International Journal of Economics & Strategic Management of Business Process (ESMB)
Vol.22 pp.103-118

Apport des modèles multiniveauxdansl'analyse de l'éducation au Maroc

Salah Eddine TAHA

LARESSGDL : LAboratoire de Recherche en Economie Sociale et Solidaire; Gouvernance et Développement

FPS, Université Cadi Ayyad, Maroc

Téléphone : +212 (0) 665 942 514

Adresse électronique taha.salah.eddine@gmail.com

Aomar IBOURK

LARESSGDL : LAboratoire de Recherche en Economie Sociale et Solidaire; Gouvernance et Développement

FSJES, Université Cadi Ayyad, Maroc

Adresse électronique: aomaribourk@gmail.com

Résumé: *Les modèles hiérarchiques ou multiniveaux ont été développés afin de remédier aux difficultés spécifiques rencontrées lors du traitement des données dites structurées ou emboitées en différents niveaux. Dès lors, ce travail tire au clair l'apport de l'analyse multiniveaux dans la recherche en éducation, et ce, en présentant les modèles hiérarchiques et les solutions apportées pour faire face à l'ensemble des contraintes imposées par les modèles linéaires classiques. Via l'analyse multiniveaux, ce travail a également comme objectif la décortication des déterminants du rendement cognitif des jeunes élèves. Les traits majeurs de notre étude qui a porté sur plus de 780 élèves révèlent entre autres que dans un contexte marocain, le niveau d'éducation des parents, la taille des groupes et la qualité de l'environnement scolaire sont parmi les facteurs clés dans l'apprentissage cognitif des élèves.*

***Mots-clés** : Modélisation économétrique (C500), éducation (I240), éducation (I210), Apprentissage (I210), performances (I250).*

ISSN:1737-9288

1. Introduction

L'analyse multiniveaux correspond à une panoplie de techniques statistiques dédiées au traitement des données dites emboitées (en termes de niveaux d'analyse). En effet, c'est une analyse qui fait intervenir des modèles statistiques permettant d'identifier les différentes sources de variance d'un coté, et d'identifier les effets des variables explicatives propres à chaque niveau d'analyse d'un autre (Ibourk et Taha, 2017 ; 2018 ; 2021). Les inférences statistiques qui sous-tendent l'analyse multiniveaux portent conséquemment sur plusieurs ensembles d'entités d'analyse correspondant aux différentes sources de variation (Snijders 2002).Les modèles multiniveaux reposent sur un postulat de base stipulant que les relations entre les variables individuelles peuvent différer selon les contextes. D'ailleurs, c'est la raison pour laquelle le modèle sera composé de variables identifiant chaque niveau d'analyse (niveau individuel et niveaux contextuels).

Du fait que les systèmes éducatifs sont connus par leur organisation hiérarchisée, les modèles multiniveaux sont souvent un outil privilégié par les chercheurs en éducation. Au Maroc, les autorités en éduction ont adopté récemment une nouvelle réforme en éducation (vision stratégique de l'éducation nationale 2015-2030). Dans cette vision, le système éducatif doit faire preuve d'implication dans la stratégie qui converge vers l'économie et la société du savoir. Dès lors, ce travail présente une initiation pratique à la régression multiniveaux, et ce, à l'aide d'une enquête ad-hoc menée auprès de 45 établissements scolaires.Egalement, via l'analyse multiniveaux, nous allons décortiquer les facteurs explicatifs des performances en Mathématiques chez les jeunes élèves marocains.

Ce travail est structuré en deux parties, la première partie aborde les modèles multiniveaux, et ce, via la présentation de leurs origines, intérêt et formulation. La deuxième partie s'attarde sur l'illustrationempirique de l'analyse multiniveaux, et ce, afin de mettre en lumière les facteurs susceptibles de contribuer potentiellement à l'acquisition des habilités en Mathématiques.

2. Présentation des modèles Hiérarchiques

2.1. Origine des Modèlesmultiniveaux

Les modèles multiniveaux connus également sous plusieurs noms, à savoir : les modèles linéaires hiérarchiques, les modèles mixtes, les modèles à coefficients aléatoires et les modèles à variance composée. Lesdits modèlesontapparusdurant la fin des années soixante-dix, et ce, via les acquis de l'analyse de contexte (Boyd et Inversen 1979 ;Burstein 1980 ;Hammond 1973) et via les avancées des modèles mixtes (Searle, Casella et McCulloch 1992)qui sont cités par Nnang (2013). Les modèles multiniveaux ont encore rayonné avec l'apparition d'une deuxième vague de travaux notamment, Aitkin et Longford (1986) ;Glodstein (1986, 2003) ;Entwisle, Mason et Hermalin (1986) etRaudenbush et Bryk (1986).Bien que le survol de la littérature permet de constater que le développement des modèles multiniveaux est assez récent, Bressoux et ses collaborateurs (1997) ont cité plusieurs travaux anciens dans lesquels les intuitions de base des modèles multiniveaux ont été abordé (Boudon 1963 ; Robinson 1950).

Le principe méthodologique de ces modèles repose sur une synthèse de trois approches complémentaires de probabilités, de la variance et des régressions. Il s'agit en fait d'une régression dont la partie indéterminée du modèle est décomposée en plusieurs variables latentes, et ce, conformément aux hypothèses sur l'hétérogénéité inobservée des variables appartenant aux différents niveaux d'analyse (Delaunay 2002).

2.2. Intérêt de l'analyse multiniveaux

L'analyse multiniveaux est utilisée souvent dans les sciences humaines et sociales, en réalité cette analyse représente une généralisation de la régression linéaire. Entre autres, elle permet d'identifier dans le même modèle des effets aléatoires de deux dimensions (inter et intra), et ce, hors mis les effets aléatoires adossés au terme d'erreur. Autrement dit, les modèles multiniveaux sont beaucoup plus adaptés pour tester l'impact des variables exogènes appartenant aux différents niveaux sur une même variable dépendante (Taha et Ibourk, 2018).

Dans notre partie empirique, nous envisageons identifier les différentes variables susceptibles d'expliquer les performances en Mathématiquesdes élèves.Sachant que cette dernière (petite enfance) est assujettie à plusieurs protagonistes appartenant aux différents niveaux, nous avons pu collecter au cours de notre enquête des données décrivant les caractéristiques propres à l'élève, à safamille, aux écoles et aux éducatrices. Ces données sont donc présentées suivant une hiérarchie à plusieurs niveaux.

