

ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
AFFILIÉE À L'UNIVERSITÉ DE MONTRÉAL

Analyse de la Mobilité Intergénérationnelle en Afrique du Sud

par

Alain-Désiré NIMUBONA

Sciences de la gestion

*Mémoire présenté en vue de l'obtention
du grade de maître ès sciences
(M.Sc.)*

Décembre 2002

©Alain-Désiré NIMUBONA, 2002

Sommaire

Traitant les questions d'équité et de justice sociale, l'analyse de la mobilité intergénérationnelle a pour objet la mesure du degré de transmission du statut socioéconomique entre générations. Le présent mémoire étudie cette problématique en Afrique du Sud. L'Afrique du Sud a connu un passé de ségrégation raciale institutionnalisée en 1948 à travers la politique d'*Apartheid*. Cette ségrégation raciale, caractérisée par la domination des Blancs, a généré de fortes inégalités socioéconomiques parmi les Sud-Africains. La fin de l'*Apartheid*, officiellement décrétée en 1994, augure une lueur d'espoir d'une plus grande égalité des opportunités. Mais, ces inégalités héritées du passé vont-elles disparaître naturellement à travers le temps ou vont-elles se répercuter de façon systématique sur les générations futures ? Quels sont les autres déterminants de la mobilité intergénérationnelle des Noirs que le niveau d'éducation de leurs parents ? Telles sont nos deux principales questions de recherche. Un autre apport de notre recherche se situe sur le plan méthodologie. Il consiste en l'introduction de la méthode des pseudo-panels dans les techniques d'analyse de la mobilité intergénérationnelle. Conçue pour pallier l'absence des données longitudinales, cette approche cumule d'autres atouts susceptibles d'améliorer l'estimation de la mobilité intergénérationnelle. Dans notre étude, le coefficient de la mobilité intergénérationnelle, correspondant au coefficient de corrélation entre le nombre d'années d'éducation d'un enfant et celui du chef de sa famille, est estimé à un niveau variant entre 0.20 et 0.29, en fonction de la procédure d'estimation utilisée. Cet intervalle est inclus dans celui allant de 0.14 à 0.45, établi par Mulligan (1999) en se référant aux principaux résultats de la littérature sur la mobilité intergénérationnelle du nombre d'années d'éducation.

Sommaire.....	1
Table des matières.....	2
Liste des tableaux.....	3
Remerciements.....	5
1. Introduction.....	6
2. Un bref survol de la littérature.....	10
2.1. Les modèles univariés.....	10
2.2. Les modèles multivariés.....	12
2.3. Approches non-linéaires.....	14
3. Les données.....	16
3.1. La construction des couples enfants - chef de famille.....	17
3.2. Analyse statistique.....	19
4. Le modèle.....	20
4.1. La spécification du modèle.....	20
4.2. Une analyse basée sur la méthodologie des pseudo-panels.....	23
4.3. La construction des cellules et variables de pseudo-panel.....	27
5. Analyse des résultats.....	29
5.1. Données individuelles.....	29
5.2. Données de pseudo-panel.....	35
6. Conclusion.....	40
Annexe A : Estimation utilisant des données de pseudo-panel....	42
Annexe B : Tableaux.....	48
Références bibliographiques.....	67

Liste des tableaux

Tableau 1 : Résumé des résultats des études de la mobilité intergénérationnelle.....	49
Tableau 2 : Intensité raciale par province des personnes âgées d'au moins 15 ans.....	50
Tableau 3 : Caractéristiques des enfants retenus dans les différents échantillons.....	51
Tableau 4 : Transformation du niveau d'éducation en nombres d'années d'éducation.....	52
Tableau 5 : Répartition des enfants retenus en fonction de leur zone de résidence....	53
Tableau 6 : Statistiques descriptives du nombre d'années d'éducation des enfants par cellule.....	54
Tableau 7 : Statistiques descriptives des variables de pseudo-panel.....	58
Tableau 8 : Résultats des estimations utilisant les données individuelles.....	59
Tableau 9 : Estimés des modèles logit et probit ordonnés.....	62
Tableau 10 : Effets marginaux des modèles logit et probit ordonnés.....	64
Tableau 11 : Résultats des estimations utilisant les données de pseudo-panel.....	65

À la mémoire de mon père.

Remerciements

Pendant tout le temps qu'a duré cette recherche, l'aide et les conseils que nous avons reçus ont atteint une ampleur peu commune. Nos efforts auraient été vains s'ils n'avaient pas bénéficié de l'appui de nombreuses personnalités qu'il nous est très agréable d'évoquer ici.

Notre conscience est principalement pleine de gratitude que nous avons l'honneur d'exprimer à l'endroit du Professeur Désiré Vencatachellum, directeur de ce mémoire. Après nous avoir proposé ce thème de recherche, il n'a rien ménagé afin que nos échanges soient empreints de son expérience et de sa rigueur scientifique.

Ce travail doit aussi beaucoup aux quatre semaines que nous avons passées à *Rhodes University*, en Afrique du Sud. Nous témoignons notre reconnaissance au personnel de cette université pour son hospitalité et son appui scientifique. Nos remerciements s'adressent tout particulièrement à Madame Brenda Erasmus, assistante administrative au département d'économie de *Rhodes University*, ainsi qu'à Madame Mweru Mwingi et Monsieur Lungelo Sali, étudiants au département d'éducation de ladite université.

Mme Corinne Berneman, professeure à HEC Montréal, nous a prêté une main forte lors de la traduction de certains textes descriptifs de nos bases de données, rédigés en Afrikaans. Nous lui en sommes très reconnaissant.

À Messieurs Pierre-Carl Michaud et Mohamed Jabir, nous disons également merci pour leur appui informatique.

Nous remercions aussi notre famille, pour les encouragements qu'elle n'a cessé de nous envoyer, de très loin.

Le financement de nos études a été assuré par l'Agence Canadienne de Développement International (ACDI), dans le cadre de son programme de bourses d'excellence de la francophonie.

1. Introduction

Plusieurs arguments économiques sont souvent formulés pour justifier l'existence de fortes inégalités dans la distribution des revenus entre les citoyens d'un pays. D'après un d'entre eux, les agents économiques prennent des décisions économiques différentes, ce qui explique les disparités de leur bien-être. Mais, cette argumentation ne serait juste que si toutes les personnes avaient des chances raisonnables d'affronter l'épreuve de la vie avec succès sous réserve de consentir les sacrifices requis. Dans la réalité, certaines personnes bénéficient, dès leur naissance, de certains avantages socioéconomiques, non en raison de leur mérite mais parce qu'elles disposent d'autres atouts liés à leur appartenance raciale ou ethnique. Parallèlement, d'autres personnes font face à des contraintes de mobilité. Dans ces conditions, le succès ou l'échec dans la vie économique et sociale peut devenir une question d'hérédité. Cette situation justifie la problématique de la mobilité intergénérationnelle qui se définit comme le degré de transmission du statut socioéconomique entre générations. Les inégalités socioéconomiques, en soi, constituent un sérieux problème¹, mais encore faut-il s'interroger sur la manière dont ces inégalités se transmettent de génération en génération. Le fait qu'un même statut socioéconomique persiste dans une même famille, sur plusieurs générations consécutives, a de fortes implications sur la distribution intertemporelle des revenus et sur l'équité socioéconomique.

Ainsi, notre mémoire se propose d'étudier la transmission intergénérationnelle du statut socioéconomique en Afrique du Sud. Nous y étudions le processus de transmission du niveau d'éducation des parents chefs de famille à leurs enfants, dans la population noire. Les particularités de l'Afrique du Sud sont très adaptées à cette problématique de la mobilité intergénérationnelle. En effet, une politique de ségrégation raciale, communément appelée *Apartheid*, y a été appliquée depuis 1948² et a généré de fortes inégalités socioéconomiques, parmi les Sud-Africains. La fin de

¹ Les carences du marché financier et l'inefficience économique qu'elles engendrent sont des caractéristiques fondamentales des sociétés fortement inégales (Galor et Zeira, 1993). Les inégalités occasionnent aussi des coûts politiques. Elles favorisent l'instabilité politique, en augmentant le niveau de criminalité et d'insécurité.

cette politique, officiellement décrétée en 1994, augure une lueur d'espoir d'une plus grande égalité des opportunités. Cependant, malgré le fait que le pays ait beaucoup progressé vers une égalité politique interraciale, les inégalités économiques persistent. En 1993, les revenus des Blancs, des Métis et des Indiens étaient estimés respectivement à 9.1, 2.2 et 4.0 fois plus grand par rapport aux revenus des Noirs (Michaud et Vencatachellum, 2003).

De nombreuses recherches économiques ont analysé les inégalités interraciales en Afrique du Sud. Paradoxalement, l'intérêt des études sur la mobilité intergénérationnelle y est récent (Thomas 1996, Burns 2000, 2001). Par ailleurs, outre la base incomplète de données qu'elle utilise, l'approche suivie par Burns (2000) souffre de plusieurs problèmes. En particulier, dans le souci d'assurer la continuité de la variable dépendante, elle effectue une transformation logarithmique du nombre d'années d'éducation des individus qui engendre un biais de sélection additionnel dans les estimations. De plus, Burns (2001) reconnaît l'impact du niveau du capital humain moyen par zone de résidence dans le processus de la mobilité intergénérationnelle, mais elle n'en tient pas compte dans son modèle. Pourtant, les transformations politiques post-*Apartheid* constituent une opportunité à ne pas rater, non seulement pour réduire les inégalités socioéconomiques structurellement héritées de l'*Apartheid*, mais aussi pour neutraliser leur transmission aux générations futures. Ainsi, dans cette recherche, nous posons les deux questions suivantes : (i) Les inégalités héritées du passé vont-elles disparaître naturellement à travers le temps ou bien vont-elles se répercuter de façon systématique sur les générations suivantes ?

(ii) Quels sont les autres déterminants de la mobilité intergénérationnelle des Noirs que le niveau d'éducation de leurs parents?

Dans le tableau (1)³, nous résumons les résultats des grands travaux effectués sur la mobilité intergénérationnelle dans différents pays, en fonction de l'indicateur du statut socioéconomique retenu et des caractéristiques des données utilisées. Les études antérieures de la mobilité intergénérationnelle retiennent soit les revenus, soit la

² Même avant cette année d'institutionnalisation de l'*Apartheid*, l'Afrique du Sud connaît une situation de ségrégation raciale caractérisée par la domination des Blancs.

³ Dans ce tableau, le coefficient estimé mesure le lien linéaire entre le statut socioéconomique d'un enfant et celui du chef de sa famille.

catégorie professionnelle, soit encore le niveau d'éducation, comme indicateur du statut socioéconomique (Sewell et Hauser, 1975, Couch et Dunn, 1997, Borjas, 1995, Mulligan, 1999). De notre côté, nous utilisons le niveau d'éducation pour deux raisons. D'une part, il s'agit d'une variable clé qui explique l'accès inégal aux opportunités de la vie (Mwabu et Schultz, 1996). D'autre part, il existe une forte corrélation entre le niveau d'éducation et le niveau de vie (Mincer, 1974). Nous adoptons une double méthodologie. D'abord, nous estimons la mobilité intergénérationnelle en utilisant les données individuelles. À ce niveau, des régressions sur les quartiles sont conduites. Ces dernières ont pour mérite d'étudier l'état de la linéarité de la mobilité intergénérationnelle. De plus, puisque notre variable dépendante est discrète, nous spécifions aussi notre modèle sous la forme d'un logit ordonné et d'un probit ordonné. Le calcul des effets marginaux que nous effectuons nous renseigne aussi sur les propriétés de linéarité du processus de transmission intergénérationnelle du niveau d'éducation.

Par rapport à la littérature existante, nous ajoutons ensuite une mesure du niveau d'inégalité dans la communauté de résidence parmi les déterminants de la mobilité intergénérationnelle. Ensuite, nous introduisons la méthode des pseudo-panel pour pallier l'absence des données longitudinales. De surcroît, cette nouvelle approche minimise les problèmes d'attrition et de biais de sélection. Son autre avantage réside dans sa façon de reconnaître la présence d'éventuelles erreurs dans les variables et d'en tenir compte explicitement lors des estimations. Notre étude est, à notre connaissance, la première à utiliser une telle méthode pour étudier la mobilité intergénérationnelle. Elle est également la première à utiliser les données issues des *October Household Survey* (OHS), pour étudier la mobilité intergénérationnelle en Afrique du Sud⁴.

Au terme de ce travail de recherche, les conclusions suivantes s'imposent par rapport aux questions que nous nous posions au départ. En premier lieu : (i) le niveau d'éducation des chefs de famille est un déterminant important de la formation du

⁴ Étant tirés aléatoirement sur une base annuelle, ces échantillons sont reconnus pour leur forte représentativité de la population sud-africaine. De plus, les données qui en sont issues sont plus complètes que celles du *KwaZulu Income Dynamics Study* (KIDS), utilisées jusqu'alors pour analyser la mobilité intergénérationnelle en Afrique du Sud.

capital humain des enfants noirs. Cette faible mobilité intergénérationnelle qui caractérise la population noire n'est pas linéaire. Les enfants issus des parents peu éduqués sont les plus immobiles; (ii) outre leur impact direct à l'intérieur de la famille, les parents influencent aussi la formation du capital humain de leurs enfants à travers le choix de la communauté dans laquelle ils vivent; (iii) le niveau de l'inégalité par district de résidence est aussi un déterminant de la mobilité intergénérationnelle entrant dans la catégorie des caractéristiques de la communauté de résidence. Plus il y a d'inégalité dans un district, moins grand est le niveau de formation du capital humain; (iv) les enfants les plus défavorisés sont ceux dont les chefs de familles sont des femmes rurales; (v) enfin, lorsque la situation financière de la famille est mauvaise, ce sont les garçons qui quittent, en premier lieu, l'école pour aller sur le marché du travail. En somme, c'est la pauvreté, par ses multiples facettes, qui se trouve au centre des problèmes de l'éducation des Noirs d'Afrique du Sud.

