autre. Y a-t-il par conséquent hétérogénéité des relations entre ces deux variables ? Supposons maintenant que les différences d'efficacité entre les classes soient plus marquées pour les élèves faibles que pour les élèves forts. Cela entraînerait que la variance des erreurs serait plus élevée pour les élèves faibles (justement parce que cela conduirait à davantage d'hétérogénéité chez ces élèves) mais moins élevée pour les élèves forts. Autrement dit, l'hypothèse d'homoscédasticité des erreurs ne serait plus tenable : il y aurait alors hétéroscédasticité. Il faudrait par conséquent que l'on dispose de modèles qui soient aptes à tester et à modéliser cette éventuelle hétéroscédasticité car elle est potentiellement un phénomène très intéressant à étudier.

### ***La question de l'agrégation ou de la désagrégation des données***

Nous l'avons vu, étudier les effets de l'environnement sur les individus conduit à traiter des données

qui relèvent de niveaux distincts : elles ne reposent pas sur les mêmes unités d'analyse. Dans le cadre des modèles MCO, deux possibilités principales s'offrent alors<sup>3</sup> : soit on désagrège les données qui caractérisent les environnements au niveau des unités individuelles, soit on agrège les données individuelles au niveau des unités d'environnement. Or, chacun des termes de cette alternative conduit à des solutions qui se révèlent insatisfaisantes.

Dans le cas de la désagrégation, chacun des individus va se voir attribuer la caractéristique de son environnement. Du coup, ladite caractéristique ne caractérise plus des environnements mais des individus, soit un nombre d'unités d'analyse beaucoup plus important. On assiste ainsi à une « multiplication miraculeuse du nombre d'unités » (Snijders & Bosker, 1999, p. 15). Les tailles d'échantillon sont donc considérablement exagérées pour les macro-unités et le risque d'erreur de première espèce est, de ce fait, considérablement accru.

Dans le cas de l'agrégation des données individuelles au niveau des environnements, on perd toute information sur les mécanismes intra-environnements et on est contraint de raisonner sur des macro-relations qui sont les relations inter-environnements. Or, on ne peut jamais inférer directement ce qui se passe au niveau individuel à partir d'observations au niveau agrégé.

Faire une telle inférence risque de conduire à ce qu'il est convenu de nommer l'erreur écologique (ou biais d'agrégation). Robinson (1950) est le premier auteur à avoir formalisé ce problème. Il avait montré par exemple que, si la corrélation entre la proportion de noirs dans différentes régions géographiques des Etats-Unis et le taux d'illettrisme dans ces mêmes régions était très forte (elle était égale à 0,946), en revanche, au niveau individuel, la corrélation entre la couleur de peau et le fait d'être illettré était beaucoup plus faible (elle était égale à 0,203). Il aurait donc été tout à fait erroné de se fonder sur la corrélation agrégée pour en inférer ce qui se passait au niveau individuel. Selon qu'on raisonne sur les individus (micro-unités) ou sur les environnements (macro-unités), les résultats peuvent varier considérablement tant en ampleur qu'en direction : on peut ainsi voir une relation négative à un niveau agrégé masquer une relation positive au niveau individuel et inversement (en ce cas, on parle d'effet *flip-flop* ou encore d'effet *see-saw*). Les variables n'ont pas

en effet la même signification selon qu'elles caractérisent des individus ou des environnements. Ainsi, le niveau initial des élèves traduit une habileté individuelle tandis que la variable agrégée correspondante, le niveau initial moyen de la classe, peut traduire un effet de contexte (compétition, norme de groupe, pression à réussir...).

### ***L'hétérogénéité des relations***

On est bien souvent intéressé, non pas seulement par la relation globale qui s'établit entre une caractéristique environnementale et le comportement (ou toute attitude, action, croyance, émotion, etc.) des individus, mais aussi par les variations éventuelles de cette relation en fonction des caractéristiques des individus.

Par exemple, dans la recherche sur l'effet-maître, l'efficacité des enseignants est susceptible de varier en fonction des caractéristiques de leurs élèves : certains sont-ils plus efficaces avec les faibles, d'autres avec les forts, d'autres encore le sont-ils globalement avec tous leurs élèves, etc. Cela revient à se poser la question de l'hétérogénéité des relations au sein des classes. Or, les modèles MCO sont très limités dans leur capacité à traiter cette question, notamment quand le nombre de groupes ou d'environnements est élevé (ce qui est typiquement le cas dans les recherches sur l'effet-maître puisqu'il y a assez peu d'élèves par classe et que le nombre de classes étudiées est en général élevé). Ils sont incapables en particulier de parvenir à une modélisation globale de ce phénomène<sup>4</sup>.

### ***Effets fixes et effets aléatoires***

Dans un modèle MCO, les effets de l'environnement sont traités comme des effets fixes. Les effets fixes sont des effets non stochastiques qui relèvent d'un ensemble fini de modalités d'un facteur. On suppose alors que ce sont ces modalités particulières qui nous intéressent en tant que telles, pour elles-mêmes. Par exemple, si l'on étudie la relation entre les contacts sociaux et le statut matrimonial (marié, célibataire, vie maritale...), on sera spécifiquement intéressé par chacune de ces modalités. De même, si l'on étudie l'effet d'un médicament sur la guérison d'une maladie en constituant un groupe expérimental, un groupe placebo et un groupe témoin, on

sera intéressé par l'effet du traitement particulier qui a été adressé à chacun des groupes de patients.

Or, il arrive bien souvent que les environnements présents dans une étude (telle classe, tel quartier, telle entreprise, etc.) ne nous intéressent pas pour eux-mêmes, mais en tant qu'unités d'une population plus vaste qu'ils sont censés représenter. Autrement dit, les macro-unités ont fait l'objet d'un tirage aléatoire. En ce cas, on n'est plus intéressé par des effets fixes, mais par des effets aléatoires, c'est-à-dire des effets qui relèvent d'un ensemble infini (ou au moins très grand) de modalités d'un facteur, dont un échantillon seulement aura été observé. Par exemple, si l'on étudie les disparités de salaires en fonction des entreprises, telle entreprise particulière n'intéressera pas nécessairement en tant que telle, mais en tant que l'une des entreprises choisies aléatoirement parmi une population d'entreprises. La distinction entre effets fixes et effets aléatoires dépend donc du fait que l'on veuille ou non émettre l'hypothèse que les modalités d'un facteur ont été choisies aléatoirement à partir d'une distribution.