Pour pouvoir détecter les facteurs de performances scolaires, il serait intéressant de déceler les effets inhérents aux élèves et à leur environnement familial (niveau I) et à l'environnement scolaire des écoles (niveau II). Dans cette situation, si nous faisons appel à la régression classique par les moindres carrées ordinaires (MCO), cela présenterait plusieurs limites et peut constituer des sources de problèmes statistiques. D'ailleurs :

- Si nous mélangeons les deux niveaux par une utilisation des mêmes données pour tous les élèves concernés, nous négligeons conséquemment l'éventualité stipulant que les observations à l'intérieur d'une même école ou de la même classe (groupe) ne sont forcément pas indépendantes. Les paramètres estimés par les MCO sont convergents, mais leur estimation n'est pas efficiente et la variance estimée ne correspond pas à la variance réelle (Michaelowa, 2000).
- Les modèles de régression MCO supposent classiquement une indépendance des erreurs. Autrement dit, la connaissance de l'erreur liée à un élève ne peut pas permettre de prédire l'erreur liée à un autre élève. Or, cela contredit les

ISSN:1737-9288

effets supposés de l'environnement sur les élèves. En effet, dans l'analyse multiniveaux et contrairement aux modèles MCO, les environnements vont « modifier » les élèves de telle manière qu'une non-indépendance va s'établir entre eux du point de vue de la variable réponse étudiée(Bressoux, 2008) qui est dans notre cas le score cognitif en Mathématiques des élèves.

- Les modèles hiérarchiques multiniveaux traitent l'hypothèse d'homocédasticité par une hypothèse plus faible selon laquelle la variance des résidus peut varier selon une fonction linéaire ou non des variables explicatives (Snijders et Bosker, 1999).

Dans ce sens, il est question également d'identifier la démarche à adopter, si les résultats de ces modèles sont opposés à ceux estimés par la méthode des moindres carrés ordinaires (MCO). Les réponses avancées généralement à ces préoccupations portent essentiellement sur le coefficient de corrélation intra-classe (intra-école) que nous allons expliquer dans la partie dédiée à la formulation des modèles multi niveaux.

Cela dit, il serait judicieux d'abandonner l'usage des modèles multiniveaux dans trois cas précis, à savoir :

- Premièrement, lorsque le coefficient de corrélation intra-classe (intra-école) est très faible en raison de différences non significatives entre les groupes.

- Deuxièmement, si nous somme en présence d'un plan de sondage compliqué ou déformé et ne présentant pas clairement la structure hiérarchique.

- Troisièmement, lorsque l'information contenue dans les niveaux hiérarchiques supérieurs est jugée peu pertinente et qu'on s'intéresse surtout à des effets fixes du modèle. Dans ce cas, l'estimation par la méthode des moindres carrés ordinaires (MCO) pourrait être préférable à une régression multiniveaux.

3. Formulation des modèles multiniveaux

La modélisation multiniveaux s'effectue en plusieurs étapes. La première étape primordiale consiste à vérifier la pertinence d'adopter une structure hiérarchique dans la modélisation des effets. Pour cette raison, nous allons procéder à l'estimation d'un modèle dit « inconditionnel ». Comme son nom l'indique, le modèle « inconditionnel » n'impose aucune variable explicative, conséquemment c'est un modèle qui ne contient aucune variable explicative. Son estimation a principalement comme objectif la décomposition de la variance totale de la variable dépendante en une part de variance appelée « intra-classe » et une part de variance appelée « interclasse ».

La part de variance intra-classe (part de variance du niveau relatif à l'élève et son environnement familial « niveau 1 ») permet de savoir dans quelle mesure le modèle explique les différences intra-classes (différences qui surgissent entre les élèves d'une même classe). Quant à la part de variance interclasse (part de variance du niveau relatif à l'école et à son environnement « niveau 2 ») reflète quant à elle dans quelle mesure le modèle proposé explique les différences interclasses (différences entre les différentes écoles considérées).

Notons Y la variable dépendante, mesurant les acquisitions scolaires (score finale en maths, français, en arabe, et sensori-motricité), X les variables explicatives relatives au niveau 1 (niveau élève et famille) et Z les variables explicatives relatives au niveau 2 (niveau école et qualité de l'environnement scolaire).

3.1. Analyse de la variance : Modèle initial

Comme déjà expliqué c'est un modèle sans variable explicative, permettant d'estimer les différentes composantes de la variance en évaluant la part expliquée par chaque niveau de l'analyse. Il est aussi appelé modèle vide.

Le modèle vide à deux niveaux peut être représenté par deux équations sans aucune variable explicative, une à chaque niveau (Bryk et Raudenbush, 1992).

La spécification du premier niveau (élève et famille) est la suivante :

$$Y_{ij} = \beta_{0j} + e_{ij} \qquad (1)$$

ISSN:1737-9288

Cette équation prédit le score qu'obtiendra l'élève i de l'école j au test cognitif.

Avec :

Y_{ij} : Score de l'élève i de l'école j

β_{0j} : Score moyen de l'école j

e_{ij} : Terme d'erreur mesurant l'écart entre le score final moyen de l'école j et celui de son élève i, il est supposé suivre une loi normale de moyenne nulle et de variance constante σ_1^2.

La spécification du deuxième niveau (école) est la suivante :

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (2)$$

Le modèle du deuxième niveau utilise le paramètre d'interception du niveau 1 (niveau élève et son environnement familial) comme variable à expliquer. Cela veut tout simplement dire que le score moyen de l'école est un résultat qui varie de façon aléatoire autour d'une moyenne générale de tous les élèves (score final moyen).

Avec :

γ_{00} : est le score final moyen de tous les élèves.

u_{0j} : Terme d'erreur mesurant l'écart entre ce score final moyen et celui de l'école j, il est supposé suivre une loi normale de moyenne nulle et de variance constante σ_2^2.

Si on remplace le score moyen de l'école par sa nouvelle expression :

De (1) et (2) on a : $Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$ (3)

Comme indiqué auparavant, l'estimation de ce modèle permet d'évaluer si l'analyse multiniveaux est nécessaire au traitement des données collectées dans notre enquête. Cela consiste à examiner si la variable dépendante est affectée par un effet – classe (effet école) ou pas. Pour cela, il faut calculer le coefficient de corrélation intra-classe (intra-école) rhô, ce dernier s'obtient en rapportant la variance interclasse (inter-école) par la variance totale.