En deuxième lieu, sur le plan économétrique : (i) le fait de ne pas tenir compte des effets spécifiques des individus n'entraîne pas un biais dirigé vers le bas du niveau estimé de la mobilité intergénérationnelle, contrairement aux conclusions de Behrman et Taubman (1989). Notre conclusion va dans le même sens que celle de Bowles et Gintis (2000) ainsi que celle de Belzil et Hansen (2002) qui utilisent les données sur les États-Unis. Ces deux derniers auteurs reconnaissent particulièrement que les habiletés individuelles influencent la mobilité intergénérationnelles des niveaux d'éducation et des salaires, mais à un degré différent. D'après eux, alors que le niveau d'éducation des parents est plus important que les habiletés individuelles dans la détermination du niveau de formation du capital humain de leurs enfants, l'inverse est vrai quant à l'explication de leurs niveaux de salaires; (ii) la construction d'un pseudo panel nous permet d'éliminer les soupçons de biais susceptibles d'être la conséquence de ce problème d'endogénéité. Cependant, l'estimateur MCO appliqué aux données de notre pseudo-panel n'est pas efficace, en raison des erreurs de mesure de ses données liées la taille et aux variances des cellules. Ces deux problèmes sont corrigés par l'estimateur de Deaton (1985) et celui de Verbeek et Nijman (1993). Ces deux derniers estimateurs nous donnent des résultats similaires, une preuve qu'il n'y a pas de problème d'endogénéité dans les variables de pseudo-panel.

La suite de notre texte est structurée comme suit. Le deuxième chapitre résume les approches utilisées jusqu'ici pour modéliser la transmission intergénérationnelle du statut socioéconomique. Dans le troisième chapitre, nous présentons les données que nous utilisons et passons en revue quelques informations sur les enquêtes dont elles proviennent. Ensuite, dans le quatrième chapitre, nous décrivons la méthodologie des pseudo-panels qui minimise les problèmes d'estimation recensés dans la revue de littérature. Dans le cinquième chapitre, nous procédons à l'analyse de nos résultats économétriques. Une brève conclusion termine ce mémoire.

2. Un bref survol de la littérature

Certains travaux antérieurs sur la mobilité intergénérationnelle étudient comment les enfants héritent de leurs parents leur position dans la distribution des revenus. D'autres s'intéressent plutôt à la persistance intergénérationnelle dans les catégories socioprofessionnelles, ou encore utilisent le niveau d'éducation comme indicateur du statut socio-économique.

Trois grandes approches de l'analyse de la mobilité intergénérationnelle peuvent être inventoriées : (i) l'approche par régression linéaire univariée, (ii) les modèles linéaires multivariés et (iii) les méthodes postulant la non-linéarité de la mobilité intergénérationnelle.

2.1. Les modèles univariés

Cette première approche spécifie le statut économique de l'enfant comme une fonction linéaire du statut de son père. Une valeur estimée du coefficient de régression proche de zéro suggère une parfaite mobilité. Par contre, plus ce coefficient se rapproche de l'unité, plus un enfant a des chances d'avoir le même statut économique que son père. La majeure partie des études basées sur cette approche utilisent les données américaines de coupe transversale sur les revenus. Elles suggèrent la présence d'une forte mobilité intergénérationnelle aux États-Unis (É-U) avec une élasticité du revenu des enfants par rapport à celui de leurs pères inférieure à 0.2

(Sewell et Hauser, 1975 ; Behrman et Taubman, 1985 ; Becker et Tomes, 1986). Dans d'autres pays, les résultats sont variés, allant d'une forte mobilité en Norvège avec une corrélation de 0.14 (Soltow, 1965), à une faible mobilité en Grande Bretagne avec une corrélation de 0.36 (Atkinson, 1981).

Les estimations utilisant ces données en coupe transversale ont été critiquées. En effet, la mobilité intergénérationnelle mesure, par définition, la relation entre le statut économique de long terme d'un enfant et celui de son père, lesquels ne sont pas directement observables. Or, ces études utilisent les revenus courants comme approximations des revenus permanents. Les corrélations ainsi obtenues sont biaisées vers le bas en raison des fluctuations transitoires affectant les données disponibles. En effet, ces corrélations sont calculées en multipliant le coefficient estimé par le ratio de l'écart-type des revenus des enfants et celui des revenus des pères. Or, la variance des revenus courants est实质iellement élevée par rapport à celle de leurs revenus permanents (Solon, 1989)⁵.

Le recours à une variable instrumentale à la place du revenu courant du père peut améliorer les résultats des estimations. En utilisant le niveau d'éducation ou l'indice de Duncan⁶ comme variable instrumentale, Zimmerman (1992) rapporte une corrélation intergénérationnelle entre 0.42 et 0.49 aux É-U. L'utilisation des données longitudinales permet aussi de réduire ce biais d'estimation en ce sens qu'elles permettent d'utiliser la moyenne des revenus courants comme approximation du revenu permanent. En effet, cette moyenne minimise la composante transitoire des revenus, lorsqu'elle est calculée sur plusieurs années. L'usage de cette dernière comme variable explicative donne une plus grande corrélation intergénérationnelle des revenus aux É-U, de l'ordre de 0.4 (Solon, 1989). Cet avantage des données longitudinales est confirmé au Canada. Les coefficients de la corrélation intergénérationnelle estimés avec les données canadiennes en coupe transversale sont inférieurs à ceux obtenus à l'aide des données longitudinales (Corak et Heisz, 1999).

⁵ Solon (1992) montre que pour obtenir la vraie corrélation intergénérationnelle, il faudrait multiplier celle qui est estimée par un facteur égal au rapport entre la variance du revenu courant du père et celle de son revenu permanent.

⁶ L'indice de Duncan est un indicateur du statut socioéconomique qui représente le niveau de prestige d'une profession, prédit à partir du nombre d'années d'éducation et du niveau salarial correspondants.

Malgré l'usage des données longitudinales, un biais dans l'estimation peut persister si l'échantillon utilisé est homogène. En effet, le fait que la variance des revenus des pères tirés d'un échantillon homogène soit inférieure à celle de la population totale engendre un biais vers le bas de la corrélation intergénérationnelle mesurée par la racine carrée du R^2 (Behrman et Taubman, 1985 ; Sewell et Hauser, 1975). Quand les données utilisées sont tirées d'un échantillon représentatif, un biais de sélection peut apparaître du fait des critères de sélection retenus. Ces dernières peuvent engendrer une certaine homogénéité dans l'échantillon (Solon, 1989), qui provoque les mêmes problèmes que précédemment. Pour tenir compte de cette sélection, la spécification d'un modèle Tobit simple, dans lequel les critères de sélection sont aussi modélisés, est suggérée par Corak et Heisz (1999).

2.2. Les modèles multivariés

L'approche univariée, précédemment décrite, restreint les déterminants de la mobilité intergénérationnelle au seul statut socioéconomique du père. Or, la corrélation intergénérationnelle des revenus varient aussi selon les caractéristiques individuelles d'un enfant et ceux de sa famille d'origine, d'où la nécessité de spécifier des modèles multivariés de la mobilité. Behrman et Taubman (1985) montrent, par exemple, que l'impact de l'éducation du père est plus important que celui de l'éducation de la mère dans la détermination du niveau d'éducation des enfants. D'après les mêmes auteurs, l'effet de l'éducation paternelle est identique pour tous les enfants peu importe leur sexe, mais l'éducation maternelle est plus bénéfique pour les filles que les garçons. Tansel (2002) trouve aussi cet avantage comparatif des filles dans le niveau de formation de leur mère en Turquie. En général, les effets du père et de la mère sur la mobilité de leurs enfants sont plutôt variables. Dans certains pays, l'éducation du père est plus bénéfique pour l'enfant que celle de la mère (Schultz, 1984); dans d'autres, c'est l'inverse qui se produit (Schultz, 1993 ; Birdsall, 1995).

D'après Datcher (1982) et Cooper (1998), les modèles qui ignorent l'influence des communautés de résidence sont mal spécifiés. En effet, les parents influencent aussi le statut socioéconomique de leurs enfants à travers le choix de la communauté

dans laquelle ils vivent. Les caractéristiques de l'entité géographique dans laquelle les individus vivent conditionnent leurs perspectives socioéconomiques à plusieurs niveaux. Leur impact s'exerce d'abord à travers le financement local du système éducatif et ensuite à travers les effets purement sociaux (Durlauf, 1997). Burns (2001) recense quatre types de modèles qui expliquent ces effets sociaux des communautés sur les individus à savoir le modèle de contagion, les modèles de la socialisation collective, les modèles de privation relative ainsi que ceux de compétition. D'après les deux premiers types de modèles, grandir dans une communauté riche confère des externalités positives aux enfants de familles pauvres. Par contre, des deux autres types, il ressort une évidence d'externalités négatives pour les enfants pauvres vivant dans une communauté de riches.

Borjas (1995) étudie aussi le rôle des communautés dans le choix du niveau d'éducation. Il souligne l'impact du niveau moyen d'éducation de la génération de ses parents sur le niveau d'éducation des enfants d'une ethnie. C'est ce qu'il a appelé le «capital ethnique». S'inspirant de l'idée de Lucas (1988) de l'impact des interactions individuelles dans la détermination de la productivité individuelle, Michaud et Vencatachellum (2003) dénotent aussi la présence d'externalités intraraciales et interraciales positives du capital humain en Afrique du Sud.

De leur côté, Becker et Tomes (1986) réfléchissent sur l'impact des interactions individuelles à l'intérieur d'une famille. D'après leur théorie, l'impact sur le niveau d'éducation d'un individu du nombre de ses frères et soeurs varie en fonction de la situation financière de ses parents. En effet, dans le souci de garantir un minimum de revenus à leurs enfants, les parents pauvres partagent de façon équitable les ressources limitées destinées à leur éducation. Cela n'est pas le cas dans les familles riches qui ont la latitude de compenser un faible niveau de revenu de certains enfants par plus de transferts financiers.

Bowles et Gintis (2000) s'interrogent, en outre, sur la place de la transmission génétique et culturelle des caractéristiques familiales dans la persistance intergénérationnelle du statut économique. Auparavant, Behrman et Taubman (1989), utilisant une adaptation du modèle de Fisher (1918), avaient attribué une grande part à la transmission génétique dans la variance des niveaux d'éducation de la population

américaine. D'après ces derniers, le fait de ne pas tenir en compte les effets spécifiques individuels sous-estime le degré de la mobilité intergénérationnelle. Mais, d'après les résultats de Bowles et Gintis (2000), l'héritage des capacités intellectuelles contribue pour peu dans la transmission intergénérationnelle du statut économique aux É-U, une conclusion qui est confirmée par Gaviria (2002). Belzil et Hansen (2001) attirent aussi l'attention sur la différenciation souvent ignorée de l'impact des habiletés individuelles avec les chocs aléatoires. Leurs résultats montrent que les habiletés individuelles influent sur la mobilité intergénérationnelle des niveaux d'éducation et des salaires, mais à un degré différent. Alors que le niveau d'éducation des parents est plus important que les habiletés individuelles dans la détermination du niveau de formation du capital humain de leurs enfants, l'inverse est vrai quant à l'explication de leurs niveaux de salaires. En guise de justification à leurs résultats, les deux auteurs soulignent la faiblesse du rendement de l'éducation qui fait en sorte que les différences individuelles dans la formation du capital humain ne se répliquent pas sur le marché du travail.

2.3. Approches non-linéaires

Les précédentes approches supposent que la mobilité intergénérationnelle est linéaire, autrement dit qu'elle est identique pour toute la population d'un pays. À la suite de nombreuses recherches, cette hypothèse a été remise en cause. Par exemple, Becker et Tomes (1986) et Mulligan (1997) mettent en évidence l'existence d'une plus forte corrélation entre les revenus des parents et ceux des enfants pour les pauvres. En effet, les parents investissent dans la formation de leurs enfants jusqu'au niveau où le rendement marginal de cet investissement soit égal à celui d'un investissement sur le marché financier. Au-delà de cet équilibre, les parents investissent sur le marché financier quitte à soutenir, par la suite, leurs enfants à travers des transferts financiers (Becker et Tomes, 1986). Or, en plus de leurs ressources limitées, les parents pauvres ont des contraintes d'emprunt. Souvent, leurs enfants n'atteignent donc pas leur niveau optimal d'investissement en capital humain, d'où la forte corrélation entre leurs salaires et ceux de leurs parents.

L'approche des matrices de transition constitue une première alternative pour vérifier cette non-linéarité de la mobilité intergénérationnelle (Peters, 1992 ; Dearden, Machin et Reed, 1997). Ces dernières permettent de déterminer la position des enfants dans la distribution du statut économique en fonction de celle de leurs parents. L'avantage de ces matrices est d'offrir une présentation détaillée de la mobilité intergénérationnelle, à travers une classification croisée des statuts économiques des fils et pères (Corak et Heisz, 1999). Toutefois, Atkinson, Maynard et Trinder (1983) suggèrent, plutôt, de spécifier des modèles quadratiques de la mobilité pour vérifier la présence des non-linéarités, en raison des limites de l'approche des matrices de transition. En effet, avec cette dernière approche, le degré de persistance au sommet et à la base de l'échelle socioéconomique peut être exagéré. Cette tendance à l'absence de la mobilité s'explique par l'impossibilité d'étudier le déplacement vers un meilleur statut économique des enfants dont les parents se trouvent au sommet de la distribution ainsi que la détérioration des conditions socioéconomiques des enfants dont les parents se trouvent au plus bas de l'échelle.

À leur tour, Corak et Heisz (1999) proposent d'adopter la méthode des estimations non paramétriques. Cette méthode permet d'analyser la variation continue de l'effet marginal d'une variable, en fonction de son niveau et de celui des autres variables d'un modèle. Le point fort de cette méthode se résume au fait qu'elle n'impose aucune forme fonctionnelle particulière, dans la spécification du modèle (Corak et Heisz, 1999). Elle permet, en outre, de visualiser graphiquement, dans un espace à trois dimensions, la variation du coefficient de la mobilité intergénérationnelle en fonction des étages de la distribution des variables d'intérêt (Burns 2000). Mais, encore une fois, en présence de biais de sélection, les résultats aux extrémités soulèvent des interrogations (Corak et Heisz, 1999). Burns (2000) utilise, tout de même, cette approche sur les données du KwaZulu Natal et démontre la présence d'une trappe de pauvreté dans le niveau d'éducation en Afrique du Sud. Ses résultats indiquent que la persistance intergénérationnelle du niveau d'éducation en Afrique du Sud est plus grande et plus significative pour les enfants dont les parents se situent dans la queue inférieure de la distribution des niveaux d'éducation.

Finalement, Belzil et Hansen (2001) suggèrent l'utilisation d'un modèle de programmation dynamique à horizon fini, dans l'analyse de la mobilité

intergénérationnelle. D'après leur recherche, l'impact des antécédents familiaux sur le niveau d'éducation, estimé à l'aide des régressions de moindres carrés ordinaires, est biaisé car ces dernières omettent l'aspect dynamique de la question. En réalité, les individus font un arbitrage périodique entre la décision d'entrer sur le marché du travail et celle de continuer leurs études, en ayant comme objectif final la maximisation de leurs utilités présente et future.