Dans le cas des effets fixes, on étudie des moyennes, tandis que dans le cas des effets aléatoires, on étudie des variances. On fait alors des hypothèses sur la distribution des effets aléatoires. Les plus courantes sont que les effets aléatoires suivent une distribution normale, de moyenne nulle et de variance quelconque (à estimer donc). Or, on ne réalise pas les mêmes inférences sur des moyennes et sur des variances. Dans un modèle MCO, où l'on estime typiquement des effets fixes, le test  $F$  a pour hypothèse nulle  $H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$ . Or, dans le cas d'effets aléatoires, on s'intéresse aux variances et donc à  $H_0 : \sigma_u^2 = 0$  (où  $\sigma_u^2$  représente la variance des effets aléatoires). Estimer des effets fixes permet seulement de faire des inférences sur les modalités qui sont effectivement manipulées dans l'étude ; ces effets sont supposés mesurés sans erreur. Au contraire, estimer des effets aléatoires permet de faire des inférences au-delà des modalités qui sont strictement observées dans l'étude ; ces effets sont supposés affectés d'erreurs d'échantillonnage.

Il arrive bien souvent, dans la recherche des effets de l'environnement sur les individus, que les différents environnements soient choisis aléatoirement parmi un ensemble d'environnements. C'est typiquement le cas des recherches sur les effets-maîtres, où l'on n'est pas intéressé par la classe de M. Untel ou

de Mme Unetelle, mais par la variance des effets dans une population d'enseignants. Les modèles MCO sont en ce cas clairement inadaptés puisqu'ils estiment des effets fixes quand on devrait rendre compte du caractère aléatoire des effets. Ce faisant, en ne « débarrassant » pas les estimations de leur variance d'échantillonnage, ils conduisent à les surestimer (cf. Bressoux, Coustère & Leroy-Audouin, 1997).

### Présentation du modèle multiniveau

Les problèmes soulevés par l'analyse MCO, mis en évidence depuis longtemps, n'ont connu de solution satisfaisante que dans le courant des années 1980, à la suite des progrès réalisés dans le domaine de la statistique (notamment les modèles mixtes ; cf. Searle, Casella & McCulloch, 1992), les réflexions sur les modèles de contexte et l'augmentation des capacités de calcul informatique. Quatre équipes de recherche sont parvenues à développer, à peu près à la même période, des modèles qui, bien que n'utilisant pas les mêmes algorithmes, proposaient des solutions satisfaisantes et produisaient des estimations proches (Aitkin & Longford, 1986 ; Goldstein, 1986 ; Mason, Wong & Entwisle, 1983 ; Raudenbush & Bryk 1986). Ces modèles sont principalement connus sous le nom de modèles multiniveaux, modèles hiérarchiques linéaires, ou encore modèles à coefficients aléatoires. Il est remarquable pour notre propos que ces modèles ont été conçus à l'origine, pour l'essentiel, dans le cadre des recherches en éducation traitant d'effets-écoles et effets-maîtres qui représentent, il est vrai, un cas très exemplaire de données hiérarchisées.

Nous présentons maintenant le modèle multiniveau. Il s'agit d'un modèle très souple qui offre de nombreux intérêts par rapport aux modèles MCO. Tout d'abord, il permet d'abandonner ou d'amender certaines hypothèses très restrictives des modèles MCO. Ainsi, le modèle multiniveau va autoriser une non-indépendance des erreurs au sein de chaque macro-unité. Le modèle multiniveau remplace l'hypothèse d'homoscédasticité par une hypothèse plus faible selon laquelle la variance des erreurs peut varier comme une fonction linéaire ou non linéaire des variables explicatives. Cela revient à étudier une possible hétérogénéité des relations entre les différents groupes. Le modèle multiniveau permet de traiter les effets des environnements comme des effets aléatoires. Le modèle multiniveau n'enjoint plus de choisir un niveau d'analyse (par agrégation

ou désagrégation des données) car les unités sont considérées à leur « bon » niveau et les estimations se font simultanément pour les différents niveaux<sup>5</sup>. Le modèle multiniveau peut être formalisé en distinguant dans un premier temps, pour simplifier la présentation, les écritures en fonction des différents niveaux impliqués dans l'analyse.

Au niveau 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (2)$$

Où les indices  $i$  et  $j$  renvoient aux individus  $i$  (les micro-unités) et aux environnements  $j$  (les macro-unités).

Les coefficients de régression  $\beta_{0j}$  et  $\beta_{1j}$  se voient affublés d'un indice  $j$  qui indique qu'ils peuvent varier d'un environnement  $j$  à l'autre.

Au niveau 2, on montre que les coefficients  $\beta_{0j}$  et  $\beta_{1j}$  sont aléatoires en introduisant les termes d'erreurs aléatoires  $u_{0j}$  et  $u_{1j}$

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \quad (3)$$

Ce qui donne en intégrant dans une seule équation

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij} \quad (4)$$

où

$\gamma_{00}$  représente la moyenne générale de  $Y$

$\gamma_{10}$  est la pente de régression moyenne pour tous les groupes

$u_{0j}$  représente une erreur aléatoire associée à chaque groupe  $j$  supposée normalement distribuée, de moyenne nulle et de variance  $\sigma_{u0}^2$

$u_{1j}$  représente l'écart de chaque groupe à la relation moyenne. C'est une variable aléatoire de moyenne nulle et de variance  $\sigma_{u1}^2$ .

$e_{ij}$  représente une erreur aléatoire associée à chaque individu  $i$  supposée normalement distribuée, de moyenne nulle et de variance  $\sigma_e^2$ .

Du fait qu'il y a maintenant deux termes d'erreurs au niveau 2, on peut estimer un paramètre supplémentaire : la covariance entre les constantes et les pentes, notée  $\sigma_{u01}$ .