On a supposé que les erreurs sont indépendantes les unes des autres, autrement dit :

$$V_{Y_{ij}} = V_{\gamma_{00}+u_{0j}+e_{ij}} = V_{u_{0j}} + V_{e_{ij}} = \sigma_1^2 + \sigma_2^2$$

Ainsi, on peut obtenir une formule de décomposition de la variance totale qui permet d'évaluer la part de celle-ci imputable aux différents niveaux d'analyse :

Proportion de la variance expliquée par niveau	
Niveau 1	**Niveau 2** : coefficient de corrélation intra-écoles
$\boldsymbol{\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}}$	$\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$

Ainsi, un coefficient de corrélation intra-classe (intra-école) s'il est égal à 0 % signifie que l'hypothèse d'indépendance des observations n'est pas violée et donc on peut utiliser un modèle linéaire classique. Par contre, si le coefficient intra-classe (intra-école) est supérieur à 0 %, l'hypothèse d'indépendance des observations n'est pas respectée, par conséquent, on peut conclure qu'une part de la variance dépendante se situerait entre les écoles. Ne pas tenir compte de cette corrélation intra-classe (intra-écoles) créera des problèmes, et ce, du fait que les estimations deviendront moins fiables.

Une autre façon d'interpréter le coefficient de corrélation intra-classe (intra-école) serait de le considérer comme étant le degré de similarité des scores des élèves au sein d'un même groupe ou d'une même école (Bressoux, 2008). Par exemple, une valeur de 10 % signifie que deux élèves d'une même école ont environ 10 % de chances de posséder les mêmes valeurs des variables qui les décrivent que s'ils avaient été sélectionnés parfaitement au hasard dans l'ensemble des élèves constituant la population. Le coefficient de corrélation intra-classe peut également être interprété comme étant le pourcentage de la variabilité des acquis qui peut être attribuée à la classe (école).

Plus le coefficient de corrélation est important, plus l'hypothèse d'indépendance entre les élèves est remise en question, et donc, plus les méthodes

ISSN:1737-9288

statistiques traditionnelles basées sur cette hypothèse risquent de donner des résultats biaisés.

Une fois le modèle appelé « vide » est estimé, le coefficient de corrélation intra-classe (intra-école) est calculé et une fois on constate sur la base du coefficient de corrélation que le score final des élèves varie dans la même classe (école) et également d'une classe (école) à l'autre, nous allons poursuivre notre étude en faisant appel à des modèles multiniveaux plus complexes qui nous aideront à expliquer les parts de variance intra-classe et interclasse du score émanant du test cognitif.

3.2. Modèle Individuel : Estimation du modèle à constante aléatoire

3.2.1. Introduction des variables du niveau I

L'étape suivante de l'analyse multiniveaux consiste à intégrer au modèle de composition de la variance (modèle vide) des variables caractérisant l'élève et son environnement familial (niveau 1). Pour simplifier la présentation, nous allons considérer que les autres variables individuelles comme étant indépendantes du deuxième niveau (classe, école et environnement).

Autrement dit, à ce stade nous nous intéresseront uniquement aux relations entre la variable Y représentant le score cognitif total et X les variables explicatives se trouvant au niveau 1.

L'équation du niveau 1 pour chaque unité statistique ij est la suivante :

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij} \qquad (4)$$

β_{0j} : est la valeur moyenne des performances du groupe j (écoles).

β_1 : sont les effets des variables explicatives. Ces effets varient d'un individu à l'autre.

e_{ij} : Terme d'erreur de la modélisation sur l'individu ij.

En effet, on peut dire que l'équation du Niveau I reflète entre autres la décomposition de la variance interindividuelle (intra-écoles). Or, l'équation du Niveau II que nous allons voir renvoie quant à elle à la décomposition de la variance intergroupes (inter-écoles).

3.2.2. Au Niveau II

Nous n'allons pas introduire les variables décrivant le Niveau II dans le modèle, et ce, du fait que nous cherchons à contrôler la variance déjà existante et par la même occasion gérer les problèmes de non indépendance des observations. Autrement dit, les effets des variables explicatives du niveau 1 ainsi que le terme constant, tous les deux deviennent des variables à expliquer par celles du niveau supérieur :

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} & (5) \\ \beta_1 = \gamma_{10} & (6) \end{cases}$$

Où

γ_{00} : Constante moyenne pour l'ensemble des écoles

γ_{10} : Pente moyenne pour l'ensemble des écoles

u_{0j} : Ecart de chaque école à la constante (effet aléatoire associé à la constante)

e_{0j} : Erreur associée à chaque individu.

Les deux termes d'erreur sont supposés être distribués aléatoirement avec une moyenne nulle. Par ailleurs sur un même niveau, on suppose qu'ils sont hétéroscédastiques et corrélés entre eux. Or, leur indépendance est présumée d'un niveau à l'autre.

En remplaçant les coefficients β_{0j} et β_1 par leurs valeurs dans l'équation du Niveau I (élève et sa famille) nous obtiendront alors le modèle suivant :

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + (u_{0j} + e_{ij}) \qquad (7)$$

La partie fixe est représentée par les *gammas* et la partie aléatoire est représentée par les deux termes d'erreur $(u_{0j} + e_{ij})$.

3.3. Estimation du modèle avec les variables du Niveau II

Dans cette étape, on autorise les variables du Niveau II parce qu'elles sont susceptibles d'avoir une incidence sur le phénomène sans toutefois jouer sur la forme de la relation mise en évidence au Niveau I. Autrement dit, seule l'intersection β_{0j} varie entre les classes et est expliquée par une variable Z_j de Niveau II (école et environnement). La forme des différents modèles est la suivante :

Au Niveau I

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij} \qquad (8)$$

Au Niveau II

Le modèle de Niveau II permet de gérer le problème de la non-indépendance des observations ainsi que l'impact des variables de contrôle de Niveau II.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \qquad (9)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \qquad (10)$$

Pour passer de la forme structurelle ci-dessus à trois équations à la forme réduite à une seule équation qui sert à l'estimation, il suffit de substituer aux coefficients β_{0j} et β_1 de l'équation de Niveau I leurs valeurs respectives, telles que spécifiées dans les équations de Niveau II. L'équation multiniveaux devient alors :

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{10} X_{ij} + \gamma_{11} Z_j X_{ij} + \left(u_{0j} + u_{1j} X_{ij} + e_{ij}\right) \qquad (11)$$