En conclusion, quelques problèmes persistent dans l'analyse de la mobilité intergénérationnelle. En dépit de l'utilisation des données longitudinales, des échantillons représentatifs et de la prise en compte des critères de sélection dans la modélisation, les erreurs de mesure dans les variables, les problèmes d'attrition et/ou d'endogénéité conduisent toujours à des estimations biaisées de la mobilité intergénérationnelle. Dans notre mémoire, nous tenons compte de ces problèmes qui restent soulevés.

3. Les données

En Afrique du Sud, il existe une seule base de données longitudinales, le *KwaZulu Income Dynamic Survey* (KIDS) qui, malheureusement, est incomplète. Mais, il existe une série de données en coupe transversale. Par exemple, depuis 1993, *Statistics South Africa* mène une enquête annuelle sur les ménages et les individus. Les données utilisées dans notre analyse proviennent de cette série d'enquêtes, dénommées *October Household Survey* (OHS).

Les OHS sont des échantillons aléatoires, tirés de façon indépendante d'une année à l'autre. Au départ, l'objectif était d'y collecter des informations sur les ménages et les individus, indispensables pour la conduite du Programme de Reconstruction et de Développement (RDP) du gouvernement post-apartheid. Plusieurs données y sont compilées, notamment celles sur l'habitat, la qualité de la vie, l'emploi, les secteurs d'activités formel et informel, les naissances, les décès, etc. Leur processus d'échantillonnage est en deux étapes. Chaque année, environ 30,000 ménages sont tirés de façon aléatoire dans les différentes zones de résidence. Ces dernières sont aussi tirées aléatoirement sur base des informations provenant des

recensements généraux de la population. Il est donc impossible de suivre les ménages et les individus à travers les enquêtes.

La distribution géographique de la population selon les races ainsi que le secteur de l'habitat (rural/urbain) sont également pris en compte lors de l'échantillonnage. À l'exception de l'OHS de 1993, ces bases de données renferment les informations sur les ex États autonomes, qui ont été absorbés à l'intérieur des 9 provinces actuelles⁷ depuis 1994. Pour cette recherche, nous utilisons les données échantillonées lors des enquêtes de 1995 à 1999⁸. Pour rendre ces bases opérationnelles, nous effectuons plusieurs transformations que nous expliquons dans les paragraphes qui suivent.

3.1. La construction des couples enfants - chef de famille

Notre première tâche est de constituer les couples enfant – chef de famille, à partir des échantillons des 5 enquêtes. Conformément à la définition de la mobilité intergénérationnelle, notre étude s'intéressent aux enfants adultes. Nous avons retenu les individus de la génération des enfants âgés d'au moins 15 ans. Notre limite d'âge maximise du nombre total des observations, indispensable en raison de la méthodologie que nous utilisons. En outre, elle se rapproche de l'âge de 16 ans, utilisé par *Statistics South Africa* pour délimiter la population active.

Les quatre races présentes en Afrique du Sud sont représentées dans les échantillons, en proportion de leurs parts dans la population sud-africaine. Selon le tableau (2), la proportion des Noirs varie entre 75% et 81% de l'échantillon, de 1995 à 1999. Les Métis représentent entre 9% et 13%, et les Blancs sont entre 4.5% et 7.1%. Enfin, les Indiens comptent pour une part comprise entre 2.3% et 3.8%. D'après le même tableau, il existe une grande concentration de certaines races dans quelques

⁷ Les 9 provinces de la RSA actuelle sont les suivantes : Western Cape, Eastern Cape, Northern Cape, Kwazulu Natal, Mpumalanga, Gauteng, Free State, North-West et Limpopo (connue sous le nom de Northern Province jusqu'à la fin de 2001).

⁸ Nous avons opté de ne pas utiliser les enquêtes de 1993 et 1994. En effet, les caractéristiques individuelles des individus en fonction des races ne sont pas conformes à la réalité dans la base de 1994.

provinces. Les Indiens sont en quasi-totalité dans la province de KwaZulu Natal tandis que les Métis prédominent dans celle de Western Cape. Les Blancs sont plus nombreux dans les provinces de Western Cape et de Gauteng. La distribution des Noirs entre les provinces est la moins dispersée. Étant donné les exigences en nombre d'observation par province de notre méthodologie, nous restreignons notre analyse aux Noirs. Par ailleurs, ce sont ces derniers qui occupent le bas de l'échelle sociale en RSA et qui ont donc plus de problèmes de mobilité.

La littérature de la mobilité intergénérationnelle insiste aussi sur le fait que c'est le statut socioéconomique de long terme des individus qui doit être étudié. La plupart des études qui utilisent le niveau d'éducation comme indicateur du statut socioéconomique ont alors exclu les données sur les enfants encore sur les bancs de l'école. Deux remarques, allant à l'encontre de cette façon de procéder, peuvent être formulées. Premièrement, elle engendre un biais de sélection en excluant les enfants qui terminent leurs études au-delà de l'âge normal de fin de scolarité. Ces derniers comprennent les personnes qui ont commencé l'école avec un retard et/ou qui ont redoublé. Deuxièmement, un enfant à l'âge adulte ou même un chef de famille peut décider de quitter son emploi pour retourner à l'école (Belzil et Hansen, 2002). À partir des informations contenues dans le tableau (3), ce phénomène de retour à l'école est mis en évidence. La part des chefs de famille qui vont à l'école va jusqu'à 4%. Par ailleurs, la part des individus de la génération des enfants qui sont à l'école va jusqu'à 51%, dans l'échantillon de 1997. De ce fait, nous n'avons pas exclu les individus qui sont encore à l'école, qu'ils soient de la génération des enfants ou de celle des chefs de famille. En agissant ainsi, nous aurions perdu jusqu'à la moitié des individus de certains de nos échantillons, avec des conséquences négatives sur l'efficacité de la méthodologie de pseudo-panels que nous proposons d'utiliser.

Par contre, il nous a été impossible de déterminer avec certitude le chef de famille dans certains ménages. Soit plusieurs personnes y sont présentées comme chef de famille, ce qui porte à confusion, soit le chef de famille est plus jeune que ses enfants. Ces observations ont alors été exclues. Ont aussi été exclus, les individus pour lesquels les informations sur le niveau d'éducation et/ou celui du chef de famille ne sont pas spécifiées durant les enquêtes. Les nombres de couples retenus à chaque étape de cette sélection sont repris dans le tableau (3).

3.2. Analyse statistique

Dans les lignes qui suivent, nous expliquons les transformations effectuées sur les variables d'intérêt de notre recherche et dégageons quelques constats descriptifs.

L'éducation des individus de la génération des enfants et celle des parents

À partir des informations disponibles sur le niveau d'éducation, nous calculons une nouvelle variable correspondant au nombre d'années d'éducation réussies. Nous nous référons aux critères établis par Moll (1993), repris dans le tableau (4). Dès lors, le nombre d'années d'éducation des individus varie de 0, pour les personnes qui ne sont pas allées à l'école, à 15 et 16 années, respectivement pour celles qui ont fait des études universitaires de premier cycle et des cycles supérieurs. Le tableau (3) indique la moyenne de cette variable pour les individus des deux générations. En moyenne, le nombre d'années d'éducation des enfants est supérieur à celui des parents. Au sein d'une même génération, il n'y a que de faibles variations intertemporelles.

Les caractéristiques des individus et des familles

Le tableau (3) résume la répartition des enfants selon leur sexe et selon celui des chefs de leurs familles. De façon générale, les enfants sont équitablement répartis selon le sexe. Le pourcentage de ceux de sexe féminin, dans les différents échantillons, varie entre 50% et 52%. Les hommes représentent entre 48% et 50%. La répartition des enfants selon le sexe de leurs chefs de famille est un peu moins équilibrée. La part de ceux dont les chefs de ménages sont des femmes varie entre 43% et 52%. Les enfants dont les familles sont dirigées par des hommes représentent entre 47% et 57% des échantillons.

Des informations additionnelles sur l'âge sont contenues dans le tableau (3). Les enfants ont en moyenne entre 23 et 24 ans. Ces résultats placent leur période de scolarité durant les décennies 1980 et 1990. Les chefs de famille sont âgés entre 52 et 55 ans, en moyenne. Donc, ils étaient à l'école durant les années 1950 et 1960. Une analyse des différences du niveau d'éducation, selon le sexe, a aussi été faite. Les filles cumulent, en moyenne, plus d'années d'éducation que les garçons. À ce sujet, l'Afrique du Sud se démarque du reste des pays en développement (Thomas, 1996).

Une variable additionnelle représentant le nombre total d'enfants dans chaque famille est également construite. Ce dernier varie en moyenne entre 3 et 5.

Les caractéristiques des communautés

D'après notre analyse statistique, il existe une grande dispersion du niveau d'éducation à l'intérieur des provinces, dans la génération des enfants. Le nombre moyen d'années d'éducation des enfants, dans les provinces, oscille entre 7 et 9 années de scolarité. On remarque une supériorité des provinces de Gauteng et Western Cape dans le niveau d'éducation. Il y a aussi un lien entre le secteur urbain/rural et le niveau d'éducation des individus. Dans la génération des enfants, les individus vivant en zone rurale ont, en moyenne, deux années d'éducation de moins que leurs congénères vivant en zone urbaine.

4. Le modèle

Nous présentons, ici, le modèle économétrique que nous allons estimer. Afin de mettre en exergue notre contribution par rapport à la littérature, nous partons du modèle utilisé dans les études antérieures. Ensuite, nous recensons les raisons de la persistance d'un biais dans l'estimation de la corrélation intergénérationnelle. Cette dernière étape nous permet de justifier les changements, dans le processus d'estimation, que nous proposons.

4. 1. La spécification du modèle

Tel que nous l'avons souligné dans la revue de littérature, l'estimateur des moindres carrés ordinaires (MCO) de la mobilité intergénérationnelle, utilisant les données en coupe transversale est biaisé pour plus d'une raison. Nous spécifions notre modèle en utilisant les données longitudinales, pour tenir compte de cette remarque.

Notre modèle prend alors la forme suivante :

$$y_{it}^e = \alpha + y_{it}^f \beta + x_{it}^e \gamma + x_{it}^f \varphi + z_{it} \delta + \sum_{t=1}^4 \lambda_t D_t + \theta_{it}^e + \varepsilon_{it} \quad (1)$$

où les variables y_{it}^e et y_{it}^f représentent, respectivement, le nombre d'années d'éducation d'un individu et celui du chef de ménage au sein d'une même famille i et à la période t . La variable x_{it}^e représente les caractéristiques de l'individu, incluant le sexe ainsi que l'âge. La variable x_{it}^f , quant à elle, correspond aux antécédents familiaux qui regroupent le sexe du chef de famille et le nombre total d'enfants dans la famille⁹. La variable z_{it} regroupe l'ensemble des caractéristiques de la communauté de résidence de la famille, à savoir le secteur urbain ou rural ainsi qu'une variable qui contrôle les externalités intraraciales. Cette dernière correspond au niveau moyen par district de résidence du capital humain de la génération parentale. Les D_t représentent la série des variables muettes correspondant aux effets fixes des périodes de collecte des données. θ_{it}^e représente les effets fixes de l'individu i observé à la période t . ε_{it} représente un aléa indépendamment et identiquement distribué. Par ailleurs, β , γ , φ et δ sont des coefficients de régression et α est une constante.

Le coefficient de régression β nous renseigne sur le degré de la mobilité intergénérationnelle dans le niveau d'éducation. Conformément à la théorie, nous anticipons un signe positif pour ce coefficient. En outre, plus il se rapprochera de zéro, plus la mobilité intergénérationnelle sera élevée. À l'inverse, un coefficient proche de 1 signifie une grande persistance intergénérationnelle des niveaux d'éducation. La variable correspondant au sexe de l'individu est une variable muette égale à 1 pour le sexe féminin. Le coefficient qui lui correspond devrait être de signe négatif. En effet, les enfants de sexe féminin ont, en pratique, un niveau d'éducation inférieur à celui de leurs frères. D'après Rosenzweig et Schultz (1982), ce dernier résultat est la conséquence du faible rendement de l'éducation des filles sur le marché du travail. Les coutumes, la tradition ainsi que les contraintes financières peuvent aussi expliquer cette différenciation. Mais, dans le cas de l'Afrique du Sud, nous nous attendons à un signe positif, conformément au résultat de notre analyse statistique. Par

⁹ Le nombre total d'enfants est certainement une variable endogène (Dessy et Vencatachellum, 2003). Mais, étant déterminé à une période qui précède celle de leur scolarité, nous pouvons l'utiliser comme variable explicative puisque sa partie stochastique n'est pas corrélée avec les aléas de notre modèle.

ailleurs, l'impact du sexe du chef de famille sur l'éducation de ses enfants fait l'objet d'un large débat. Dans notre modèle, nous intégrons une variable muette égale à 1 pour les chefs de famille de sexe féminin. Comme nous l'avons mentionné dans la revue de littérature, le signe de son coefficient est incertain (Schultz, 1984, 1993, Birdsall, 1995).

L'ajout de l'âge des enfants dans les variables du modèle est, à son tour, motivée par le souci de prendre en compte leur stade du cycle de vie. En effet, l'âge conditionne la position de chaque individu dans la distribution du statut socioéconomique de sa génération (Solon, 1989). En général, plus un individu est âgé, plus il devrait avoir d'années d'éducation. Le signe attendu pour le coefficient de la variable âge est donc positif. Nous intégrons aussi le nombre total d'enfants dans la famille. Rappelons que l'impact du nombre d'enfants dans une famille varie en fonction de la situation financière des parents (Becker et Tomes, 1986). Pour les individus nés de parents pauvres, le nombre total d'enfants de la famille a un effet négatif sur leur niveau de scolarité. Dans les familles riches, le nombre total d'enfants n'affecte par contre pas le niveau de scolarité.

Pour mesurer l'impact des caractéristiques des communautés, nous définissons, d'abord, une variable muette égale à 1 pour le secteur rural. Le signe attendu du coefficient de cette variable est négatif (Burns, 2000 ; Tansel, 2002). Ensuite, poursuivant l'idée de Borjas (1989) ainsi que de Michaud et Vencatachellum (2003), nous ajoutons une variable pour le contrôle des externalités intraraciales. Cette dernière variable est représentée par le niveau moyen du capital humain de la génération parentale par district de résidence. Nous intégrons aussi les effets fixes des périodes que nous représentons par les variables muettes D_t , en utilisant l'année 1995 comme année de référence.