L'écriture suivante montre de manière condensée la conception de la structure des paramètres aléatoires à estimer

au niveau 2

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix} \right)$$

au niveau 1,  $e_{ij} \sim N(0, \sigma_e^2)$

où le signe  $\sim$  indique « suit une distribution ». Par exemple, les erreurs de niveau 2, à savoir  $u_{0j}$  et  $u_{1j}$ , suivent une loi normale ( $N$ ) bivariée, de moyenne nulle, de variances  $\sigma_{u0}^2$  et  $\sigma_{u1}^2$  et de covariance  $\sigma_{u01}$ .

On montre bien ici que la structure des erreurs du modèle implique l'estimation de quatre paramètres aléatoires, trois au niveau 2 et un au niveau 1.

Remarquons que, dans l'équation (4), le terme  $u_{1j}X_{ij}$  représente un terme aléatoire d'interaction entre l'environnement et la variable  $X$ . Nous verrons plus loin les implications importantes de ce terme lorsque nous détaillerons les composants de la variance du modèle.

Il est évidemment possible d'intégrer des variables explicatives de niveau 2 pour expliquer les constantes et les pentes ; il est de même possible de complexifier encore la structure des paramètres aléatoires car on peut intégrer davantage de termes aléatoires au niveau 1. On se retrouve alors avec une (ou plusieurs) variance(s) supplémentaire(s) à estimer et autant de covariances avec les autres paramètres aléatoires. On sera donc davantage limité *in fine* par notre propre capacité à interpréter les résultats obtenus que par la capacité du modèle à estimer des structures de plus en plus complexes (sous réserve que la taille de l'échantillon soit suffisante, en particulier le nombre d'unités de niveau supérieur).

## Un exemple d'application

Nous présentons maintenant un exemple d'analyse multiniveau en pointant ses apports en comparaison avec les estimations du modèle MCO.



### Présentation de l'étude

Nous utilisons pour ce faire une étude qui a porté sur 332 élèves de CE2 répartis dans 16 classes. Dans cette étude, il s'agissait de déterminer quelles sont les variables qui influencent le jugement que les enseignants portent sur la valeur de leurs élèves en mathématiques. L'hypothèse était que l'enseignant ne fonde pas seulement son jugement sur les performances avérées des élèves mais sur tout un système d'informations qui fait intervenir des éléments contextuels tels que le niveau moyen de la classe et des éléments individuels caractérisant les élèves du point de vue de leurs attributs sociodémographiques, de leur scolarité et de leur intégration de certaines normes sociales générales. Concernant ce dernier élément, nous avons travaillé sur la norme d'internalité. Rappelons que celle-ci est conceptualisée comme une norme sociale générale de jugement qui consiste en la valorisation des explications internes pour expliquer ce que font les individus ou ce qui leur arrive. Les explications internes sont celles qui renvoient aux forces de l'individu lui-même dans ce qu'il fait ou dans ce qui lui arrive : ses efforts, son potentiel, son humeur, etc. Les explications externes renvoient aux forces extérieures à l'individu qui déterminent ce qu'il fait ou ce qui lui arrive : le hasard, la chance, le destin, autrui, etc. La théorie de la norme d'internalité postule que les explications causales internes sont les plus valorisées (même pour expliquer des événements négatifs) et que les individus qui se montrent les plus internes dans leurs explications sont les mieux jugés (Beauvois, 1984).

Le protocole de l'étude était le suivant. Les enseignants devaient prononcer un jugement sur la valeur scolaire en mathématiques de chacun de leurs élèves sur une échelle allant de 0 (élève très faible) à 10 (élève très fort). Nous avons collecté des renseignements sociodémographiques sur les élèves (sexe, profession des parents) et scolaires (déroulement de la scolarité). Nous avons également saisi les scores obtenus par les élèves aux épreuves nationales de CE2 en mathématiques. Nous avons enfin fait remplir aux élèves un questionnaire d'internalité (cf. Bressoux & Pansu, 2003). Ce questionnaire proposait des événements scolaires (comportements et renforcements) auxquels étaient associées deux explications, l'une interne, l'autre externe, entre lesquelles les élèves devaient choisir celle qui leur convenait le mieux. Par exemple :

« Lorsqu'ils ont une bonne note [renforcement positif] : Certains élèves disent : " C'est parce que le maître est gentil" [explication externe]

Certains élèves disent : " C'est parce que j'ai bien travaillé" [explication interne] »

Les réponses fournies par les élèves permettent de les ranger sur une échelle d'internalité allant de 0 (aucune explication interne) à 16 (toutes les explications sont internes).

### Modélisation MCO

Nous commençons d'abord par une modélisation MCO afin de voir ses limites et d'apporter des comparaisons concrètes avec la modélisation multiniveau, à la fois en termes de spécification de modèles et en termes d'estimations.

Si l'on veut estimer les écarts de jugement entre les classes, la modélisation MCO consiste à introduire une batterie de variables muettes<sup>6</sup> qui représentent les classes. Il faut signaler que, en ce cas, on peut contrôler les caractéristiques individuelles des élèves qui jouent un rôle sur le jugement mais qu'il est impossible d'y introduire des variables contextuelles (celles-ci devenant alors redondantes avec la batterie de variables muettes).

Les variables incluses dans le modèle sont le score aux épreuves standardisées, la profession du père, le retard scolaire et le score d'internalité, auxquelles on ajoute la batterie de variables muettes qui représentent les classes. Le modèle explique 60,66 % de la variance totale du jugement. Le modèle qui inclut les mêmes variables explicatives mais duquel on a retranché la batterie de variables muettes explique quant à lui 50,98 % de la variance du jugement. Autrement dit, d'après la modélisation MCO, les classes expliquent  $60,66 - 50,98 = 9,68$  % de la variance totale du jugement.

Les écarts entre les classes, indiqués par la batterie de variables muettes (i.e. les différences de constantes entre les classes) se révèlent assez importants puisqu'ils varient de -1,01 à 1,63, et que l'écart-type de la distribution égale 0,703. Rappelons que le jugement des enseignants varie sur une échelle de 0 à 10. C'est dire que l'écart « typique » entre deux classes, à score aux épreuves standardisées, profession des parents, retard scolaire et internalité donnés, est d'environ 0,7 point. Cela révèle des différences de sévérité/indulgence entre les classes qui ne sont pas minces.