Dans le modèle ci-dessus, ce sont les paramètres γ_{00}, γ_{01}, γ_{10}, γ_{11}, $\sigma^2_{u_{0j}}$, $\sigma^2_{e_{ij}}$ et $\sigma^2_{u_{1j}}$que nous allons tenter d'estimer. Ce modèle multiniveaux peut être décomposé en deux parties, une partie fixe qui reflète les coefficients de pente des variables explicatives et une partie aléatoire composée des différentes variances. D'ailleurs, c'est cette partie aléatoire qui distingue le modèle multiniveaux du modèle de régression par les moindres carrés ordinaires (MCO). En effet, dans le modèle multiniveaux, la partie aléatoire apparaît à chacun des niveaux de la structure hiérarchisée des données, ce qui fait référence aux niveaux 1 et 2 (élève, famille, école et environnement). Tandis que dans le modèle des moindres carrés ordinaires (MCO), l'effet aléatoire ne surgit qu'au niveau des élèves et leur environnement familial, ce qui signifie qu'on a fait abstraction aux effets de l'école et son environnement.

A noter également, dans la formulation de ce modèle, nous avons varié et la constante et les coefficients des variables explicatives (modèle à constantes et à coefficients aléatoires). En d'autres termes, dans la formulation ci-dessus, nous avons rajouté un effet aléatoire au niveau de la pente des variables explicatives libérées. Toutefois, dans d'autres formulations des modèles multiniveaux, seule la constante du modèle qui est variée (modèle à constante aléatoire).D'ailleurs, dans notre simulation empirique visant l'identification des déterminants de performances en Mathématique chez l'élève, nous avons opté pour un modèle à constante aléatoire, et ce, du fait que l'effet aléatoire des pentes des variables explicatives est non significatif.

Avant de présenter les résultats de l'analyse multiniveaux, il est important de rappeler les hypothèses fondamentales sur lesquelles se base la technique de régression multiniveaux pour une application correcte (Bressoux 2008) :

Hypothèse 1 : Additivité qui prévoit la linéarité de la relation entre a variable à expliquer et la variable explicative.

Hypothèse 2 : La variable à expliquer est censée être continue, non bornée mesurée sans erreur.

Hypothèse 3 : Absence de colinéarité entre les variables explicatives.

Hypothèse 4 : A chaque niveau, les erreurs suivent une loi normale : $e_{ij} \rightarrow N\left(0, \sigma^2_{e_{0j}}\right), u_{0j} \rightarrow N(0, \sigma^2_{u_{0j}})$

Hypothèse 5 : Les erreurs entres les niveaux sont indépendantes : $Cov(\sigma^2_{e_{ij}}, \sigma^2_{u_{0j}}) = 0$

Hypothèse 6 : Les perturbations des observations de mêmes niveaux sont indépendantes : $Cov\ (e_{ij}, e_{i''j''}) = 0\ si\ i \neq i''\ ou\ j \neq j''$.

Hypothèse 7 : Les variables explicatives à tous les niveaux sont observées sans perturbation, donc indépendantes des erreurs aléatoires (hypothèse d'exogénéité) :

$$Cov\ (X_{qij}, e_{ij}) = Cov\ (Z_j, u_{0j}) = Cov\ (X_{ij}, u_{0j}) = Cov\ (Z_j, \varepsilon_{1j}) = 0$$

Hypothèse 8 : Les variables explicatives de mêmes niveau ne sont pas corrélées entre elles : $Cov\ (X_q, X_{q''}) = Cov\ (Z_l, Z_{l''}) = 0\ si\ q \neq q''\ ; l \neq l''$

4. Etude empirique

Via la modélisation multiniveaux, nous allons vérifier dans ce qui suit l'impact des variables caractérisant l'environnement familial et l'environnement de l'école sur les performances mathématiques des élèves en âge scolaire.

4.1. Présentation des données

Les données mobilisées dans cette analyse portent exactement sur 784 élèves répartis sur 45 écoles. Ces données proviennent essentiellement des informations soulevées des formulaires et questionnaires administrés auprès des établissements et des familles des élèves en scolaire et des informations extraites des tests en mathématiques que les élèves en scolaire ont dû passer afin d'évaluer leurs connaissances cognitives en fin du cycle scolaire.

Les données issues de cette étude, bien qu'elles soient riches en informations, présentent quelques difficultés liées à la présence de valeurs manquantes qui peuvent fausser les estimations. Afin de dépasser cette difficulté, nous avons privilégié la technique d'imputation multiple. En effet, Rubin (1987) a introduit pour la première fois cette technique qui correspond davantage à la nature de nos données. La méthode d'imputation multiple peut être décrite en trois étapes. La première consiste à créer des groupes de valeurs plausibles pour les données manquantes (notre cas m=10). Chacun de ces groupes de valeurs sert à remplir les valeurs manquantes et créer ainsi m bases de données. Ensuite, les bases complètes sont analysées avec les méthodes utilisées traditionnellement (différentes méthodes économétriques d'estimation). Enfin, les résultats obtenus des analyses réalisées à partir des m bases complètes sont combinées selon une procédure spécifique dans le but d'obtenir des estimateurs non biaisés.

4.2. Estimationsmultiniveaux

Pour l'estimation des modèles multiniveaux, il existe de nombreux logiciels spécialisés tels que HLM, et MLwiN[1], mais des logiciels généralistes comme SAS, SPSS ou STATA peuvent être néanmoins utilisés. Dans le cadre de ce travail, les estimations ont été réalisées avec STATA. Nous avons adopté les procédures d'estimation « xtmixed » et « xtreg », et ce, avec la méthode du maximum de vraisemblance préconisée dans la littérature économétrique. D'autres techniques d'estimation peuvent aussi être utilisées. Il s'agit principalement de la méthode du maximum de vraisemblance (ML), de la méthode des moindres carrés généralisés (GLS), de la méthode des moindres carrés généralisés restrictifs (RGLS), de la méthode du « bootstrap », et des méthodes bayésiennes.

La spécification du modèle économétrique et le descriptif des variables mobilisées dans la modélisation du score en Mathématiques sont indiqués au niveau des Annexes I et II.L'analyse des acquisitions en mathématiques passe par un processus de trois estimations dans lesquelles nous avons essayé d'alimenter les modèles par les variables pertinentes permettant à la fois un bon ajustement et une meilleure parcimonie.