Enfin, des effets spécifiques individuels sont incorporés dans ce nouveau modèle. Parmi ces effets spécifiques θ^e_i , il y a lieu de citer le sexe de l'enfant et celui du chef de famille, ainsi que les variables indicatrices de la province de naissance. Hormis ceux-là, d'autres composantes des effets individuels sont non observables. Ces derniers comprennent, d'une part, les habiletés intellectuelles des individus, qui

sont positivement corrélées avec le nombre d'années de scolarité du chef de famille (Behrman et Taubman, 1989 ; Bowles et Gintis, 2000 ; Belzil et Hansen, 2002). D'autre part, ils contiennent les effets de l'apprentissage en bas âge. Ces derniers sont d'une importance capitale pour le développement des enfants. Ils sont perceptibles jusqu'à l'âge adulte et sont aussi corrélés avec le niveau d'éducation du chef de famille (Belzil et Hansen, 2002). Rappelons que, d'après Berhman et Taubman (1989), le fait de ne pas tenir en compte les effets spécifiques individuels non observables sous-estime le degré de la mobilité intergénérationnelle.

Pour mettre en exergue la nature des données à notre disposition, il importe d'indexer les variables individuelles par $i(t)$. Nous obtenons un nouveau modèle dont la forme est la suivante :

$$y^e_{i(t)} = \alpha + y^f_{i(t)} \beta + x^e_{i(t)} \gamma + x^f_{i(t)} \varphi + z_{i(t)} \delta + \sum_{t=1}^4 \lambda_t D_t + \theta^e_{i(t)} + \varepsilon_{i(t)} \quad (2)$$

Cet indice $i(t)$ montre que des individus échantillonnés d'une enquête à l'autre sont différents. Nous pouvons bien estimer ce modèle, en utilisant la technique des pseudo-panels.

4.2. Une analyse basée sur la méthodologie des pseudo-panels

Puisque les effets spécifiques non observés des individus peuvent être corrélés avec les variables explicatives du modèle (2), il est primordial d'en tenir compte dans la modélisation pour estimer de façon convergente l'impact de ces dernières sur la variable dépendante. Sinon, l'indépendance entre les régresseurs et les aléas, une des hypothèses de base des moindres carrés ordinaires, est alors violée, entraînant un biais dans les estimations. En général, cette correction nécessite l'utilisation des données longitudinales. Or, pour l'Afrique du Sud, nous ne disposons que des données en coupe transversale sur 5 années successives. Cette situation nous amène à utiliser pour la première fois la méthode des pseudo-panels dans l'estimation de la corrélation intergénérationnelle. Cette approche que nous ajoutons dans la littérature a pour premier objectif de pallier l'absence des données longitudinales. Mais, comme nous

l'expliquons plus loin, elle cumule d'autres atouts susceptibles de donner de meilleurs résultats par rapport aux travaux antérieurs.

Deaton (1985) a été le premier à appliquer cette méthode. Intuitivement, un pseudo-panel est obtenu en agrégeant les observations des individus à l'intérieur des groupes ayant des caractéristiques communes dans chaque enquête, que nous appelons par la suite «les cellules». En appliquant cette méthode, à notre situation, une manière alternative d'estimer le modèle (2) consiste à définir de nouvelles variables, correspondant à l'espérance - conditionnelle à l'appartenance à une cellule - des variables utilisées précédemment. L'espérance conditionnelle de ce modèle correspond à l'expression suivante :

$$E(y^e_i/g_i \in I_c)_{(t)} = \alpha + E(y^f_i/g_i \in I_c)_{(t)}\beta + E(x^e_i/g_i \in I_c)_{(t)}\gamma + E(x^f_i/g_i \in I_c)_{(t)}\phi + E(z_i/g_i \in I_c)_{(t)}\delta + E(\theta^e_i/g_i \in I_c)_{(t)} + \sum_{t=1}^4 \lambda_t D_t + E(\varepsilon_i/g_i \in I_c)_{(t)} \quad (3)$$

où :

- l'indice c dénote les cellules : $c = 1, \dots, C$
- l'indice t dénote les années : $t = 1995, \dots, 1999$.
- g est l'ensemble des caractéristiques qui déterminent l'appartenance aux cellules. Autrement dit, une famille $i \in c$ si et seulement si $g_i \in I_c$.

Il importe de signaler que la variable correspondant au niveau moyen du capital humain par district n'apparaît plus dans le modèle utilisant les données de pseudo-panel. En effet, les valeurs agrégées par cellule de cette variable et du nombre d'années d'éducation des chefs de famille par district sont presque identiques et donc très fortement corrélés. Le modèle (3) peut aussi s'écrire de façon simplifiée comme suit :

$$y^{e*}_{ct} = \alpha + y^{f*}_{ct}\beta + x^{e*}_{ct}\gamma + x^{f*}_{ct}\phi + z^{*}_{ct}\delta + \sum_{t=1}^4 \lambda_t D_t + \theta^{e*}_{ct} + \varepsilon^{*}_{ct} \quad (4)$$

Dans cette expression, les variables avec un astérisque correspondent aux valeurs moyennes de la population d'une cellule. Dès lors, les informations sur les cellules peuvent être suivies à travers le temps.

Cependant, les effets spécifiques agrégés des individus d'une cellule, représentés par θ^{*e}_{ct} , deviennent problématiques. Contrairement aux effets spécifiques individuels θ^e_{it} du modèle (1), ils ne sont pas constants dans le temps. Or, il est important de conserver les effets spécifiques aux individus dont l'impact sur la variable expliquée nous intéresse, à savoir le sexe de l'enfant, celui du chef de famille, ainsi que la province de naissance. Pour ce faire, il est possible que ces effets spécifiques observables soient utilisés pour la construction des cellules. Nous reviendrons sur la construction de ces cellules, plus tard. Ainsi, les effets spécifiques individuels agrégés peuvent être décomposés entre leurs composantes observables et non-observables, selon la relation suivante :

$$E(\theta^e_i / g_i \in I_c) = \theta^{*e}_{ct} = \psi_1 \text{ sexe}^e + \psi_2 \text{ sexe}^f + \sum_{k=1}^8 \psi_k D_k + v^{*}_{ct} \quad (5)$$

Les variables sexe^e , sexe^f et D_k représentent, respectivement, le sexe de l'individu, celui du chef de sa famille et sa province de naissance¹⁰. v^{*}_{ct} est une variable aléatoire représentant les effets spécifiques agrégés non-observables des individus.

En substituant l'expression de θ^{*e}_{ct} dans le modèle (4), nous obtenons le modèle suivant :

$$\begin{aligned} y^{*e}_{ct} = & \alpha + y^{*f}_{ct} \beta + x^{*e}_{ct} \gamma + x^{*f}_{ct} \delta + \sum_{t=1}^4 \lambda_t D_t + \psi_1 \text{ sexe}^e + \psi_2 \text{ sexe}^f \\ & + \sum_{k=1}^8 \psi_k D_k + v^{*}_{ct} + \varepsilon^{*}_{ct} \end{aligned} \quad (6)$$

¹⁰ L'indice k représente la province de naissance et varie entre 1 et 8, la province de Gauteng étant utilisée comme province de référence.

Pour ce qui est des effets non-observables v_{ct}^* , trois options sont possibles : ils peuvent être considérés comme (i) des effets fixes ; (ii) des effets aléatoires non-corrélés; (iii) des effets aléatoires corrélés; et enfin (iv) ils peuvent être ignorés.

D'après Browning, Deaton et Irish (1985), Verbeek et Nijman (1992), la première option n'est acceptable que lorsque la taille des cellules est au moins égale à 100. En effet, les v_{ct}^* tendent vers 0 à mesure que la taille des cellules augmente. Si cette option est retenue alors que les cellules ne sont suffisamment grandes, un problème d'identification surgit dans le modèle (Deaton, 1985). La deuxième alternative, qui correspond aux effets aléatoires non-corrélés, présente certains avantages par rapport à la première. D'abord, elle diminue le nombre de paramètres à estimer. Ensuite, elle permet de tirer des conclusions intéressantes quant à l'impact des effets fixes observables des individus, qui sont déterminants dans la formation du capital humain. Advenant que ces effets non-observables soient considérés comme étant aléatoires alors qu'ils sont corrélés avec y_c^* , un biais d'endogénéité apparaît dans l'estimation du coefficient de la mobilité intergénérationnelle, à l'instar des estimations qui utilisent les données individuelles. Cette situation correspond à la troisième option. Enfin, les v_{ct}^* peuvent aussi être ignorés. En effet, ce sont les effets spécifiques des cellules qui ont une grande probabilité d'être corrélés avec le niveau d'éducation moyen des chefs de famille par cellule, plutôt que les effets spécifiques individuels agrégés. En fonction de l'une des alternatives, les modèles définitifs que nous allons estimer s'écrivent comme suit :

-Modèle à effets spécifiques individuels agrégés fixes :

$$y_{ct}^{e*} = \alpha + y_{ct}^{f*} \beta + x_{ct}^{e*} \gamma + x_{ct}^{f*} \varphi + z_{ct}^* \delta + \sum_{t=1}^4 \lambda_t D_t + \sum_{c=1}^{35} \psi_c D_c + \varepsilon_{ct}^* \quad (7)$$

où les D_c représentent les effets fixes des cellules.

-Modèle à effets spécifiques individuels agrégés aléatoires : il correspond au modèle (6). L'estimation de ce modèle sous cette alternative suppose que les v_{ct}^* ne sont pas corrélés avec les variables explicatives.

-Modèle où les effets spécifiques individuels agrégés sont ignorés :

$$\begin{aligned}
 y^{e*}_{ct} = & \alpha + y^{f*}_{ct} \beta + x^{e*}_{ct} \gamma + x^{f*}_{ct} \varphi + z^{*}_{ct} \delta + \sum_{t=1}^4 \lambda_t D_t + \psi_1 \text{sexe}^e + \psi_2 \text{sexe}^f \\
 & + \sum_{k=1}^8 \psi_k D_k + \varepsilon^{*}_{ct}
 \end{aligned} \tag{8}$$

À notre connaissance, cette approche n'a jamais été utilisée, auparavant, pour analyser la mobilité intergénérationnelle. Pourtant, en plus de pallier l'absence des données de panel, elle minimise les sources de biais dans les estimations. En effet, dans la pratique, des limites s'imposent à l'utilisation des données longitudinales. Au fil du temps, certains individus sont perdus de vue, d'autres refusent de coopérer. Ces limites de maintenance systématique des données occasionnent les problèmes d'attrition, une source fréquente de biais que la méthode des pseudo-panels permet de juguler (Baltagi, 2001; Verbeek et Nijman, 1990). Qui plus est, puisque les échantillons utilisés par cette méthode sont des coupes transversales qui sont tirés annuellement, leur représentativité est constamment assurée (Deaton, 1985). À fortiori, la méthodologie des pseudo-panels cumule les atouts des techniques proposées dans la littérature de la mobilité intergénérationnelle. D'une part, elle minimise les erreurs dans les variables à travers son processus d'agrégation. D'autre part, elle peut être considérée comme un cas particulier de la méthode d'estimation par variables instrumentales (Deaton, 1985 ; Moffitt, 1993), proposée par Solon (1989), pour améliorer l'estimation de la corrélation intergénérationnelle.

Dans ce qui suit, nous expliquons les démarches suivies pour la construction de notre pseudo-panel avant de présenter les techniques d'estimation d'un modèle utilisant ce genre de données.

4.3. La construction des cellules et variables de pseudo-panel

Afin d'utiliser la méthode des pseudo-panels, nous identifions, d'abord, les variables présentes dans chacune des enquêtes, pouvant servir à la constitution de nos cellules. En général, il est recommandé de choisir des critères qui sont invariants dans le temps (Deaton, 1985; Verbeek et Nijman, 1992). Dans notre recherche, nous optons

pour un choix naturel des critères de construction des cellules. Outre le degré de mobilité intergénérationnelle, rappelons que nous désirons vérifier l'impact du sexe des individus et des chefs de famille, de la province de naissance et de la zone de résidence (urbaine/rurale), dans la formation du capital humain. De ces objectifs, quatre variables potentielles de référence dans la construction des cellules ressortent directement. Mais, tel que nous l'avons déjà énoncé, les critères de construction des cellules comprennent le sexe de l'enfant, celui du chef de famille ainsi que la province de résidence. Tenant compte des exigences de notre méthodologie par rapport à la taille des cellules, nous ne conservons que ces trois.

Remarquons que la province de résidence est utilisée à la place de celle de naissance, car cette information n'est pas disponible. Ce remplacement est acceptable si les flux migratoires entre les provinces sont négligeables. En principe, nous aurions exclu les individus qui ont émigré, si cette information était disponible. Pour les années 1993, 1994 et 1995, où cela est possible, nous calculons l'intensité des flux migratoires. Le pourcentage des individus qui émigrent reste en dessous de 2%, ce qui rend plausible notre hypothèse. Par ailleurs, la délimitation des cellules sur de grandes étendues, à savoir les provinces, minimise la probabilité de la mobilité géographique. De même, nous ne saurions nier l'avantage d'utiliser le secteur de l'habitat (urbain/rural) comme variable de repère dans la construction de notre pseudo-panel. Cependant, tel que le tableau (5) le montre, les effectifs de gens vivant dans le secteur rural dans les différentes provinces sont très bas. Nous continuons à tenir compte de cet élément, en intégrant, dans le modèle, une variable correspondant au pourcentage, par cellule, des individus vivants dans le secteur rural.

Les trois variables retenues nous permettent de construire 36 cellules que nous définissons dans la légende du tableau (6). Par exemple, la première cellule regroupe les enfants de sexe masculin vivant dans la province de Western Cape et dont les chefs de leurs familles sont des hommes. En suivant ces cellules sur 5 ans, de 1995 à 1999, nous obtenons ainsi un pseudo-panel de 180 observations. Le tableau (6) renferme aussi les informations sur la taille de ces cellules. Sur les 180 cellules de notre pseudo-panel, 29 ont moins de 100 individus. Ce dernier nombre correspond à l'effectif minimal par cellule, pour l'estimateur MCO soit sans biais (Browning, Deaton, et Irish, 1985). En outre, pour obtenir un bon estimateur MCO, il est essentiel

que les variances du nombre d'années d'éducation des enfants à l'intérieur des cellules dites aussi «variables intracellulaires» soient inférieures à la variance entre les cellules du nombre moyen d'années d'éducation des enfants que nous appelons par la suite «variance intercellulaire», calculées sur l'ensemble de la population (Deaton, 1985 ; Verbeek et Nijman, 1992). Dans le tableau (7), nous avons résumé les informations sur les statistiques descriptives des données de notre pseudo-panel, y compris la dispersion intercellulaire, calculée sur les échantillons, du nombre d'années d'éducation des enfants. Nous l'avons comparée aux dispersions intracellulaires contenues dans le tableau (6). Ces dernières sont largement supérieures à l'écart-type intercellulaire du niveau d'éducation des individus de la génération des enfants.