Concernant l'étude de l'éventuelle hétérogénéité des relations entre scores aux épreuves standardisées et jugement, nous n'avons d'autre choix, en modélisation MCO, que de réaliser des régressions séparées pour chaque classe. Cela empêche bien entendu une modélisation globale de ce phénomène. De plus, attendu que le nombre d'élèves par classe est relativement réduit, il ne nous est pas possible d'intégrer d'autres variables que le score aux épreuves standardisées. On note évidemment d'emblée la restriction en termes de contrôle d'autres variables et les limites imposées à la clause « toutes choses égales par ailleurs ».

Les analyses révèlent des disparités importantes dans la relation entre le score aux épreuves standardisées et le jugement de l'enseignant. En termes de coefficients standardisés (en régression simple, le coefficient de régression standardisé est égal au coefficient de corrélation), les valeurs des coefficients varient de 0,12 à 0,92. Autrement dit, on dispose d'une classe pour laquelle le jugement de l'enseignant n'est quasiment pas corrélé avec les performances des élèves (0,12) tandis que, dans une autre classe, cette corrélation est quasiment parfaite (0,92). Toutefois, on relève que la classe pour laquelle la corrélation est faible ne comptait que peu d'élèves de CE2 (seulement 4). On se doute bien sûr que cette faible corrélation peut être due à ce très faible effectif. Le problème c'est que, le modèle MCO considérant cette relation comme fixe, il est incapable de prendre en compte la variance d'échantillonnage, très importante ici et sans doute responsable de cette très faible corrélation. Le modèle multiniveau permettra au contraire de débarrasser l'estimation de cette classe (et de toutes les autres) de sa part de variance d'échantillonnage.

Si on calcule la moyenne des coefficients de régression (non standardisés)<sup>7</sup> et qu'on calcule l'écart à la

moyenne pour chaque classe, on relève que l'écart-type de la distribution vaut 0,054, avec une valeur minimale de -0,075 et une valeur maximale de 0,115. Il s'agit là d'une autre manière de présenter des disparités importantes dans la relation entre le jugement et les scores aux épreuves standardisées des élèves. Autrement dit, l'hétérogénéité des relations entre ces deux variables apparaît assez forte.

### **Modélisation multiniveau**

Les données présentent une structure hiérarchisée puisque les élèves (niveau 1) sont groupés dans des classes (niveau 2). Une modélisation multiniveau s'impose donc, ce que nous nous proposons de faire maintenant. Il est d'usage de procéder par étapes dans la modélisation multiniveau. On estime alors, dans un premier temps, un modèle vide. Ce modèle est dit « vide » car il ne contient aucune variable explicative (cf. modèle 1, tableau 1). Il s'agit d'une simple décomposition de la variance en une part de variance inter-classes et une part de variance intra-classes. A ce stade, le coefficient de corrélation intra-classe *Rho* est très informant car il pose le rapport de la variance inter-classes sur la variance totale et donne par conséquent une estimation du degré de « ressemblance » des élèves au sein des classes. Dans notre cas,  $Rho = 0,478 / (3,678 + 0,478) = 0,115$ . Autrement dit, 11,5 % de la variance totale du jugement des enseignants réside entre les classes. Il y a bien des classes où les jugements sont en moyenne plus élevés que dans d'autres. Il faut signaler à cet égard que le modèle MCO estime quant à lui cette part de variance inter-classes à 15,4 %<sup>8</sup>. Il conduit par conséquent à une surestimation notable qui est due au fait que la variance d'échantillonnage n'est pas prise en compte.

Paramètres	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
Effets fixes					
Constante	6,652 (0,205)	8,991 (1,628)	11,399 (1,555)	10,443 (1,571)	10,232 (1,511)
Score (centré)		0,127 (0,007)	0,130 (0,013)	0,130 (0,013)	0,120 (0,013)
Score moyen par classe		-0,057 (0,039)	-0,113 (0,037)	-0,090 (0,037)	-0,092 (0,034)
Profession du père : (réf = cadre sup)					
Artisan/commerçant					-0,192 (0,299)
Intermédiaire					-0,351 (0,254)
Employé					-0,218 (0,235)
Ouvrier					-0,594 (0,223)
Autre					-0,283 (0,271)
Retard scolaire					-0,713 (0,206)
Internalité					0,071 (0,032)
Effets aléatoires					
Niveau 2 (classes) :					
Variance des constantes	0,478 (0,235)	0,429 (0,198)	0,369 (0,175)	0,356 (0,167)	0,279 (0,136)
Variance des pentes			0,002 (0,001)	0,002 (0,001)	0,002 (0,001)
Covariance constantes-pentes			-0,014 (0,010)		
Niveau 1 : variance intra-classes	3,678 (0,286)	1,847 (0,147)	1,694 (0,136)	1,662 (0,135)	1,560 (0,129)
-2 Log V	1395,37	1184,81	1165,39	1167,50	1150,87

Tableau 1 : Modèles multiniveaux expliquant le jugement des enseignants (entre parenthèses figurent les erreurs-types des coefficients). Logiciel utilisé : SAS. Méthode d'estimation : maximum de vraisemblance restreint

Le modèle 2 (cf. tableau 1) introduit le score aux épreuves standardisées sous deux formes : le score individuel et le score agrégé moyen par classe. On cerne donc par là l'effet intra-classe du score (l'effet produit par le fait d'être plus ou moins fort au sein de la classe) et l'effet de contexte (l'effet produit par le fait d'être dans une classe plus ou moins forte, à score individuel donné). L'effet est positif et significatif pour le score individuel<sup>9</sup> ; cela traduit, comme on pouvait s'y attendre, que les élèves sont d'autant mieux jugés qu'ils ont de meilleures performances au sein de leur classe. L'effet de la variable agrégée n'apparaît pas significatif dans le modèle 2. Nous avons cependant maintenu cette variable dans les analyses car, nous allons le voir, son impact va se révéler significatif dans les autres modèles, c'est-à-dire quand on introduira un effet aléatoire du score individuel (i.e. quand on modélisera la variance des pentes). Il faut de suite noter le signe négatif du coefficient de cette variable<sup>10</sup> qui, nous le confirmerons dans les modèles suivants, tend à indiquer que le jugement est d'autant plus sévère que la classe est plus forte (pour une discussion approfondie de ces effets, voir Bressoux & Pansu, 2003). On relève aussi que l'introduction de ces variables a modifié la partie aléatoire du modèle. La variance inter-classes (i.e. la variance des constantes  $\sigma_{i0}^2$ ) n'a guère évolué, passant de

0,478 à 0,429, soit 10,3 % de variance expliquée. En revanche, la variance intra-classes a été considérablement réduite, passant de 3,678 à 1,847. C'est dire que le score individuel<sup>11</sup> explique 49,8 % de la variance intra-classes du jugement. Autrement dit, les variations de jugement enregistrées au sein d'une classe sont bien largement déterminées par les performances des élèves les uns par rapport aux autres au sein de cette classe, même si une assez large part de variance demeure inexpliquée.