[1] C'est l'un des instruments les plus complets puisqu'il permet l'analyse des variables dépendantes continues, dichotomiques, multinomiales ou ordonnée, et il offre la possibilité d'analyser les séries chronologiques, les modèles multivariés tout comme les modèles de survie et des événements historiques. A cela s'ajoute le fait qu'il n'impose pas de restriction quant au nombre de niveaux présents dans les données.

VARIABLES	Modèle (01) Score en Mathématique	Modèle (02) Score en Mathématique	Modèle (03) Score en Mathématique
Variables Niveau I (Élève et sa famille) **:**			
Age de l'élève = 1, Plus de 5 ans		4.200584 (6.123774)	4.13224 (6.17538)
Genre de l'élève = 1, Garçon		-12.54136* (6.361093)	-12.74445** (6.390782)
Taille de fratrie = 1, 3 Élèves et plus		-10.89505 (10.05823)	-10.77518 (10.15022)
Langue parlée à la maison = 1, Français		12.29904 (11.52859)	11.85707 11.71364
Niveau d'instruction de la mère = 1, Primaire		18.94774 (11.74559)	19.08792 (11.70694)
Niveau d'instruction de la mère = 2, Collège		23.42099** (9.943821)	23.78254** (9.951355)
Niveau d'instruction de la mère = 3, Secondaire		37.85244*** (10.40365)	37.08918*** (10.61711)
Niveau d'instruction de la mère = 4, Post secondaire		49.17254*** (15.93355)	46.99842*** (16.65121)
Structure de la famille=1, famille monoparentale		3.038179 (7.565443)	2.923835 (7.755361)
Nombre de manuels en possession=1, Entre 1-5 manuels		24.06384* (12.82935)	24.01769* (12.82635)
Nombre de manuels en possession=2, Entre 6-15 manuels		21.01857 (15.8883)	20.38581 (15.99428)
Nombre de manuels en possession=3, Plus de 15 manuels		17.49784 (14.21558)	17.32335 (14.17029)
Activité et devoirs = 1, Présence du suivi parental		54.38668*** 18.01124	54.25973*** (17.92155)
Variables Niveau II (Ecole et Environnement) :			
Qualité environnementale scolaire= 1, Minimale			-10.03061 (20.20158)
Qualité environnementale scolaire = 2, Bonne			-1.5458 (23.05964)
Nombre d'élèves par groupe = 1, 15-20 élèves			-36.62986** (14.31832)

ISSN:1737-9288

Nombre d'élèves par groupe = 2, 21-30 élèves			-38.80948** (16.86014)
Nombre d'élèves par groupe = 3, +30 élèves			-76.73992*** (25.86851)
Constante	508.02095*** (8.42756)	418.3411 *** (23.33312)	468.9313*** (25.43623)
Effets aléatoires :			
Niveau II :			
Variabilité inter-écoles	52.43572*** (7.041328)	48.29904*** (7.310164)	45.06656*** (6.56925)
Niveau I :			
Variance intra-écoles	77.21279*** (6.013351)	75.000821*** (6.037506)	74.97155*** (6.03844)
Rho (CCI)	40.45 %	39.15 %	37.38 %
Observations	784	784	784
Nombre d'écoles	45	45	45

Entre parenthèses figurent les erreurs-types robustes des coefficients
*** Significativité à 1% (p<0.01) ; ** Significativité à 5% (p<0.05) ; * Significativité à 10% (p<0.1)

D'après le tableau récapitulant les estimations effectuées du score en mathématiques, le modèle vide (Modèle 01) reflétant la décomposition de la variance permet de dégager le coefficient de corrélation intra-écoles (Rho). De ce coefficient, nous déduisons la proportion de la variance inter-écoles occupée dans la variance totale et conséquemment nous pouvons avoir une information sur le degré de ressemblance des scores en mathématiques des élèvesscolairisés au sein de leurs écoles. Dans le cas dudit modèle,

$$Rho = \frac{52.43572}{77.21279 + 52.43572} = 40.45$$

Cela revient à dire tout simplement que 40.45 % de la variance totale du score en mathématique des élèvesscolarisés réside entre les écoles. Autrement dit, la variance inter-écoles du score cognitif total occupe 40.45 % dans la variance totale. Bien évidemment.

Dans le deuxième modèle du tableau, nous avons introduit les variables explicatives caractérisant l'élève et sa famille. De ce, nous essayons de cerner les éventuels effets de certaines spécificités proprement liées à l'élève et son environnement familial notamment. Ceci dit, nous constatons que le genre de l'élève est une variable dont l'impact sur le score mathématique est négatif et statistiquement significatif. Conséquemment, la significativité du signe négatif de l'effet du genre révèle qu'en moyenne, le passage d'une fille à un garçon engendrera une diminution du score cognitif total. En d'autres termes, les filles ont plus de chances d'acquérir un niveau de compétences mathématique supérieur à celui des garçons. Compte tenu de l'effet du suivi des devoirs et activités par les parents sur ledit score, nous remarquons qu'il est positif et statistiquement significatif. Cela revient à dire qu'en moyenne lorsque les parents assurent un suivi régulier des devoirs et activités des élèvesscolairisés, ces derniers parviennent à réalisé un score en mathématique supérieur à celui des élèves dont les parents n'assurent pas de suivi des devoirs et activités. Quant au nombre de manuels en possession, le modèle à révélé que cette variable peut impacter significativement le score en mathématiques. L'examen du niveau d'instruction des parents a révélé également que l'impact du niveau d'instruction de la maman est positivement significatif, et ce, contrairement à l'impact non significatif du niveau d'instruction des pères qu'on a préféré de le retirer du modèle. En ce qui concerne l'âge, la taille de fratrie et la structure de la famille de l'élèvescolairisé, nous pouvons constater que leurs effets ne se sont pas révélés significatifs dans le deuxième modèle. Ceci dit, nous avons gardé ces variables tout de même dans notre modèle par ce que nous verrons l'ajustement de leurs coefficients au fur et à mesure que nous introduisons d'autres variables explicatives.

A souligner également qu'au niveau du deuxième modèle, les deux composantes de la variance totale à savoir la variance inter et intra-école ont changé différemment. En effet, nous remarquons que la variance intra-école a été revue à la baisse en passant de 77 à 75, ce qui revient à dire que l'ajout des variables explicatives de l'élève et sa famille permet d'expliquer une partie de la variance intra-écoles. Quant à la variance inter-écoles, elle a également baissé en passant de 52 à 48. Ceci dit, le changement de la composante aléatoire constaté en passant du modèle vide (modèle (01)) au modèle (02) reste un peu timide.