Dans ces conditions, tel que nous l'expliquons dans l'annexe A, il est probable que l'estimateur MCO de notre modèle utilisant les données de pseudo-panel soit aussi biaisé. Nous corigeons ces problèmes liés à la construction des cellules à l'aide des estimateurs proposés par Deaton (1985) ainsi que Verbeek et Nijman (1993)¹¹.

5. Analyse des résultats

Cette section contient les résultats de notre investigation économétrique. Leurs interprétations y sont également présentées, à la fois sur le plan empirique et économique. Dans un premier temps, nous estimons le modèle (2), qui utilise les données individuelles issues de nos 5 enquêtes OHS. Ensuite, des régressions utilisant les données de pseudo-panel sont conduites.

5.1. Données individuelles

Dans un premier temps, nous estimons le modèle (2) sur les données individuelles, décrit dans le précédent chapitre. Les résultats de l'estimateur MCO de ce modèle sont résumés dans la colonne (1) du tableau (8). Le coefficient du nombre

¹¹ La définition des estimateurs de Deaton (1985) ainsi que Verbeek et Nijman (1992) et une présentation de leurs processus d'estimation se retrouvent à l'annexe A.

d'années d'éducation du chef de famille qui renseigne sur le degré de la mobilité est estimé de façon significative à 0.21. Notre résultat est comparable à ceux trouvés dans les travaux antérieurs.¹² Le coefficient de la mobilité intergénérationnelle du nombre d'années d'éducation varie entre 0.17 et 0.4 aux États-Unis (Couch et Dunn, 1997 ; Borjas, 1995). De son côté, Thomas (1996) trouve des coefficients de mobilité variant entre 0.3 par rapport aux mères et 0.25 par rapport aux pères pour l'Afrique du Sud. Burns (2000) trouve un coefficient de la mobilité intergénérationnelle compris entre 0.33 par rapport aux mères et 0.17 par rapport aux pères, pour le même pays. En Allemagne, les résultats sont plus variables, allant de 0.03 à 1.33 (Couch et Dunn, 1986 ; Gang et Zimmerman ; 2000).

Le coefficient associé au niveau du capital humain est aussi significativement estimé à 0.21. D'après ce résultat, vivre dans un district de gens ayant une formation poussée confère une externalité positive considérable aux enfants, en ce qui concerne leur niveau d'éducation. Cela concorde avec les résultats des études antérieures. Datcher (1982) trouve, par exemple, qu'une augmentation de 10% du revenu moyen des parents par district de résidence augmente d'un dixième le nombre d'années d'éducation de leur enfants. Ainsi, outre leur impact direct à l'intérieur de la famille, les parents influencent la formation du capital humain de leurs enfants à travers la communauté dans laquelle ils vivent. Plus précisément, les individus bénéficient de façon équivalente d'une année de plus d'éducation de leur parents ou de la moyenne du niveau du capital humain par district.

Quant à l'effet du sexe des chefs de famille, deux constats sont à faire :

- (i) les individus issus de familles dirigées par les femmes ont moins d'années d'éducation que les autres. Ce résultat n'est pas surprenant si l'on se rappelle que la plupart des familles dirigées par les femmes sont monoparentales. En Afrique du Sud, les femmes chefs de ménages sont soit des veuves soit des filles-mères, ayant quitté l'école très tôt à cause d'une grossesse précoce. En général, ces deux catégories de femmes se retrouvent dans une situation matérielle très vulnérable.
- (ii) l'impact marginal du niveau d'éducation d'un chef de famille de sexe féminin est positif. En effet, l'effet croisé du sexe et du niveau d'éducation du chef de famille est

¹² Rappelons que les résultats de travaux antérieurs sont résumés dans le tableau (1).

estimé de façon significative à 0.03. Autrement dit, une année de scolarité additionnelle d'un chef de famille est plus profitable pour ses enfants, lorsque ce dernier est une femme.

Les résultats concernant l'impact du sexe de l'enfant dans la mobilité intergénérationnelle sont aussi intéressants. La variable dichotomique égale à 1 pour les enfants de sexe féminin a un coefficient estimé de 0.46. Cette supériorité du niveau d'éducation des filles, qui ressort de nos résultats, contredit la littérature, mais, elle confirme les résultats de notre analyse statistique ainsi que les conclusions de Thomas (1996). Contrairement à la littérature, l'âge de l'individu a un coefficient de signe négatif qui est fortement significatif. En s'inspirant de l'historique du système éducatif sud-africain, une explication plausible peut être formulée. En effet, notre résultat montre que la formation du capital humain des Noirs s'est amélioré avec le temps.

À l'exception de l'âge et du sexe des enfants, les signes du reste des coefficients estimés sont conformes à nos attentes. Ainsi, plus un individu a de frères et sœurs, moins il cumulera d'années d'éducation. D'après la littérature, cette situation caractérise les familles pauvres. Ce résultat va dans le bon sens si l'on se rappelle que la quasi-totalité de la population sud-africaine en dessous du seuil de pauvreté se retrouve parmi les Noirs. De plus, d'après nos estimations, un enfant vivant dans le secteur rural est plus souvent amené à quitter l'école par rapport à celui vivant en ville. Enfin, les coefficients des variables indicatrices pour les provinces sont tous significatifs. Ce résultat peut éventuellement s'expliquer par la présence d'une forte décentralisation du système éducatif sud-africain.

Dans un deuxième temps, nous effectuons des régressions sur les quartiles du modèle (2), dans le but d'étudier la linéarité de la mobilité intergénérationnelle. En empruntant la démarche de Thomas (1996), nous séparons les individus en fonction de leur sexe. Dans chacune de ces régressions, les coefficients associés au niveau d'éducation du chef de famille et au niveau du capital humain sont positifs et fortement significatifs. Qui plus est, la persistance du niveau d'éducation est forte chez les individus du premier quartile, indépendamment de leur sexe. Puis, elle décroît progressivement pour les quartiles supérieurs. Cependant, contrairement aux

résultats de Burns (2000), la persistance est plus forte chez les individus de sexe masculin. De même, l'effet positif du niveau moyen du capital humain par district diminue lorsqu'on se déplace du premier quartile vers les quartiles supérieurs. De par ce dernier résultat, les enfants dont les parents occupent le bas de l'échelle sociale retirent donc plus externalités positives du fait d'habiter dans un district où les individus ont en moyenne une bonne formation.

D'après nos estimations, l'impact du nombre total d'enfants sur l'éducation des filles n'est pas significatif, mais il l'est pour les garçons. Ainsi, lorsque une famille a beaucoup d'enfants, ce sont surtout les garçons qui en subissent les conséquences. Ils abandonnent leurs études pour aller sur le marché du travail, en vue de soutenir financièrement leurs parents. Cet impact sur l'éducation des garçons est plus perceptible pour le 3^{ème} quartile. Nous interprétons ce résultat par le fait que les parents des enfants du premier quartile n'ont pas de marge de manœuvre puisque la formation de leurs enfants est déjà minimale. Les parents du premier quartile sont donc tellement pauvres de sorte que le fait d'avoir peu ou beaucoup d'enfants n'a plus d'effet sur l'éducation de leurs enfants. L'effet du sexe du chef de famille, quant à lui, ne semble pas dépendre du quartile dans lequel se trouve l'enfant. Mais, nos résultats indiquent que les garçons écourtent leurs études plus que les filles lorsqu'ils grandissent dans un ménage dirigé par une femme. En outre, selon l'effet croisé du sexe et du niveau d'éducation des chefs de famille, l'éducation d'une femme chef de ménage est plus profitable pour les filles. Enfin, le fait d'habiter dans le milieu rural affecte négativement les filles plus que les garçons dans les deux premiers quartiles, mais cette tendance se renverse dans le troisième quartile.

Troisièmement, dans le but d'améliorer la robustesse de nos résultats, nous refaisons les estimations antérieures en ajoutant une nouvelle variable, parmi les caractéristiques de la communauté de résidence, mesurant le niveau d'inégalité de l'éducation des parents par district de résidence. Nous utilisons le coefficient de variation du nombre d'années d'éducation comme indicateur de ce niveau d'inégalité. L'ajout de cette variable constitue un des apports de notre recherche à la littérature sur la mobilité intergénérationnelle. Jamais, auparavant, le niveau d'inégalité n'avait été utilisé dans la modélisation en vue de déterminer la mobilité. Pourtant, quelques récents travaux mettent en évidence la relation entre le niveau de l'inégalité et la

mobilité intergénérationnelle du niveau d'éducation (Checchi, 2000 ; Binder et Woodruff, 2002). Mais, ces derniers se penchent plus sur l'étude de la causalité allant du niveau d'éducation vers le niveau de l'inégalité et n'abordent pas la causalité allant dans l'autre sens. D'autres travaux traitent plutôt de l'impact de l'inégalité sur le niveau d'éducation au niveau macroéconomique (Dessy et Vencatachellum, 2003).

Les résultats de cette nouvelle série de régressions sont aussi contenus dans le tableau (8). D'emblée, nos résultats indiquent que le niveau de l'inégalité dans la zone de résidence est un déterminant à part entière de la mobilité intergénérationnelle, au même titre que les autres caractéristiques de la communauté. L'estimateur MCO du coefficient de cette variable est significativement égal à -0.19. Ainsi, une réduction du niveau de l'inégalité à l'intérieur des districts devrait conduire à une amélioration du niveau d'éducation des enfants de leur ressort. Dans l'ensemble, les coefficients des autres variables explicatives du modèle ne changent pas et restent significatifs sauf celui du niveau moyen du capital humain par district qui passe de 0.21 à 0.17¹³.

Les nouvelles régressions sur les quartiles nous donnent aussi des résultats intéressants. D'abord, nous remarquons que le niveau de l'inégalité influence uniquement le niveau d'éducation des enfants du premier quartile. Ce résultat est conforme à nos attentes. En effet, ce sont ceux-là qui se trouvent en bas de l'échelle sociale qui souffrent plus des méfaits de la distribution non équitable des ressources du pays. Ensuite, une fois que nous tenons compte du niveau de l'inégalité par district, nous constatons que l'impact positif du niveau moyen du capital humain décroît du premier quartile vers le 2^{ème} quartile. Le coefficient de cette variable passe, de façon significative, de 0.15 à 0.20. Selon ce nouveau résultat, ce sont les enfants des classes sociales intermédiaires qui semblent profiter, beaucoup plus, des externalités positives d'habiter dans une zone où les gens sont en moyenne bien formés. Cette réalité explique, en partie, la persistance des inégalités socioéconomiques en Afrique du Sud, malgré la fin de l'*Apartheid*.

¹³ D'après la statistique de Fisher ($F (1, 83136) = 3.30$), la différence entre ces deux valeurs estimées est significative à un seuil de 10%.

Quatrièmement, puisque la variable dépendante de notre modèle peut être considérée comme étant discrète, nous l'estimons aussi, d'abord, comme un logit ordonné, puis comme un probit ordonné. Nous modélisons la probabilité que le niveau d'éducation des individus de nos échantillons prenne les différents nombres d'années de scolarité compris entre 0 et 16. Les résultats sont contenus dans le tableau (9). La valeur des coefficients estimés ne peut pas être comparée à celles des estimations précédentes. Ils concordent, en outre, avec ceux de Tansel (2002), qui utilise la même procédure sur les données de la Turquie. Mais, pour ce qui est du sens des effets des variables explicatives sur le niveau d'éducation des individus de la génération des enfants, ces résultats corroborent avec nos précédentes estimations. Ils confirment nos précédentes conclusions qui supposent la continuité de notre variable dépendante. Cependant, en ajoutant la mesure du niveau d'inégalité, les résultats trouvés diffèrent de ceux des régressions qui supposent la continuité de la variable du nombre d'années d'éducation. Dans les modèles logit et probit ordonnés, le niveau d'inégalité n'a plus d'impact significatif sur le nombre d'années d'éducation des enfants.

Nous présentons aussi, dans le tableau (10), les effets marginaux des variables indépendantes significatives de notre modèle, calculés au niveau moyen de ces dernières dans nos échantillons. Ces effets marginaux représentent l'impact d'une unité additionnelle de la variable indépendante correspondante, sur la probabilité qu'un individu se retrouve à chacun des niveaux d'éducation. Selon nos calculs, une année additionnelle du niveau d'éducation d'un chef de famille diminue la probabilité de ses enfants d'appartenir à chacun des 8 premiers niveaux d'éducation. En même temps, elle augmente la probabilité d'appartenir aux 11 autres niveaux supérieurs d'éducation. La probabilité d'appartenir à la classe supérieure est presque nullement affectée. Autrement dit, les enfants situés en bas de l'échelle des niveaux d'éducation sont plus sensibles aux niveaux d'éducation de leurs chefs de famille. Ainsi, nos résultats confirment la non-linéarité de la mobilité intergénérationnelle, mise en évidence précédemment à l'aide des régressions sur les quartiles.

5.2. Données de pseudo-panel

Nos estimations antérieures peuvent être suspectées d'être biaisées, pour les raisons que nous avons déjà énumérées. Ainsi, avant d'entériner nos résultats, il convient d'estimer notre modèle (6) qui utilisent les données de pseudo-panel. Les résultats de ces nouvelles estimations se trouvent dans le tableau (11).

Dans un premier temps, nous considérons l'hypothèse que les effets spécifiques agrégés des individus sont fixes, ce qui revient à estimer le modèle (7). Le test conjoint de Fisher¹⁴ que nous avons effectué pour vérifier l'hypothèse de nullité des coefficients associés à ces effets fixes nous invite à rejeter cette hypothèse. En retenant cette option selon laquelle les effets spécifiques des individus sont fixes, le coefficient de régression correspondant au nombre d'années d'éducation du chef de famille est estimé à 0.29. La valeur de ce coefficient qui renseigne sur le degré de la mobilité intergénérationnelle n'est pas inférieure à celle de 0.21 obtenu à l'aide des données en coupe. Ainsi, contrairement aux conclusions de Behrman et Taubman (1989), la prise en compte des effets spécifiques individuels agrégés n'augmente pas la mobilité intergénérationnelle des Noirs en Afrique du Sud. Cela confirme la présence l'immobilité des Noirs d'Afrique du Sud. Un autre constat différent par rapport aux résultats précédents et surprenant est le fait que la variable correspondant au nombre total d'enfants dans la famille n'est plus significative.