Les modèles 3 et 4 (cf. tableau 1) ont été réalisés de manière à spécifier au mieux la partie aléatoire du modèle. On se pose notamment la question d'une éventuelle hétérogénéité des relations entre le score des élèves et le jugement de l'enseignant, d'une classe à l'autre. C'est pourquoi on a ici spécifié un effet aléatoire de la relation entre le score des élèves et le jugement des enseignants, autorisant ainsi les pentes des droites de régression à varier aléatoirement d'une classe à l'autre. On cherche donc à estimer la variance des pentes  $\sigma_{i1}^2$ . Le modèle 3 estime aussi la covariance entre les pentes et les constantes  $\sigma_{i01}$ .

La première information qu'on peut tirer du modèle 3 est qu'il ajuste significativement mieux les données que le modèle 2. Il faut pour cela observer la valeur  $-2 \text{ Log } V$ , que l'on nomme la déviance du modèle. Quand on passe d'un modèle à un autre plus

complexe (les deux modèles doivent être hiérarchisés, c'est-à-dire que l'un doit être inclus dans l'autre, qui en est une extension ; il n'est pas possible qu'un paramètre soit présent dans le modèle le moins complexe et exclu du modèle plus complexe), la décroissance de la déviance d'un modèle à l'autre suit une loi du Chi2 qui a pour nombre de degrés de liberté le nombre de paramètres supplémentaires à estimer dans le modèle le plus complexe. Dans notre cas, il y a deux paramètres supplémentaires à estimer dans le modèle 3 par rapport au modèle 2 (i.e. une variance et une covariance à estimer en plus). La décroissance de la déviance vaut  $1184,81 - 1165,39 = 19,42$ , une valeur significative pour deux degrés de liberté. L'examen des erreurs-types semble toutefois indiquer que la variance des pentes est significative mais que la covariance ne l'est pas. Cependant, attendu que les coefficients aléatoires peuvent s'éloigner beaucoup de la normalité, il est conseillé de ne pas se fonder sur le rapport entre leur valeur et leur erreur-type pour en tester la significativité. La décroissance de la déviance est alors le seul test fiable<sup>12</sup>.

De ce point de vue, le modèle 4 (cf. tableau 1) présente deux intérêts : d'une part, il va permettre de s'assurer de la significativité ou non de ces deux coefficients et, d'autre part, il permet de spécifier un modèle plus parcimonieux puisqu'on y a ôté la covariance entre les constantes et les pentes. La décroissance de la déviance peut être calculée par rapport au modèle 3 mais aussi par rapport au modèle 2. Dans ce dernier cas, elle indique l'éventuelle amélioration de la qualité de l'ajustement aux données produite par l'introduction d'un effet aléatoire du score sur le jugement (et de ce seul paramètre). La décroissance de la déviance vaut  $1184,81 - 1167,50 = 17,31$ , une valeur significative pour un degré de liberté. Dans le cas du calcul de la décroissance de la déviance du modèle 3 par rapport au modèle 4 (là, on inverse le sens du calcul parce que le modèle 3 est plus complexe que le modèle 4), cela indique l'éventuelle amélioration de la qualité de l'ajustement aux données produite par l'introduction de la covariance entre les pentes et les constantes (et de ce seul paramètre). La décroissance de la déviance vaut  $1167,50 - 1165,39 = 2,11$ , une valeur non significative pour un degré de liberté. Autrement dit, les pentes varient bien d'une classe à l'autre, mais on n'enregistre pas de lien entre les pentes et les constantes. Les jugements d'une classe indulgente ne reposent pas davantage sur les scores individuels et inversement. Autrement

dit, si l'on prend la corrélation entre le score aux épreuves standardisées et le jugement comme un indicateur de « l'exactitude » du jugement de l'enseignant, on peut conclure que les enseignants les plus « exacts » ne sont pas plus indulgents que les autres et inversement.

Le modèle 4 a permis de spécifier au mieux la partie aléatoire du modèle : la variance des pentes est utile mais la covariance constantes-pentes est inutile. Le modèle 5 (cf. tableau 1) conserve par conséquent la même structure aléatoire que le modèle 4 mais il intègre l'ensemble des variables pertinentes dans l'analyse afin de parvenir à une modélisation « complète ». Examinons donc les effets fixes de ces variables. On peut voir que la profession des parents joue un rôle significatif. La catégorie des cadres supérieurs ayant été placée en référence, tous les coefficients s'entendent comme un écart avec cette catégorie. On relève d'emblée que tous les coefficients sont négatifs et qu'ils sont d'autant plus forts que la profession marque une position sociale plus faible. Ainsi, les enfants d'ouvriers sont, toutes choses égales par ailleurs, moins bien jugés que les enfants de cadres supérieurs, la différence s'établissant à un peu plus d'un demi-point. Rappelons-nous que le jugement s'établit sur une échelle allant de 0 à 10. Ce demi-point n'est donc pas négligeable. Les élèves en retard scolaire sont, toutes choses égales par ailleurs, moins bien jugés que ceux qui sont « à l'heure » avec une différence de 0,7 point. Enfin, on relève un effet positif de l'internalité : les élèves qui manifestent plus d'explications internes sont mieux jugés que ceux qui manifestent davantage d'explications externes. On marque ainsi l'effet positif de cette norme sociale générale et cela très tôt puisqu'il s'agit ici d'enfants de CE2, soit des enfants de 8-9 ans. Conformément à notre hypothèse, on voit que le jugement de l'enseignant sur la valeur scolaire des élèves, en l'occurrence il s'agit de leur valeur en mathématiques, est sensible à l'expression de l'internalité par les élèves.