Au niveau du troisième modèle et afin de cerner davantage l'effet intra-écoles et inter-écoles sur les acquisitions mathématiques des élèves, nous avons introduit d'autres variables caractérisant l'école et son environnement. En effet, dans le modèle (03) nous avons rajouté des variables caractérisant l'école et la qualité de son environnement. Cela dit, nous constatons dans le modèle (03) que le rajout des variables du niveau II (école et son environnement) a amélioré davantage l'ajustement de l'effet des variables niveau I (élève et sa famille) déjà introduites, et ce, tout en améliorant leur signification statistique. Par rapport au modèle vide, nous remarquons également un changement au niveau des deux composantes de la variance, dans ce modèle où nous avons introduit les variables explicatives du deuxième niveau, nous assistons à une diminution relativement significative en termes de la variance inter-école, et ce, en passant de 52 (modèle vide) à 45 (modèle 03). Concernant les variables nouvellement introduites au modèle (03), nous remarquons que seul l'effet de la taille de classe (nombre d'élèves par groupe) qui s'est révélé très significatif, le résultat de cette variable permet de constater qu'en cycle scolaire, plus la taille de classe est élevée moins bons sont les résultats des scores en mathématiques, et plus la taille est réduite plus les performances en mathématiques s'améliorent.

4.3. Pertinence et validité du modèle à constante aléatoire

Avant de procéder aux interprétations des coefficients et paramètres estimés du modèle final (modèle 03), nous avons utilisé certains tests permettant de juger la pertinence du choix du modèle à constante aléatoire et sa validité. Dans ce sens, les résultats du test d'hausman, du ratio de vraisemblance et l'analyse du pouvoir explicatif (Pseudo R^2) du modèle final sont regroupés au niveau de l'Annexe I.

4.4. Interprétations des résultats

Il ressort de l'estimation finale du score en mathématiques (modèle 03) que plusieurs variables des deux niveaux (élève et école) expliquent très significativement le score en mathématiques des élèvesscolairisés.

Le genre de l'élèvescolairisé joue un rôle décisif en matière d'amélioration du score en mathématiques, d'ailleurs l'estimation finale révèle que le passage d'une fille à un garçon réduit significativement le score en mathématiques de 13 points, toutes choses étant égales par ailleurs. Cela revient à dire que les esquisses des disparités d'apprentissage entre les filles et les garçons surgissent bien avant l'âge du primaire.

Les élèves dont le niveau d'instruction de la mère évolue au primaire, collège, secondaire et postsecondaire voient leurs habilités en mathématiques augmentent respectivement de 20 points ; 24 points, 37 points et 47 points par rapport aux élèves dont la maman est sans instruction, et ce toutes choses étant égales par ailleurs. Contrairement aux niveaux d'instruction de la maman qui ont un impact significatif sur les acquisitions en mathématiques, les niveaux d'instruction du père n'ont pas d'incidence statistiquement significative. Cela revient à dire que l'éducation des mamans des élèves en scolaire est très décisive en matière d'apprentissage des habilités en mathématiques, et ce, pour tous les niveaux d'instruction.

Quant au suivi des devoirs et activités des élèves en cycle scolaire par les parents, les résultats obtenus montrent que cette variable est statistiquement significative et impacte positivement les acquis en mathématiques, d'ailleurs les élèves dont les parents assurent un suivi des devoirs et activités ont en moyenne un score en math supérieur de 55 points par rapport aux élèves dont les parents n'assurent pas un suivi d'activité et devoirs. Autrement dit, le suivi des devoirs et activité par les parents contribue significativement à l'amélioration des habilités de calcul des élèves en cycle scolaire.

En ce qui concerne le nombre de manuels en possession, l'estimation finale révèle que cette variable peut impacter significativement l'acquisition des habilités en mathématiques. En effet, en moyenne les élèves ayant de un à cinq manuels revoient leur score en mathématiques augmente significativement (au seuil de 10 %) d'un peu près 25 points comparativement aux élèves qui n'ont aucun manuel en leur possession. Egalement, il faut signaler qu'au fur et à mesure que le nombre de manuels augmente l'impact est toujours positif sur les acquis en math, mais au-delà de cinq manuels cet impact sur le score en math est non significatif.

Les résultats ont révélé également que le nombre frères et sœurs n'as pas d'impact statistiquement significatif sur les habilités acquises en mathématiques. En revanche la négativité du coefficient de cette variable laisse comprendre que l'élève disposant d'une taille de fratrie dépassant 2 membres voit son score en mathématiques diminué de 12 points (ceteris paribus). Le modèle final montre également que l'âge, la langue parlée à la maison (Dialecte, Amazigh ou Français) et la structure de la famille (famille traditionnelle/monoparentale) des élèves en âge précoce n'ont pas d'impact statistiquement significatif sur l'acquisition des habilités en mathématiques.

Le deuxième niveau qui s'intéresse à l'école et son environnement, seule la variable relative à la taille de classe qui a révélé un effet statistiquement très significatif sur le score en mathématiques. D'ailleurs, plus la taille de classe est élevée moins bons sont les résultats du score en mathématiques, et plus la taille est réduite plus le score en mathématiques s'améliorent. Pour illustrer cela, il suffit de comparer les écarts des scores en mathématiques constatés entre l'effectif de moins de 15 élèves d'un cotés (effectif de référence) et les autres effectifs (entre 15-20 ; entre 21-30 ; plus de 30) d'un autre. En effet, les élèves dont l'effectif de groupe est inférieur à 15 élèves par groupe améliorent significativement leur score en mathématiques de 37 points par rapport à ceux dont l'effectif est compris entre 15-20. Les élèves appartenant à l'effectif de référence (moins de 15 élèves) améliorent encore plus leur score en mathématiques de 39 points par rapport aux élèves dont la taille du groupe est comprise entre 21-30 élèves par groupe. Egalement, les élèves constituant le taille de référence améliorent plus encore leurs score en mathématiques de 77 points comparativement aux élèves dont l'effectif dépasse 30 élèves par groupe. Et ce, toutes choses étant égales par ailleurs. Cela revient à conclure que plus la taille de groupe augmente moins les élèves en cycle scolaire apprennent les habilités en mathématiques.