En raison de la taille de certaines de nos cellules qui n'est pas assez grande, la possibilité d'avoir des problèmes d'identification nous suggère l'estimation, dans un deuxième temps, du modèle (6) selon l'hypothèse que les effets spécifiques agrégés sont aléatoires. Ces effets aléatoires peuvent être corrélés avec les variables explicatives du modèle. Dans ce cas, l'estimateur MCO est biaisé. En considérant les effets individuels comme étant aléatoires, une hypothèse additionnelle sur l'absence de corrélation entre la partie non-observée des effets individuels et les variables explicatives de ce modèle (6) doit donc être vérifiée. Selon cette deuxième alternative, la corrélation intergénérationnelle du niveau d'éducation est toujours estimée à une

¹⁴ La valeur du test de l'hypothèse nulle selon laquelle tous les effets fixes sont nuls, calculée avec 35 et 140 degrés de liberté, est de 17.69.

valeur qui n'est pas inférieure à celle obtenue avec les données en coupe. Elle est égale à 0.27. L'estimation en faisant cette deuxième hypothèse nous permet, encore une fois, d'analyser l'impact des caractéristiques fixes des individus et des ménages. À ce niveau, nous constatons que, contrairement à nos attentes, l'effet du sexe du chef de famille n'est plus significatif, de même que l'impact de certaines provinces. Mais, le test du multiplicateur de Lagrange de Breusch-Pagan¹⁵ ne nous confirme pas la validité de cette deuxième option. En effet, la variance de ces effets aléatoires n'est pas significativement différente de zéro.

En vue de déterminer lesquels des effets fixes ou aléatoires devraient être utilisés, nous avons aussi effectué le test de Hausman¹⁶. Néanmoins, la valeur du test ne nous permet pas de rejeter l'hypothèse nulle selon laquelle les coefficients estimés selon les deux spécifications sont différents. Autrement dit, le fait que les effets spécifiques agrégés soient spécifiés comme étant aléatoires ou fixes importe peu. Ce résultat nous a amené à vérifier la dernière option qui suggère d'ignorer ces effets en estimant le modèle (8). Le fait d'enlever les effets spécifiques agrégés des individus suppose qu'ils ne sont pas corrélés avec les variables explicatives de ce modèle.

En ne tenant pas compte de ces effets spécifiques agrégés non-observables des individus, le coefficient de la mobilité intergénérationnelle estimé par MCO est de 0.25. La persistance intergénérationnelle dans le niveau d'éducation est toujours estimée à un niveau qui n'est pas significativement inférieur à celle obtenue à partir des données individuelles. Ce résultat nous permet de rejeter encore une fois l'idée selon laquelle l'utilisation des données en coupe transversale sous-estime le degré de la mobilité intergénérationnelle des Noirs en Afrique du Sud. Cependant, même avec cette dernière hypothèse, certaines contradictions par rapport à nos attentes subsistent. En particulier, alors que l'impact du secteur rural est accentué, le nombre total d'enfants dans les ménages n'est pas un déterminant significatif de la mobilité, de même que le sexe du chef de famille. De plus, les variables indicatrices de certaines provinces de résidence ne sont pas significatives.

¹⁵ La valeur du test Chi- carré, calculée pour 1 degré de liberté est de 1.74. L'hypothèse nulle à vérifier stipule que la variance des effets aléatoires est nulle.

¹⁶ La valeur du test de Chi- carré de Hausman, calculée pour 4 degrés de liberté, est de 2.53.

D'après la littérature sur la méthode des pseudo-panels, les contradictions de nos attentes, trouvées dans nos trois estimations précédentes, sont le résultat des erreurs de mesure dans les variables de pseudo-panel. Ces dernières ne sont pas à négliger dans notre cas, puisque la taille de certaines cellules est inférieure à 100. Heureusement les estimateurs mis au point par Deaton (1985) et Verbeek et Nijman (1993) tiennent compte explicitement des erreurs présentes dans les données de pseudo-panel. D'après les deux dernières colonnes de notre tableau (11), l'estimateur de Deaton et celui de Verbeek et Nijman donnent des résultats presque similaires, pour tous les coefficients. Les coefficients de la mobilité intergénérationnelle sont maintenant estimés de façon significative à 0.24. Mais, contrairement à nos précédentes estimations utilisant les données de pseudo-panel, la variable du nombre total d'enfants dans une famille redevient significative. L'augmentation du nombre total d'enfant dans une famille exerce un effet significativement négatif sur la formation du capital humain. De plus, les variables indicatrices des provinces de résidence sont toutes significatives, comme cela est le cas en utilisant les données en coupe.

Il est à noter que la formule pour le calcul de la variance de ces deux derniers estimateurs, développée par Verbeek et Nijman (1993), suppose une même taille pour toutes les cellules. Le fait que les cellules que nous avons constituées n'ont pas la même taille ne nous permet pas d'utiliser cette formule. Ainsi, pour avoir une idée sur l'efficience de ces estimateurs et sur l'intervalle de confiance des coefficients estimés, nous mettons à profit la technique du *bootstrapping*. D'après MacKinnon (2003), il existe plusieurs situations - la nôtre en est une - où le degré de significativité des coefficients estimés et leurs intervalles de confiance ne peuvent être calculés que par cette procédure. Les étapes requises par cette procédure sont les suivantes :

- (i) effectuer un tirage avec remise, dans l'échantillon initial, du même nombre d'observations que celui de cette échantillon initial ; (ii) faire les estimations voulues à partir du nouvel échantillon constitué et sauvegarder les résultats ; (iii) répéter ces deux premières étapes plusieurs fois pour générer une distribution de probabilité de chacun des paramètres estimés ; (iv) calculer la variance de ces paramètres à partir de leur distribution de probabilité.

Bref, l'idée de base du *bootstrapping* est d'effectuer un rééchantillonnage des observations, en générant d'autres échantillons à partir d'un tirage avec remise dans l'échantillon initial. Chacun des nouveaux échantillons obtenus, constitué de données appelées *bootstrap*, a la même taille que l'échantillon initial et peut être constitué de plusieurs rangées identiques résultant de la remise des données. L'estimateur $\beta(\alpha)$ est ensuite calculé pour chaque échantillon, de nouvelles valeurs étant obtenues pour chaque paramètre estimé. Cela permet, par la suite, d'identifier la tendance selon laquelle chaque paramètre est estimée à partir des données initiales.

Ainsi, le *bootstrapping* est un processus répétitif dont le nombre d'itérations doit être décidé par l'utilisateur. Pour le calcul de la variance des estimateurs, il est généralement recommandé de répéter le processus 200 fois. Dans ce travail, nous augmentons à 1000 le nombre de réplications. En effet, pour les intervalles de confiance des estimations, de bons résultats ne peuvent être obtenus que si l'on augmente le nombre d'itérations à 1000. Le calcul de la variance de l'estimation de chacun des paramètres de notre modèle s'effectue donc à partir de 1000 échantillons obtenus à l'aide des 1000 itérations effectuées. Au niveau pratique, nous définissons, à l'aide du logiciel STATA, un petit programme de calcul par *bootstrapping* de chacun des estimateurs.

Comme nous venions de le souligner, la raison première qui nous amène à utiliser la procédure du *bootstrapping* réside dans le fait que le calcul de la variance, tel que suggéré par Verbeek et Nijman (1993), implique une hypothèse forte sur la taille des cellules. En plus de cela, en rééchantillonnant la variable dépendante en même temps que les régresseurs, cette procédure nous permet de corriger le problème d'hétéroscédasticité susceptible de surgir avec l'utilisation les données de pseudo-panel (Freedman, 1981). Selon les valeurs de la statistique t de Student, calculés par cette méthode du *bootstrapping*, les coefficients obtenus des deux derniers estimateurs sont tous significatifs, ce qui concorde encore plus avec les enseignements de la littérature. Ces résultats confirment les doutes émis par Deaton (1985) et Verbeek et Nijman (1993) sur l'efficacité de l'estimateur MCO, lorsque les variances intracellulaires sont grandes et que la taille des cellules est petite. La ressemblance des résultats des deux derniers estimateurs confirme aussi notre hypothèse d'absence de

corrélation entre les effets spécifiques agrégés et non-observables des individus et les variables explicatives du modèle (6).

En définitive, les résultats de ces deux derniers estimateurs nous confirment que le niveau de la persistance intergénérationnelle du nombre d'années d'éducation n'est pas surestimé lorsqu'on le calcule à partir des données individuelles. Une explication plausible à cette sous-estimation est la présence d'une corrélation négative entre les habiletés spécifiques des individus et le niveau d'éducation de leurs chefs de familles. Ce constat rejoue l'idée de Belzil et Hansen (2002). Selon ces derniers, les effets spécifiques des individus représentent aussi l'impact de l'apprentissage à domicile, reçu avant l'âge normal de début de scolarité. Or, d'après ces deux auteurs, les chefs de famille très éduqués peuvent avoir un coût d'opportunité élevé en terme de temps consacré à l'encadrement de leurs enfants, ce qui entraîne une corrélation négative entre leur niveau d'éducation et les habiletés de leurs enfants.

Dans un dernier temps, nous refaisons tous nos précédents calculs impliquant les données de pseudo-panel, en intégrant notre mesure d'inégalité par district. Les résultats obtenus, résumés également dans le tableau (11), sont cohérents avec nos attentes. D'abord, les coefficients de la mobilité intergénérationnelle estimés varient entre 0.20 et 0.23. Ils nous amènent encore une fois à rejeter l'idée selon laquelle la non prise en compte des effets spécifiques individuelles sous estime la mobilité intergénérationnelle. Ensuite, le coefficient correspondant au niveau de l'inégalité par district est négatif et devient fortement significatif, à mesure que nous utilisons les procédures qui améliorent l'estimation à l'aide des données de pseudo-panel. Pour l'estimateur de Deaton et celui de Verbeek et Nijman, tous les autres coefficients sont significatifs et vont dans le bon sens.

6. Conclusion

Depuis 1994, des dispositifs visant à inscrire l’Afrique du Sud sur la ligne de la justice sociale et de l’équité ont afflué. Dans la foulée des réformes initiées, le secteur de l’éducation a été particulièrement visé. Les réformistes ont été incités à se tourner vers ce secteur par la place qu’il occupe dans la réduction de la pauvreté et des inégalités. Néanmoins, la concrétisation de ces réformes et leur adaptation au fur et à mesure qu’en apparaissent les limites et les imperfections dépendent largement de l’ampleur à laquelle les inégalités socioéconomiques, héritées de l’Apartheid, se transmettent de génération en génération.

En étudiant le degré de transmission intergénérationnelle du nombre d’années d’éducation en Afrique du Sud, ce mémoire s’inscrit dans cette perspective. D’après nos estimations, le coefficient de la mobilité intergénérationnelle du nombre d’années d’éducation des Noirs d’Afrique du Sud varie entre 0.23 et 0.29. Outre leur impact direct à l’intérieur de la famille, les parents influencent aussi la formation du capital humain de leurs enfants à travers le choix de la communauté dans laquelle ils vivent. Cette faible mobilité intergénérationnelle qui caractérise la population noire n’est pas linéaire. Les enfants issus des parents peu éduqués sont les plus immobiles. Par ailleurs, le niveau d’inégalité par district de résidence, mesuré par le coefficient de variation du nombre d’années d’éducation dans la génération des parents, s’avère être un déterminant important du niveau d’éducation des enfants. De plus, la prise en compte des effets spécifiques individuels que nous effectuons grâce à la méthode des pseudo-panels n’augmente pas le niveau estimé de la mobilité intergénérationnelle, contrairement à la littérature sur le sujet.

La conclusion sur laquelle débouche cette recherche est que la majorité des Noirs se trouvent dans une trappe de pauvreté. Dans ces conditions, les différents changements positifs qui ont été enregistrés jusqu’ici vont seulement profiter aux classes socioéconomiques intermédiaires. À terme, les familles qui ont pu se déplacer vers les zones résidentielles des Blancs vont donc bénéficier des externalités positives. Entre temps, la situation de ceux qui sont restés dans les zones défavorisées risque de

s'empirer. En somme, c'est la pauvreté, par ses multiples facettes, qui est au centre de la problématique de l'éducation des Noirs d'Afrique du Sud.

En définitive, la promotion de la mobilité intergénérationnelle des enfants noirs qui vivent dans des familles rurales dirigées par les femmes est recommandée. Cette recommandation permettrait à cette catégorie de la population qui occupe le bas de l'échelle sociale, de retrouver une place dans la représentation de la société. Pour ce faire, un programme général de lutte contre la pauvreté s'impose, en vue de parvenir à un mieux-être collectif et d'améliorer la qualité de vie des générations futures. En effet, pour que l'argument de l'intelligence dans le succès scolaire soit considéré comme juste, il importe que l'épreuve scolaire soit considérée comme ne mesurant que la performance scolaire et non les ressources économiques des parents ou le milieu social dans lequel évolue l'enfant. C'est en relevant ce défi, que le gouvernement post-*Apartheid* parviendra, à terme, à faire disparaître les inégalités sociales dans ce qu'elles ont de négatif.

Annexe A : Estimation utilisant des données de pseudo-panel

Estimation d'un modèle en utilisant des données de pseudo-panel

S'il était possible de calculer l'espérance conditionnelle des variables du modèle (6), à partir de la population totale de chaque cellule, ce modèle pourrait être estimé de façon convergente par MCO. Or, ce sont les espérances conditionnelles des variables, calculées à partir des échantillons, qui sont utilisées comme approximations des vraies espérances apparaissant dans la spécification du modèle. Lorsqu'elle est biaisée, cette approximation peut contaminer l'estimateur MCO. À cela s'ajoute le problème de corrélation entre les effets spécifiques individuels agrégés et les caractéristiques des cellules, dont l'absence est sujette à une hypothèse assez forte.

D'aucuns pourraient donc penser que l'estimation de ce modèle est moins efficace par rapport à celle utilisant les données longitudinales. Mais, cette limite de la méthode des pseudo-panels n'est qu'une apparence (Deaton, 1985). Les données longitudinales ne peuvent-elles pas contenir, elles aussi, des erreurs de mesure ? Cela est le cas, notamment, lorsque les informations sur les individus sont cueillies auprès de tierces personnes. Par ailleurs, il existe un moyen de corriger l'estimateur des MCO utilisant les données de pseudo-panel. Cette correction utilise la matrice des variances-covariances des erreurs de mesure des variables du modèle, estimée à partir des données individuelles (Deaton, 1985; Verbeek et Nijman, 1992, 1993). Ainsi, l'utilisation des données de pseudo-panels reconnaît la présence d'éventuelles erreurs de mesure et en tient compte, explicitement, lors de l'estimation.