Du fait de l'introduction de l'effet aléatoire du score aux épreuves standardisées des élèves sur le jugement des enseignants, la variance inter-classes du jugement n'est plus constante mais elle peut varier en fonction du score des élèves. En effet, la variance inter-classes est égale à

$$\text{var}(u_{0j} + u_{1j}X_{ij}) = \sigma_{u0}^2 + \sigma_{u1}^2 X_{ij}^2 + 2\sigma_{u01} X_{ij} \quad (5)$$

On relève que la variance inter-classes devient une fonction quadratique de  $X$  (à cause de la présence du terme  $X^2$ ). On modélise donc une variance complexe au niveau 2. On voit qu'on vient là de lever l'hypothèse d'homoscédasticité des modèles MCO, très restrictive, pour l'adapter à notre objet de recherche. En effet, la question de l'homoscédasticité devient alors un problème de recherche et même une question très intéressante à laquelle on peut maintenant répondre empiriquement. Pour quelle catégorie d'élèves, les variations de jugement des enseignants sont-elles les plus fortes ? La fonction est représentée dans la figure 1. Notons que, dans le cas du modèle 5, la covariance entre la constante et les pentes ayant été fixée à zéro, le terme  $2\sigma_{\nu 01} X_{ij}$  de l'équation (5) égale 0.

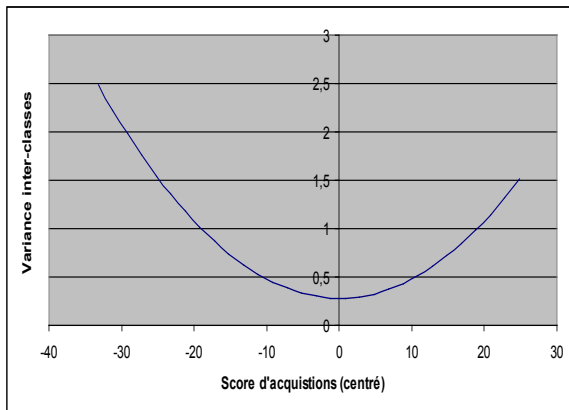


Figure 1 : Variance inter-classes du jugement des enseignants en fonction du score des élèves aux épreuves de mathématiques

On relève d'après la figure 1 que la variance inter-classes du jugement est élevée chez les élèves forts et les élèves faibles. C'est pour les élèves moyens qu'il y a le moins de variabilité en matière de sévérité/indulgence entre les classes. Les élèves les plus faibles sont quant à eux ceux qui subissent les plus grands écarts de sévérité/indulgence entre les classes. Pour eux, le jugement des enseignants varie de manière très notable d'une classe à l'autre. C'est aussi le cas, dans une moindre mesure, pour les élèves forts. Cette information apparaît tout à fait importante et elle est, à tout dire, inaccessible à la modélisation MCO. Tout d'abord parce que les régressions MCO séparées d'une classe à l'autre nuisent à une appréhension globale du phénomène de l'hétérogénéité des relations entre scores aux épreuves standardisées et jugement de l'enseignant (autrement dit, comme nous l'avons déjà signalé, la question de la plus ou

moins grande « exactitude » du jugement des enseignants), et ensuite parce que les estimations MCO surestiment fortement la variabilité des pentes de régression d'une classe à l'autre.

Nous illustrons ce dernier phénomène en représentant dans les figures 2 et 3 les droites de régression estimées pour chaque classe, en modélisation MCO (figure 2) et en modélisation multiniveau (figure 3).

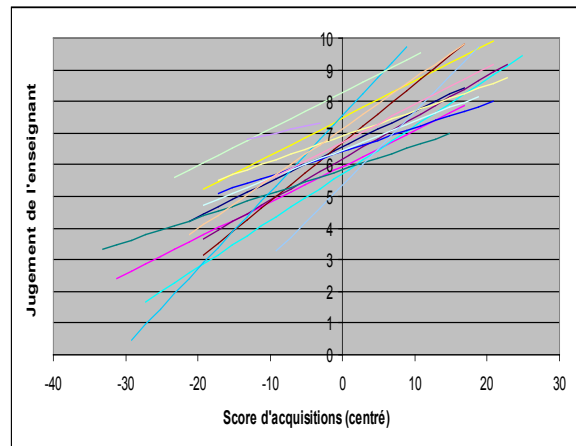


Figure 2 : Droites estimées par des régressions simples estimées séparément pour chaque classe

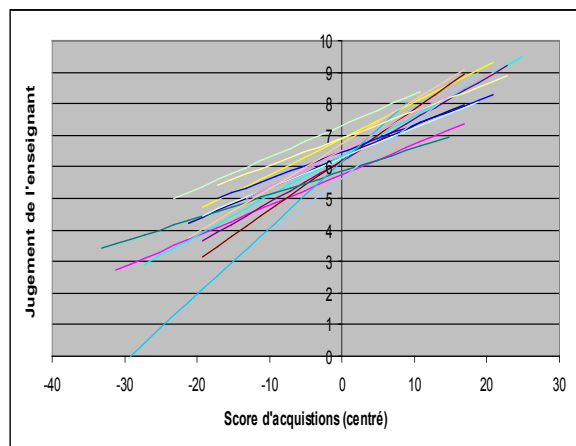


Figure 3 : Droites estimées par le modèle multiniveau (modèle 5) pour un enfant de cadre supérieur, « à l'heure » dans sa scolarité et dans une classe de niveau moyen

Pour prendre la pleine mesure de ce que représentent les figures 2 et 3, il convient de rappeler que la spécification du modèle multiniveau est différente de celle du modèle MCO. Dans les modèles MCO, seul le score aux épreuves standardisées est pris en compte comme variable explicative du fait du faible nombre d'élèves par classe (faible nombre de

degrés de liberté). En revanche, dans la modélisation multiniveau, l'ensemble des variables explicatives pertinentes ont été introduites dans le modèle. Cela conforte bien sûr la capacité à exprimer les résultats en termes de relations causales « toutes choses égales par ailleurs ». De plus, dans la modélisation multiniveau, c'est la variance des pentes qui est estimée (et non pas le coefficient de chaque pente qui est estimé séparément des autres) en tenant compte de la variance d'échantillonnage.