En ce qui concerne la qualité de l'environnement, le type de l'établissementscolaire (public/privé), la formation des éducatrices et la zone d'appartenance géographique, les résultats ont révélé que leurs impacts ne sont pas statiquement significatifs et nous les avons retiré de la régression vue leur contribution très timide en matière d'explication du score en math (à l'exception de la qualité de l'environnement que nous avons gardé dans la régression finale). De ce, nous pouvons conclure qu'en cycle scolaire, l'apprentissage des habilités en mathématiques (calcul, comparaison de grandeurs, etc.) est moins sensible à ces spécificités décrivant l'environnement de l'école et les éducatrices.

5. Conclusion

Ce travail avait comme finalité la mise en valeur de la contribution des modèles multiniveaux dans la recherche en éducation scolaire. En effet, il a permis de présenter d'une part les origines et les spécifications économétriques des modèles multiniveaux. D'autre part, ce travail a tiré au clair l'intérêt des modèles multiniveauxface aux données emboitées. En effet, les régressions multiniveaux sont privilégiées par rapport aux régressions ordinaires, et ce, du fait qu'ils présentent une solution à l'ensemble des contraintes imposées par les modèles linéaires classiques (MCO) notamment, la non indépendance des erreurs, l'hétérogénéité des relations, l'homoscédasticité et les problèmes d'agrégation ou de désagrégation des données hiérarchisées (BRESSOUX, 2008).

Ce travail via la modélisation multiniveaux avait également comme objectif la décortication et l'analyse des déterminants des performances en Mathématiquesdes élèves se trouvant dans la province de Marrakech. Les résultats des différentes modélisations élaborées dans cette recherche révèlent que dans un contexte marocain, l'apprentissage des habilités en Mathématiques durant le scolaire ne dépend pas uniquement des caractéristiques relatives aux élèves et à leurs familles mais également à un ensemble de caractéristiques décrivant l'environnement de l'établissement scolaire.

En effet, concernant l'éducation des parents, nos résultats révèlent que l'apprentissage cognitif des élèves est relativement indifférent vis-à-vis du niveau d'instruction du père. Or, il est fort de constater qu'au fur et à mesure que le niveau d'instruction de la mère évolue, l'élève améliore de plus en plus leurs performances cognitives. Cela revient à conclure que l'éducation des mères des élèves en scolaire est plus décisive en matière d'apprentissage cognitif surtout à l'âge précoce, et ce, pour tous les niveaux d'instruction. De même, le genre de l'élève s'est également révélé décisif en termes d'acquisition cognitive, cela permet de constater que les disparités de genre (garçon/fille) surgissent bien avant le cycle primaire. Par rapport au statut socioprofessionnel du père et la structure de la famille (traditionnelle/monoparentale) de l'élèvescolarisé, les résultats obtenus démontrent que ces deux variables n'ont pas d'influence statistiquementsignificative sur les performances acquises durant le scolaire.

Autre facteur concourent également à l'explication des performances mathématiques acquises en âge scolaire, notamment la taille de classe des élèves.Force est de constater qu'en cycle scolaire, plus la taille de classe est élevée moins bons sont les résultats du score cognitif en Mathématiques, et plus la taille est réduite plus les performances en Mathématiques s'améliorent. Quant à la variable caractérisant la qualité environnementale, les résultats révèlent que l'appartenance à un environnement de qualité élevée ou réduite n'a pas d'impact statistiquement significatif sur l'acquisitiondes habilités en Mathématiques, cela peut conduire à conclure que l'apprentissage des compétences en Mathématiques est moins sensible au score de la qualité de l'environnement scolaire.

BIBLIOGRAPHIE

AITKIN, M. LONGFORD, N. (1986). Statistical modelling issues in school effectiveness studies. Journal of the Royal Statistical Society. Series A (General), 1-43.

BOUDON, R. (1963). Propriétés individuelles et propriétés collectives: un problème d'analyse écologique. Revue française de sociologie, 275-299.

BOYD, L. H., & IVERSEN, G. R. (1979), Contextual analysis: Concepts and statistical techniques. Wadsworth Pub Co.

BRESSOUX, P. (1993), Les performances des écoles et des classes : le cas des acquisitions en lecture, Les dossiers éducation et formation, 30.

BRESSOUX, P. (2008), Modélisation statistique appliquée aux sciences sociales, Bruxelles : De Boeck.

BRESSOUX, P. et PANSU, P. (2003), Quand les enseignants jugent leurs élèves, Paris: PUF.

BURSTEIN, L. (1980). Chapter 4: The Analysis of Multilevel Data in Educational Research and Evaluation. Review of research in education, 8(1), 158-233.

BRYK, A.S. et RAUDENBUSH, S.W. (1992), Hierarchical linear models for social and behavioral research: Applications and data analysis methods.

DELAUNAY, J. G. (2002). Chapitre 6. Absence et présence dans la médiation pédagogique ou comment faire circuler les signes de la présence. In Pratiquer les TICE (pp. 103-113). De Boeck Supérieur.

DUNCAN, G, J. NICHD. (2003), Modeling the impacts of child care quality on children's preschool cognitive development, Child Development, Early Child Care Research Network 74(5), 1454-1475.

ENTWISLE, B. MASON, W. M. HERMALIN, A. I. (1986). The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength. Demography, 23(2), 199-216.

GOLDSTEIN, H. (2003), Multilevel Statistical Models. Arnold, London.

HAMMOND, C. B., BORCHERT, L. G., TYREY, L., CREASMAN, W. T., et PARKER, R. T. (1973). Treatment of metastatic trophoblastic disease: good and poor prognosis. American journal of obstetrics and gynecology, 115(4), 451-457.

HARMS, T. CLIFFORD, R. M. et CRYER, D, (1998), Early childhood environment scale-revised edition.

IBOURK, A., TAHA, S. E. (2017). « Les Déterminants de la Qualité dans l'Education de la Petite Enfance ». Critique économique n°36. Eté-automne.

IBOURK, A., TAHA, S. E. (2018). « Key Factors of Cognitive Performance in Moroccan Preschool: Evidence from Random Slope Model ». International Business Research. Vol. 11, N° 11. ISSN 1913-9004. E-ISSN 1913-9012.

IBOURK, A. & TAHA, S. (2021). De la différenciation des contextes à la création des écoles fragiles : sources des inégalités d'apprentissage de la petite enfance marocaine. *Spécificités*, 15, 98-129. https://doi.org/10.3917/spec.015.0098

IVERSON, G. R. (1991), Contextual Analysis Sage, Newbury Park, CA.