Pour décrire cette procédure de correction, réécrivons le modèle (6) sous la forme matricielle simplifiée qui suit:

$$Y_{ct} = X_{ct}\beta + \varepsilon_{ct} \quad (9)$$

c dénote toujours les cellules et t la période de réalisation de l'enquête. X_{ct} est une matrice qui regroupe l'ensemble des régresseurs du modèle. Elle comprend deux sous-matrices Z_{ct} et D_{ct} . Ces dernières contiennent les caractéristiques des cellules qui sont variables dans le temps et les effets fixes observables des cellules, respectivement. Y_{ct} est un vecteur colonne correspondant à la variable dépendante.

Donc, les variables qui composent Y_{ct} et Z_{ct} renferment une marge d'erreurs, mais cela n'est pas le cas pour D_{ct} car elle contient les effets fixes spécifiques aux cellules qui sont mesurés sans erreur. ε_{ct} est un vecteur colonne des aléas du modèle (9).

En l'absence de corrélation entre les effets spécifiques agrégés des individus et les régresseurs du modèle et lorsque la taille des cellules est suffisamment grande, le modèle (9) est estimé, de façon convergente, par MCO (Deaton, 1985 ; Browning, Deaton, et Irish, 1985). Dans le cas contraire, Deaton (1985) suggère un estimateur tenant compte des erreurs de mesure dans les variables. Son estimateur se calcule comme suit :

$$\beta_{deaton} = (X^T X - CT\Sigma)^{-1} (X^T Y - CT\sigma) \quad (10)$$

C correspond au nombre total des cellules et T à celui des périodes. Σ est la matrice des variances-covariances des erreurs de mesure des régresseurs du modèle. Si k est le nombre de régresseurs du modèle, le rang de cette matrice est $(k+1, k+1)$ ¹⁷. σ est un vecteur colonne, de rang $(k+1, 1)$, des covariances entre les erreurs de mesure de la variable dépendante et celles des variables explicatives.

Pour déterminer la matrice Σ et le vecteur colonne σ , il suffit de compiler l'ensemble des données disponibles, puis d'estimer la matrice des variances-covariances des variables explicatives du modèle ainsi que le vecteur de la covariance entre la variable dépendante et ces variables explicatives. Cette procédure repose sur une hypothèse qui stipule que la variance de la variable dépendante à l'intérieur de chaque cellule que nous appelons variance intracellulaire ne dépend ni de la cellule dans laquelle on se trouve, ni de la période d'analyse. Signalons que les variables mesurées sans erreur ont une variance intracellulaire et des covariances qui sont nulles.

¹⁷ Nous ajoutons une unité au nombre total des régresseurs k , pour tenir compte de la constante.

Verbeek et Nijman (1992) proposent, quant à eux, une classe d'estimateurs différenciés par un paramètre $\alpha \in [0, 1]$, dont l'expression mathématique est la suivante :

$$\beta(\alpha) = (X^T X - \alpha C T \Sigma)^{-1} (X^T Y - \alpha C T \sigma) \quad (11)$$

Dans cette classe d'estimateurs, l'estimateur de Deaton correspond à un $\alpha = 1$ et celui des MCO à un $\alpha = 0$. Selon ces deux auteurs, en présence d'une corrélation entre les effets spécifiques agrégés des individus et les régresseurs du modèle, l'estimateur $\beta(\alpha)$ est sans biais, si et seulement si $\alpha = (T - 1)/T$. T correspond toujours au nombre total de nos bases de données de coupe transversale. Par cela, les deux auteurs démontrent que l'estimateur de Deaton est aussi biaisé que celui des MCO, dans ces conditions. Même lorsque la taille des cellules est grande, l'estimateur de Deaton reste biaisé si T est petit (Verbeek et Nijman, 1992). Par contre, la valeur de α importe peu, en l'absence de corrélation entre les effets spécifiques agrégés et les variables explicatives du modèle sont corrélés.

Le choix de cette valeur de α affecte aussi la variance de l'estimateur $\beta(\alpha)$. Cette dernière est une fonction croissante de α . Selon Verbeek et Nijman (1993), la valeur optimale de α , en terme d'efficience, est inférieure à $(T-1)/T$. Ainsi, d'une part, l'estimateur de Deaton est non seulement biaisé mais aussi inefficient. D'autre part, malgré le fait qu'il soit convergent l'estimateur de Verbeek et Nijman (1993) correspondant à $\alpha = (T - 1)/T$ est lui-même inefficient. Outre la valeur de α , la variance de $\beta(\alpha)$ dépend aussi des caractéristiques des cellules. L'efficacité dans la classe d'estimateurs utilisant les données de pseudo-panel dépend du nombre de bases de données disponibles, des critères de constitution des cellules, de la taille de ces dernières ainsi que de leur nombre.

En résumé, notre processus d'estimation utilisant la procédure de Deaton (1985) ainsi que celle de Verbeek et Nijman (1993) implique les étapes suivantes :

(i) compiler les données individuelles de coupe transversale dans une même base de données ; (ii) calculer la matrice des variances-covariances des régresseurs du modèle qui est une estimation de la matrice Σ des variances-covariances des erreurs de mesure

de ces régresseurs ; (iii) calculer à partir des mêmes données individuelles le vecteur des covariances entre la variable dépendante et les variables explicatives du modèle qui est une estimation du vecteur colonne σ ; (iv) calculer l'estimateur $\beta(\alpha)$ à partir des données de pseudo-panel, une valeur de $\alpha = 1$ donnant lieu à l'estimateur de Deaton et une valeur $\alpha = T/(T+1)$ soit $4/5$ correspondant à l'estimateur de Verbeek et Nijman.

Dans ce qui suit, nous rapportons les valeurs des éléments de la matrice Σ dans le cas de nos estimations utilisant une mesure du niveau d'inégalité par district de résidence comme variable explicative additionnelle :

La première ligne (et la première colonne) correspond à la variable représentant le nombre d'années d'éducation du chef de famille. Puis, vient successivement l'âge de l'enfant, le secteur de l'habitat (rural/urbain), le nombre d'enfant dans la famille, le niveau d'inégalité dans le district de résidence, le sexe de l'enfant et celui du chef de famille, les variables indicatrices de la province de résidence (sauf Gauteng considéré comme province de référence) ainsi qu'une constante. Les variances et les covariances intracellulaires des erreurs de mesure des variables utilisées pour la construction des cellules, à savoir le sexe de l'enfant et celui du chef de famille ainsi que les variables muettes indiquant la province de résidence, qui sont nulles nous

indiquent que ces variables sont mesurées sans erreur. Les variables restantes sont mesurées avec erreur tel que le montrent les valeurs de leurs variances et covariances.

Les valeurs des éléments du vecteur σ , quant à elles, sont les suivantes :

4.280
-3.020
-0.138
-0.010
-0.146
0
0
0
0
0
0
0
0
0
0
0

L'ordre par ligne des variables de ce vecteur correspond à celui de la matrice précédente. Il apparaît toujours que les covariances entre la variable expliquée du modèle et les variables explicatives utilisées pour la construction des cellules sont toutes nulles puisque ces dernières variables sont mesurées sans erreur.

Annexe B : Tableaux

Tableau 1: Résumé des résultats des études de la mobilité intergénérationnelle

Auteurs	Pays	Indicateur du statut socioéconomique	Données	Coefficient estimé
Soltow (1965)	Norvège	Revenus	Données en coupe	0.14
Sewell et Hauser (1975)	États-Unis	Revenus	Données en coupe	0.15
Atkinson (1981)	Royaume-Uni	Revenus	Données de panel	0.36
Atkinson (1983)	Royaume-Uni	Revenus	Données en coupe	0.43
Becker et Tomes (1986)	États-Unis	Revenus	Données en coupe	0.20
Behrman et Taubman (1985)	États-Unis	Revenus	Données en coupe	0.18
Behrman et Taubman (1990)	États-Unis	Revenus	Données de panel	0.40-0.60
Solon (1992)	États-Unis	Revenus	Données de panel	0.41-0.53
Zimmerman (1992)	États-Unis	Revenus	Données en coupe	0.25-0.42
Zimmerman (1992)	États-Unis	Revenus	Données de panel	0.40-0.54
Zimmerman (1992)	États-Unis	Salaires	Données en coupe	0.26-0.49
Zimmerman (1992)	États-Unis	Salaires	Données de panel	0.37-0.39
Peters (1992)	États-Unis	Revenus	Données de panel	0.17-0.23
Peters (1992)	États-Unis	Salaires	Données de panel	0.14-0.23
Corcoran et al (1992)	États-Unis	Revenus	Données de panel	0.37
Gustafsson (1994)	Suède	Revenus	Données de panel	0.20
Borjas (1995)	États-Unis	Salaires	Données de panel	0.22-0.46
Borjas (1995)	États-Unis	Années d'éducation	Données de panel	0.17-0.36
Thomas (1996)	Afrique du Sud	Années d'éducation	Données de panel	0.2-0.4
Dearden, Machin et Reed (1997)	Royaume-Uni	Salaires	Données de panel	0.22-0.59
Dearden, Machin et Reed (1997)	Royaume-Uni	Années d'éducation	Données de panel	0.42-0.44
Couch et Dunn (1997)	États-Unis	Salaires	Données de panel	0.13
Couch et Dunn (1997)	Allemagne	Salaires	Données de panel	0.12
Couch et Dunn (1997)	États-Unis	Années d'éducation	Données de panel	0.26-0.37
Couch et Dunn (1997)	Allemagne	Années d'éducation	Données de panel	0.03-0.48
Björklund et Jäntti (1997)	Suède	Revenus	Données de panel	0.22-0.36
Björklund et Jäntti (1997)	États-Unis	Revenus	Données de panel	0.29-0.52
Mulligan (1999)	États-Unis	Salaires	Données de panel	0.35-0.61
Mulligan (1999)	États-Unis	Niveau de consommation	Données de panel	0.41-0.71
Corak et Heisz (1999)	Canada	Revenus	Données de panel	0.196
Gang et Zimmermann (2000)	Allemagne	Années d'éducation	Données de panel	0.82 - 1.33
Burns (2000)	Afrique du Sud	Années d'éducation	Données de panel	0.17-0.33
Nimubona (2002)	Afrique du Sud	Années d'éducation	Données en coupe	0.21
Nimubona (2002)	Afrique du Sud	Années d'éducation	Pseudo-panel	0.25-0.28

Tableau 2 : Intensité raciale par province des personnes âgées d'au moins 15 ans

Provinces	Effectifs	OHS 95				OHS 96				OHS 97				OHS 98				OHS 99			
		Indiens	Noirs	Métis	Blancs																
Western Cape	Total	2558	2558	2558	2558	1126	1126	1126	2965	2965	2965	2965	1794	1794	1794	1794	2070	2070	2070	2070	
	Pourcentage	2.58	16.73	66.18	14.5	0.53	26.47	63.68	9.33	1.79	19.46	71.53	7.22	1.67	21.74	63.71	12.88	1.16	17.68	70.29	10.48
Eastern Cape	Total	5402	5402	5402	5402	2895	2895	2895	2895	3909	3909	3909	3909	2338	2338	2338	2338	3118	3118	3118	3118
	Pourcentage	0.56	84.17	11.88	3.39	0.35	89.33	7.01	3.32	0.38	88.85	9.18	1.59	0.04	86.66	10.56	2.57	0.48	84.89	9.85	4.33
Northern Cape	Total	993	993	993	993	536	536	536	536	1214	1214	1214	1214	840	840	840	840	850	850	850	850
	Pourcentage	0.3	32.33	56.29	11.08	0	38.62	58.21	3.17	0.08	28.09	66.72	5.11	0	37.02	56.19	5.24	1.06	37.65	54	7.29
Free State	Total	2273	2273	2273	2273	897	897	897	897	2201	2201	2201	2201	1239	1239	1239	1239	1554	1554	1554	1554
	Pourcentage	0.09	84.03	8.58	7.3	0	86.85	1.9	11.26	0.14	93.55	2.32	4	0.08	92.33	1.86	5.65	0	89.45	2.7	7.79
KwaZulu/Natal	Total	6422	6422	6422	6422	3018	3018	3018	3018	6262	6262	6262	6262	2955	2955	2955	2955	3666	3666	3666	3666
	Pourcentage	9.36	84.32	2.66	3.66	15.24	82.14	0.36	2.25	9.66	86.91	1.17	2.27	13.06	81.69	1.66	3.59	9.17	86.91	1.39	2.43
North-West	Total	2243	2243	2243	2243	1153	1153	1153	1153	2714	2714	2714	2714	1700	1700	1700	1700	2117	2117	2117	2117
	Pourcentage	1.69	88.32	5.22	4.77	0	93.5	4.08	2.43	0	95.65	1.81	2.54	0.41	94.71	1.18	3.71	0.43	94.38	1.94	3.26
Gauteng	Total	2602	2602	2602	2602	2484	2484	2484	2484	3698	3698	3698	3698	1835	1835	1835	1835	2615	2615	2615	2615
	Pourcentage	7.23	64.8	9.15	18.83	4.55	79.43	3.7	12.32	2.57	76.04	5.73	15.66	2.89	76.08	3.81	17.06	2.6	78.09	5.28	13.96
Mpumalanga	Total	3166	3166	3166	3166	1062	1062	1062	1062	2709	2709	2709	2709	1488	1488	1488	1488	2023	2023	2023	2023
	Pourcentage	1.58	91.31	2.31	4.8	0	93.31	0	6.69	0.92	95.57	0.74	2.77	0.13	94.22	0.74	4.91	0.54	94.71	0.49	4.25
Limpopo	Total	863	863	863	863	2164	2164	2164	2164	3276	3276	3276	3276	2025	2025	2025	2025	2725	2725	2725	2725
	Pourcentage	1.16	90.38	1.74	6.72	0	97.74	0.42	1.85	0	99.57	0.12	0.31	0.15	98.77	0.2	0.89	0.18	97.54	0.11	2.17
Afrique du Sud	Total	26522	26522	26522	26522	15335	15335	15335	15335	28948	28948	28948	28948	16214	16214	16214	16214	20738	20738	20738	20738
	Pourcentage	3.73	75.25	13.96	7.06	3.84	81.55	9.18	5.43	2.75	79.97	12.78	4.49	2.98	78.28	12.58	6.03	2.3	79.67	12.08	5.8

N.B :