En conséquence, la figure 3 exprime un effet de « resserrement » autour de la relation moyenne. D'une part, les droites sont beaucoup moins « écartées » de la relation moyenne que dans la figure 2. Autrement dit, la variabilité des constantes est plus forte dans les modèles MCO que dans le modèle multiniveau. D'autre part, on voit que, dans la figure 3, les droites sont beaucoup moins « enchevêtrées » qu'elles ne le sont dans la figure 2. Cela traduit le fait que la variabilité des pentes est beaucoup plus forte dans les modèles MCO que dans le modèle multiniveau. Ces deux constats visuels (« resserrement » et réduction de « l'enchevêtrement ») sont en fait le produit d'un même phénomène d'écrasement (*shrinkage effect*) : du fait de la prise en compte de leur variance d'échantillonnage, les estimations des constantes et des pentes sont ramenées vers la moyenne générale des estimations. Elles sont d'autant plus fortement ramenées vers la moyenne générale que la variance d'échantillonnage est plus forte (i.e. que les effectifs par classe sont plus restreints). Reprenons le cas de la classe où l'effectif est très faible (seulement 4 élèves de CE2). Pour cette classe, l'estimation MCO de la pente égale 0,053. La pente est donc très faible par rapport à la moyenne des pentes de régression pour chaque classe (écart de -0,075 par rapport à la moyenne des pentes). En ce qui concerne le modèle multiniveau, la pente de cette classe est estimée à 0,108 (écart de -0,013 par rapport à la pente générale). Si l'on examine les constantes, l'estimation multiniveau de la constante égale 10,295 (pour une constante moyenne de 10,232, soit un écart de 0,063 point). Dans le modèle MCO qui inclut la batterie de variables muettes, l'estimation fournit un écart par rapport à la constante moyenne de 1,259 point. Les estimations pour cette classe ont donc été « ramenées » fortement vers la moyenne du fait d'une très grande variance d'échantillonnage dans les estimations MCO.

## Conclusion

Les modèles que nous avons estimés montrent que, conformément à nos hypothèses, le jugement est loin de n'être que le simple enregistrement des performances avérées des élèves. Bien qu'il se fonde en grande partie sur ces performances, il s'agit plutôt d'un processus complexe qui fait intervenir divers facteurs tels que les caractéristiques sociodémographiques des élèves, leur scolarité, le contexte de la classe et l'expression de normes sociales générales. Nous avons aussi montré une hétérogénéité des relations entre les diverses classes. Tous les enseignants ne prennent pas en compte de la même façon les performances des élèves pour former leur jugement (leurs jugements sont plus ou moins « exacts » si on prend le score aux évaluations standardisées comme un indicateur des performances réelles des élèves). De même, certaines catégories d'élèves (en particulier les élèves les plus faibles) voient le jugement porté sur eux varier de manière très notable selon la classe dans laquelle ils se trouvent.

De cette analyse, il apparaît que la modélisation multiniveau offre de nombreux avantages par rapport à la modélisation MCO. Non seulement les estimations MCO conduisent à des surestimations de taille d'effet ou de significativité, mais les modèles MCO sont limités dans leur capacité heuristique. Leurs fondements sont sapés par les questions de recherche qui portent sur les relations entre les individus et leur environnement (hypothèses d'homoscédasticité et d'indépendance). Comme nous l'avons vu, les modèles MCO sont impropres à accéder à une modélisation globale de l'hétérogénéité des relations entre les individus et leur environnement. Les modèles multiniveaux offrent quant à eux un cadre d'analyse propice à l'analyse de tels effets et, du fait qu'ils permettent de modéliser des variances complexes, étendent les questions que le chercheur est amené à se poser et à complexifier l'analyse de son objet de recherche.

Il faut signaler que les modèles multiniveaux connaissent de nombreuses extensions. Leur généralité et leur grande souplesse leur permettent de couvrir un champ d'application très vaste (cf. Bressoux, à paraître 2008). Il est en effet possible maintenant de modéliser des structures où les unités ne sont pas strictement hiérarchisées les unes par rapport aux autres. Par exemple, tous les élèves d'un même quartier ne vont pas dans la même école et une école n'accueille pas que des enfants d'un même quartier. De même, tous les élèves d'une classe donnée ne

proviennent pas d'une même classe de l'année précédente et tous les élèves d'une classe donnée ne se retrouvent pas l'année suivante dans la même classe. Il est possible de modéliser de telles structures avec des modèles multiniveaux aléatoires croisés. Il est de même possible de modéliser des courbes de croissance dans un cadre multiniveau. Lorsqu'un phénomène fait l'objet de mesures répétées pour chaque individu, on est intéressé à étudier l'évolution de ce phénomène dans le temps. Par exemple, si on mesure le poids des enfants à différents âges, on sera en mesure de modéliser le rythme de croissance de ce poids. Pour nous rapprocher du domaine scolaire, tester à plusieurs reprises les acquis des élèves, par exemple en lecture, permet d'étudier le rythme d'évolution de ces acquisitions dans le temps au fil des apprentissages (et non pas seulement le niveau d'acquisitions atteint). Ces phénomènes ne sont qu'en apparence éloignés de la modélisation multiniveau telle que nous l'avons présentée jusqu'alors, où les individus (élèves, employés, détenus, etc.) étaient groupés dans des environnements (écoles, entreprises, prisons, etc.). Dans le cas des modèles multiniveaux de croissance, les mesures (de poids, d'acquisition en lecture, etc.) sont considérées comme les unités d'analyse de niveau 1 et les individus (enfants,

élèves, etc.) sont considérés comme les unités d'analyse de niveau 2. Autrement dit, les individus constituent les « environnements » des mesures réalisées, comme si ces mesures étaient « groupées » dans les individus. On se retrouve donc bien dans un cadre d'analyse multiniveau.