MICHAELOWA, K. (2000). Dépenses d'éducation, qualité de l'éducation et pauvreté.

NNANG, E. (2013), L'APPROCHE PAR COMPETENCES DANS LES PAYS EN DEVELOPPEMENT EFFETS DES REFORMES CURRICULAIRES EN AFRIQ, Science de l'éducation, Bourgogne : Université de Bourgogne, 17 Décembre, Np 333.

RABE-EESKETH, S. et Skrondal, A. Multilevel and Longitudinal Modeling Using Stata, Volume I, Continuous Responses, Edition n°3.

RAUDENBUSH, S. BRYK, A. S. (1986). A hierarchical model for studying school effects. Sociology of education, 1-17.

Robinson, F. P. (1950). Principles and procedures in student counseling.

RUBIN, D. B. (1987), Multiple imputation for nonresponse in surveys, Wiley Series in probability and statistics.

SEARL, S. R., CASELLA, G. et MCCULLOCH, C. E. (1992). Variance components. Wiley, New York.

SNIJDERS, T. (2002), Markov chain Monte Carlo estimation of exponential random graph models, Journal of Social Structure, 3 (2), 1-40.

SNIJDERS, T. BOSKER, R. (1999), Multilevel analysis : An introduction to basic and advanced multilevel modeling, Washington : SAGE.

TAHA, S.E. et IBOURK, A. (2018). L'usage du numérique dans les écoles comme déterminant de l'apprentissage préscolaire. Inetrnational Journal of Economics & Strategic Management of Business Process (ESMB), Vol.14, pp.45-53.

International Journal of Economics & Strategic Management of Business Process (ESMB)
Vol.22 pp.103-118

ANNEXES I

Test d'hausman : Modèle à effets aléatoires vs Modèle à effets fixes

```
. hausman Fixed Random
```

	Coefficients (b) Fixed	Coefficients (B) Random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Age	9.535447	7.291879	2.243568	1.181859
SX	-11.03669	-11.51358	.4768835	.7158002
NFS	-7.967925	-9.731857	1.763932	.7621703
LPM	6.174472	8.559573	-2.385101	3.264027
NIM				
1	15.84303	15.80861	.0344124	1.147969
2	19.82749	19.71351	.1139746	1.409339
3	31.55621	31.6718	-.1155919	1.591011
4	42.73084	40.09947	2.631373	2.350396
SF	1.264087	2.757247	-1.49316	1.298544
NMP				
1	29.79634	29.82289	-.0265521	2.880497
2	21.64754	22.65867	-1.011138	3.1713
3	24.98945	24.26447	.7249801	3.573354
SDAP	62.57154	60.68134	1.8902	3.283049

```
                    b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

              chi2(13) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                       =       17.76
             Prob>chi2 =      0.1667
```

Nous constatons que le test d'Hausman est non significatif et donc nous retenons l'hypothèse nulle permettant l'adoption du modèle à effets aléatoires. Conséquemment, le résultat de ce test corrobore la pertinence du modèle à effets aléatoires en général et le modèle à constante aléatoire en particulier.

Test du ratio de vraisemblance

```
. lrtest Modèle03

Likelihood-ratio test
(Assumption: Modèle02 nested in Modèle03)
```

D'après ce test, au niveau de 5 % et un degré de liberté de 5 et une p-value = 0.0447. Nous pouvons largement rejeter l'hypothèse nulle stipulant que l'ajout des variables décrivant l'école et son environnement n'apportera pas d'explication au score cognitif des mathématiques. Cela dit, on peut conclure que la décroissance de la déviance est significative (p-value = 0.0447) entre les deux modèles comparés. Conséquemment nous

ISSN:1737-9288

retiendrons le modèle le contenant plus de variables explicatives (modèle 03).

Le pouvoir explicatif du modèle final

D'après le tableau récapitulant l'ensemble des estimations du score en mathématiques, la variance totale estimée du modèle vide (modèle 01) est de :

$$\widehat{\varphi}_0 + \widehat{\theta}_0 = 52.43572^2 + 77.21279^2 = 8711.32$$

Alors que la variance totale estimée dans le modèle final (modèle 06) est de :

$$\widehat{\varphi}_F + \widehat{\theta}_F = 45.06656^2 + 74.97155^2 = 7649.77$$

D'où :

$$\text{Pseudo R}^2 = \frac{\left[\left(\widehat{\varphi}_0 + \widehat{\theta}_0\right) - \left(\widehat{\varphi}_F + \widehat{\theta}_F\right)\right]}{\widehat{\varphi}_0 + \widehat{\theta}_0} = \frac{8711.32 - 7649.77}{8711.32} = 0.1219744$$

Ce qui signifie que les variables explicatives niveau I et II qui ont été ajouté au modèle final du score cognitif total (modèle 03) ont pu expliquer un peu près 13 % de la variance totale.

ANNEXES II

Spécification du modèle économétrique relatif au score en mathématiques

Spécification des variables du niveau I (élève et sa famille) pour une école fixée :

$$SS_{ScoreenMath} = \alpha_0 + \alpha_1 Age + \alpha_2 SX + \alpha_3 NFS + \alpha_4 Nim2 + \alpha_5 Nim3 + \alpha_6 Nim4 + \alpha_7 Nim5 + \alpha_8 SDAP + \alpha_9 LPM + \alpha_{10} NMP + \varepsilon_1$$

Spécification des variables du niveau II (école et son environnement) :

$$\alpha_0 = \beta_0 + \beta_1 sqte2 + \beta_2 sqte3 + \beta_3 Neg2 + \beta_4 Neg3 + \beta_5 Neg4 + \varepsilon_{0j}$$

Nous avons supposé que tous les coefficients ($\alpha_1, \alpha_2, \alpha_3,$ $etc.$) varient non aléatoirement, contrairement à α_0(Constante aléatoire). Conséquemment, nous déduisons le modèle final à constantes aléatoire comme suit :

$$SS_{ScoreenMath} = \beta_0 + \alpha_1 Age + \alpha_2 SX + \alpha_3 NFS + \alpha_4 Nim2 + \alpha_5 Nim3 + \alpha_6 Nim4 + \alpha_7 Nim5 + \alpha_8 SDAP + \alpha_9 LPM + \alpha_{10} NMP + \beta_1 sqte2 + \beta_2 sqte3 + \beta_3 Neg2 + \beta_4 Neg3 + \beta_5 Neg4 + \varepsilon_1 + \varepsilon_{0j}$$