Les modèles multiniveaux offrent par conséquent un cadre d'analyse particulièrement riche et souple, avec des capacités heuristiques très intéressantes pour les chercheurs en sciences sociales en général et les chercheurs en éducation en particulier, d'autant que ces modélisations peuvent être combinées entre elles en fonction des structures analysées (modélisation multiniveau de croissance avec modélisation multiniveau aléatoire croisée par exemple). Même s'il convient bien sûr de ne pas accorder de primauté aux méthodes par rapport à la construction des objets d'étude, il reste que ces techniques permettent d'explorer les objets au-delà de ce que peuvent faire les techniques MCO, qu'elles conduisent à se poser de nouvelles questions de recherche et que, en cela, elles participent aussi à la construction des objets de recherche et, par conséquent, aux connaissances que nous élaborons.

## NOTES

1. Le verbe « concevoir » a bien ici un sens propre et fort car il serait tout à fait erroné d'attribuer a priori un statut ontologique quelconque aux différents niveaux impliqués dans l'analyse. Ils ne s'imposent pas d'eux-mêmes aux chercheurs comme s'ils étaient un « déjà-là » dans la réalité. La définition des niveaux doit relever de considérations théoriques en fonction de l'objet de recherche. Tel niveau jugé pertinent pour analyser un phénomène donné peut se révéler tout à fait non pertinent dans l'étude de tel autre phénomène.
2. De par le fait de l'existence de corrélations intra-groupes non nulles.
3. Entre l'agrégation simple et la désagrégation simple, quelques solutions intermédiaires existent, mais qui ne sont pas pleinement satisfaisantes et qui ne suppriment jamais cette dialectique agrégation *versus* désagrégation, ramenant toujours le problème à la question du « bon » niveau d'analyse (voir à ce sujet Bressoux, à paraître 2008; Bressoux, Coustère & Leroy-Audouin, 1997).
4. On est en général contraint dans le cadre des MCO, lorsqu'on veut analyser l'hétérogénéité des effets, de réaliser des régressions séparées pour chaque environnement ce qui pose 1) la question de la globalité de l'analyse et 2) la question du traitement que l'on fait ensuite subir aux estimations obtenues. Si on les récupère pour les expliquer par des variables de niveaux 2, ces *estimations* sont alors traitées comme des *observations* (supposées sans erreur de mesure donc), ce qu'elles ne sont pas.
5. Pour être tout à fait précis, les estimations sont itératives et le modèle se sert d'une première estimation à un niveau pour estimer l'autre niveau. Cette estimation permet elle-même de réestimer le premier niveau, ce qui va conduire à une réestimation de l'autre niveau et ainsi de suite jusqu'à ce qu'il y ait convergence, c'est-à-dire jusqu'à ce que les estimations ne varient quasiment plus (en fonction d'une limite fixée).
6. Une variable muette est une indicatrice qui signifie l'appartenance (code 1) ou la non-appartenance (code 0) à une catégorie déterminée. Par exemple, l'indicatrice de la classe 1 (nommons-la *CLA1*) vaudrait 1 pour tous les élèves qui appartiennent à la classe 1 et 0 pour tous les autres élèves. De même, *CLA2* vaudrait 1 pour tous les élèves qui appartiennent à la classe 2 et 0 pour tous les autres élèves. Etc.
7. Ce qui ne revient pas à trouver la relation moyenne sur l'ensemble de l'échantillon des élèves puisqu'on ne tient pas compte ici des variations d'effectifs entre les classes.
8. Il s'agit en ce cas d'un modèle MCO qui introduit la batterie de variables muettes représentant les classes mais sans inclure aucune autre variable explicative.
9. Le rapport du coefficient sur son erreur-type doit être au moins égal à 1,96 pour être jugé significatif au seuil  $\alpha = 0,05$  (i.e. pour avoir moins de 5 chances sur 100 de se tromper en affirmant que les variables sont liées).
10. Il est à noter que le modèle MCO sous-estime beaucoup l'erreur-type de cette variable agrégée (elle est estimée à 0,020) ce qui traduit un risque considérablement accru d'erreur de première espèce. Dans ce modèle MCO, on a ôté la batterie de variables muettes indicatrices des classes et on a introduit, en plus des variables explicatives individuelles, la variable agrégée du score moyen par classe. Du point de vue des variables explicatives introduites, ce modèle est donc tout à fait identique au modèle 5 (multiniveau) présenté dans le tableau 1. Pour autant, on le voit, les estimations des erreurs-types des coefficients diffèrent de manière notable.
11. Si les deux variables sont potentiellement explicatives de la variance inter-classes, seule la variable individuelle peut expliquer la variance intra-classes car il n'y a pas de variabilité intra-classes sur la variable agrégée.
12. Pour être tout à fait précis, la décroissance de la déviance est aussi le seul test exact pour les effets fixes. Le rapport du coefficient fixe d'une variable sur son erreur-type donne cependant en général une valeur fiable très proche de celle donnée par la décroissance de la déviance. De plus, la méthode d'estimation utilisée ici étant le maximum de vraisemblance restreint, il n'est pas possible de tester la signification des effets fixes en étudiant la décroissance de la variance.



## BIBLIOGRAPHIE

- Aitkin, M., & Longford, N. T. (1986). Statistical modelling in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, 149, 1-43.
- Beauvois, J.-L. (1984). *La psychologie quotidienne*. Paris : PUF
- Bressoux, P. (à paraître, 2008). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles : De Boeck.
- Bressoux, P., Coustère, P., & Leroy-Audouin, C. (1997). Les modèles multiniveaux dans l'analyse écologique : le cas de la recherche en éducation. *Revue Française de Sociologie*, 38(1), 67-96.
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris : PUF
- Erbring, L., & Young, A. A. (1979). **Individuals and social structure : contextual effects as endogenous feedback**. *Sociological Methods and Research*, 7, 396-430.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.
- Mason, W. M., Wong, G. M., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 72-103). San Francisco : Jossey-Bass.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York : Wiley.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis : an introduction to basic and advanced multilevel modeling*. London : Sage.
- Willms, D. J. (1985). The balance thesis : contextual effects of ability on pupils' O-grade examination results in Scotland. *Oxford Review of Education*, 11, 33-41.