OHS = *October Household Survey*

Tableau 3: Caractéristiques des enfants retenus dans les différents échantillons

	OHS 95	OHS 96	OHS 97	OHS 98	OHS 99
Nombres total d'observations					
Nombre d'enfants ayant plus de 15 ans	19962	12506	23151	12693	16523
Nombre de couples enfant - chef de famille utilisés	19548	11983	23035	12664	15941
Caractéristiques des couples retenus					
Enfants de sexe féminin en %	51.6	52.1	52.1	52.2	50.3
Enfants dont les chefs de famille sont de sexe féminin (%)	43.4	49.7	52.6	51.6	51.7
Enfants qui vont à l'école en %	48.9	51.4	48.6	47.4	46.4
Enfants dont les chefs de famille vont à l'école en %	0.5	1.6	3.6	2.1	1.3
Âge moyen des enfants	23.77	23.67	24.18	23.7	23.88
Nombre moyen d'années d'éducation des enfants	8.57	8.51	8.341	8.48	8.62
Âge moyen des chefs de famille	55.28	54.25	55.04	54.75	54.71
Nombre moyen d'années d'éducation des chefs de famille	4.7	4.53	4.35	4.41	4.71
Nombre total moyen d'enfants par famille	4.01	3.98	3.833	3.71	3.69
<u>Nombre moyen d'années d'éducation :</u>					
- des enfants de sexe féminin	8.74	8.71	8.58	8.74	8.87
-des enfants de sexe masculin	8.38	8.29	8.08	8.19	8.36
-des femmes chefs de ménage	4.63	4.4	4.17	4.35	4.54
-des hommes chefs de ménage	4.75	4.67	4.55	4.48	4.88
<u>Nombre moyen d'années d'éducation des enfants vivant à :</u>					
-Western Cape	8.77	9.43	9.05	9.25	8.83
-Eastern Cape	8.31	7.81	7.53	7.84	8.1
-Northern Cape	7.26	7.7	7.93	8.35	8.16
-Free State	8.6	8.6	8.35	8.62	8.65
-Kwazulu/Natal	8.63	8.46	8.29	8.4	8.28
-North-West	8.21	8.03	8	8.15	8.28
-Gauteng	9.81	9.47	9.34	9.56	9.75
-Mpumalanga	8.31	8.56	8.29	8.52	8.68
-Limpopo	9.15	8.69	8.65	8.5	8.9
<u>Nombre d'années d'éducation des enfants vivant en :</u>					
-zone urbaine	9.25	9.24	9	9.18	9.32
-zone rurale	8.06	7.97	7.88	7.97	8.04

N.B :

OHS = *October Household Survey*

Tableau 4: Transformation du niveau d'éducation en nombres d'années d'éducation

Niveau d'éducation ¹⁸	Nombre d'années d'éducation
N'est jamais allé à l'école	0
Grade 0	0
Grade 1 / Sub A	1
Grade 2 / Sub B	2
Sub A / Sub B / Grade 1 / Grade 2 / Standard 1	2
Grade 4 / Standard 2	4
Grade 5 / Standard 3	5
Grade 6 / Standard 4	6
Grade 7 / Standard 5	7
Grade 8 / Standard 6 / Form I	8
Grade 9 / Standard 7 / Form II	9
Grade 10 / Standard 8 / Form III	10
Grade 11 / Standard 9 / Form IV	11
Grade 12 / Standard 10 / Form V / Matric	12
NTC I	10
NTC II	11
NTC III	12
Diploma / Certificate with Std 9 or lower	13
Diploma / Certificate with less than Grade 12 / Std 10	13
Diploma / Certificate with Std 10	14
Diploma / Certificate with Grade 12 / Std 10	14
Degree	15
Postgraduate degree or diploma	16
Autre	*
Non spécifié	*

Source:

- Moll, Peter G. (1993), p.244.
- October Household Surveys 1995, 1996, 1997, 1998, 1999.

¹⁸ Les niveaux d'éducation sont définis selon la nomenclature du système éducatif sud-africain, d'où leur dénomination en anglais.

Tableau 5: Répartition des enfants retenus en fonction de leur zone de résidence

		OHS 95		OHS 96		OHS 97		OHS 98		OHS 99	
Provinces	Effectifs	Urbain	Rural								
Western Cape	Total	421	421	277	277	574	574	385	385	349	349
	Pourcentage	98.34	1.66	96.75	3.25	98.26	98.26	97.92	2.08	91.4	8.6
Eastern Cape	Total	4515	4515	2539	2539	3468	3468	2025	2025	2564	2564
	Pourcentage	31.67	68.33	31	69	25.92	25.92	30.81	69.19	35.65	64.35
Northern Cape	Total	297	297	202	202	336	336	306	306	293	293
	Pourcentage	80.13	19.87	91.09	8.91	89.29	89.29	83.66	16.34	92.49	7.51
Free State	Total	1901	1901	744	744	2045	2045	1142	1142	1344	1344
	Pourcentage	68.12	31.88	64.52	35.48	70.51	70.51	65.15	34.85	64.06	35.94
Kwazulu/Natal	Total	5286	5286	2334	2334	5415	5415	2412	2412	3072	3072
	Pourcentage	28.98	71.02	32.73	67.27	29.99	29.99	35.45	64.55	29.59	70.41
North-West	Total	1942	1942	1069	1069	2590	2590	1603	1603	1898	1898
	Pourcentage	39.55	60.45	30.22	69.78	28.03	28.03	30.26	69.74	38.41	61.59
Gauteng	Total	1621	1621	1862	1862	2789	2789	1390	1390	1941	1941
	Pourcentage	93.71	6.29	97.26	2.74	98.31	98.31	98.99	1.01	97.89	2.11
Mpumalanga	Total	2830	2830	962	962	2576	2576	1402	1402	1891	1891
	Pourcentage	32.58	67.42	26.82	73.18	32.84	32.84	35.52	64.48	43.52	56.48
Limpopo	Total	735	735	1994	1994	3242	3242	1999	1999	2589	2589
	Pourcentage	27.76	72.24	10.98	89.02	10.27	10.27	7.55	92.45	18.08	81.92
Afrique du Sud	Total	19548	19548	11983	11983	23035	23035	12664	12664	15941	15941
	Pourcentage	42.57	57.43	42.51	57.49	41.14	41.14	42.37	57.63	45.13	54.87

N.B: OHS = *October Household Survey*

Tableau 6: Statistiques descriptives du nombre d'années d'éducation des enfants par cellule

Cellules	OHS 95			OHS 96			OHS 97			OHS 98			OHS 99		
	Nbre	Moy.	Écart-type	Nbre	Moy.	Écart-type	Nbre	Moy.	Écart-type	Nbre	Moy.	Écart-type	Nbre	Moy.	Écart-type
M-WC-M	103	8.22	3.02	66	9.06	3.04	129	8.71	2.74	108	8.76	2.85	70	8.77	2.59
M-WC-F	103	8.77	3.16	66	8.65	2.86	133	8.27	2.85	68	9.15	2.77	78	8.15	2.84
M-EC-M	1107	8.20	3.21	590	7.57	3.27	698	7.16	3.28	443	7.53	3.44	607	7.83	3.29
M-EC-F	1131	7.76	3.25	678	7.10	3.53	1025	7.07	3.17	549	7.06	3.33	722	7.41	3.41
M-NC-M	89	7.12	3.68	46	7.39	3.69	101	7.57	3.63	88	8.45	3.02	65	7.77	2.87
M-NC-F	59	7.03	3.64	45	8.44	3.47	76	7.46	3.75	62	7.84	3.53	84	8.18	2.78
M-FS-M	560	8.65	3.06	189	8.74	2.77	505	8.03	3.09	288	8.61	2.73	316	8.50	2.86
M-FS-F	357	8.61	2.96	165	8.44	3.02	455	8.05	3.15	256	8.31	3.06	322	8.39	3.16
M-KN-M	1475	8.63	3.19	566	8.50	3.24	1229	8.13	3.20	515	8.45	3.11	748	8.11	3.35
M-KN-F	1073	8.26	3.37	470	7.84	3.57	1321	8.03	3.30	585	7.89	3.27	748	7.89	3.47
M-NW-M	545	8.13	3.36	298	8.10	3.62	615	7.89	3.38	400	8.21	3.41	503	8.17	3.53
M-NW-F	359	7.40	3.68	232	7.64	3.62	606	7.69	3.45	367	7.71	3.59	437	7.91	3.56
M-G-M	464	9.79	2.65	469	9.49	2.82	800	9.29	2.66	386	9.51	2.51	535	9.83	2.70
M-G-F	334	9.37	2.90	436	9.08	2.83	556	8.81	2.84	310	8.98	2.84	443	9.24	2.81
M-Mp-M	943	8.27	3.41	275	8.84	3.21	642	8.09	3.32	359	8.38	3.21	461	8.71	3.06
M-Mp-F	408	8.18	3.57	189	8.46	3.43	589	8.01	3.38	314	8.28	3.24	456	8.45	3.32
M-L-M	221	9.27	3.12	403	8.56	3.39	635	8.65	3.12	420	8.29	3.09	589	8.88	2.88
M-L-F	136	8.82	3.37	560	8.61	3.22	928	8.47	3.05	537	8.13	2.97	741	8.55	3.07
F-WC-M	93	8.95	2.63	70	10.00	2.42	157	9.47	2.35	98	9.42	2.71	94	9.19	2.65
F-WC-F	122	9.09	2.69	75	9.92	2.11	155	9.58	2.14	111	9.64	2.69	107	9.04	2.42
F-EC-M	1035	8.69	3.05	572	8.35	3.19	696	8.12	3.12	464	8.55	2.93	550	8.75	2.99
F-EC-F	1242	8.60	3.10	699	8.28	3.20	1049	7.86	3.11	569	8.24	3.02	685	8.53	3.20
F-NC-M	98	7.34	3.63	63	7.25	4.06	86	8.29	3.21	86	8.76	3.26	70	7.93	2.79
F-NC-F	51	7.65	3.39	48	7.88	3.73	73	8.51	2.95	70	8.19	3.35	74	8.69	3.20
F-FS-M	571	8.54	3.02	224	8.47	3.05	575	8.51	2.92	316	8.86	2.59	362	8.79	2.89
F-FS-F	413	8.60	3.07	166	8.79	3.00	510	8.76	2.76	282	8.64	2.86	344	8.88	2.97
F-KN-M	1484	8.99	3.19	672	8.78	3.00	1224	8.52	3.27	595	8.78	2.96	711	8.79	3.33
F-KN-F	1254	8.53	3.55	626	8.55	3.65	1641	8.46	3.27	717	8.47	3.19	865	8.35	3.38
F-NW-M	582	8.73	3.05	298	8.21	3.51	648	8.16	3.49	434	8.38	3.30	472	8.31	3.51
F-NW-F	456	8.29	3.40	241	8.08	3.71	721	8.21	3.35	402	8.23	3.44	486	8.69	3.31
F-G-M	495	10.14	2.57	509	9.75	2.56	818	9.76	2.43	360	9.96	2.28	518	10.27	2.46
F-G-F	328	9.81	2.69	448	9.51	2.79	615	9.34	2.64	334	9.74	2.62	445	9.56	2.68
F-Mp-M	976	8.45	3.30	294	8.53	3.23	709	8.68	3.11	375	8.66	2.99	474	8.66	2.96
F-Mp-F	503	8.25	3.71	204	8.33	3.43	636	8.30	3.34	354	8.73	3.26	500	8.88	3.01
F-L-M	222	9.51	3.00	424	8.69	3.35	648	8.95	2.93	392	8.61	2.90	557	9.33	2.72
F-L-F	156	8.76	3.78	607	8.87	3.18	1031	8.63	3.04	650	8.87	2.79	702	8.94	3.14

N.B :

- Nbre : nombre d'enfants par cellule
- Moy. : nombre moyen d'années d'éducation des enfants par cellule
- Écart-type : écart-type du nombre d'années d'éducation des enfants par cellule
- OHS : October Household Survey

- M-WC-M : Cellule des enfants de sexe masculin habitant dans la province de Western Cape et dont le chef de famille est un homme.
- M-WC-F : Cellule des enfants de sexe masculin habitant dans la province de Western Cape et dont le chef de famille est une femme.
- M-EC-M : Cellule des enfants de sexe masculin habitant dans la province de Eastern Cape et dont le chef de famille est un homme.
- M-EC-F : Cellule des enfants de sexe masculin habitant dans la province de Eastern Cape et dont le chef de famille est une femme.
- M-NC-M : Cellule des enfants de sexe masculin habitant dans la province de Northern Cape et dont le chef de famille est un homme.
- M-NC-F : Cellule des enfants de sexe masculin habitant dans la province de Northern Cape et dont le chef de famille est une femme.
- M-FS-M : Cellule des enfants de sexe masculin habitant dans la province de Free State et dont le chef de famille est un homme.
- M-FS-F : Cellule des enfants de sexe masculin habitant dans la province de Free State et dont le chef de famille est une femme.
- M-KN-M : Cellule des enfants de sexe masculin habitant dans la province de KwaZulu/Natal et dont le chef de famille est un homme.
- M-KN-F : Cellule des enfants de sexe masculin habitant dans la province de KwaZulu/Natal et dont le chef de famille est une femme.
- M-NW-M : Cellule des enfants de sexe masculin habitant dans la province de North-West et dont le chef de famille est un homme.
- M-NW-F : Cellule des enfants de sexe masculin habitant dans la province de North-West et dont le chef de famille est une femme.
- M-G-M : Cellule des enfants de sexe masculin habitant dans la province de Gauteng et dont le chef de famille est un homme.
- M-G-F : Cellule des enfants de sexe masculin habitant dans la province de Gauteng et dont le chef de famille est une femme.
- M-Mp-M : Cellule des enfants de sexe masculin habitant dans la province de Mpumalanga et dont le chef de famille est un homme.
- M-Mp-F : Cellule des enfants de sexe masculin habitant dans la province de Mpumalanga et dont le chef de famille est une femme.
- M-L-M : Cellule des enfants de sexe masculin habitant dans la province de Limpopo et dont le chef de famille est un homme.

- M-L-F : Cellule des enfants de sexe masculin habitant dans la province de Limpopo et dont le chef de famille est une femme.
- F-WC-M : Cellule des enfants de sexe féminin habitant dans la province de Western Cape et dont le chef de famille est un homme.
- F-WC-M : Cellule des enfants de sexe féminin habitant dans la province de Western Cape et dont le chef de famille est une femme.
- F-EC-M : Cellule des enfants de sexe féminin habitant dans la province de Eastern Cape et dont le chef de famille est un homme.
- F-EC-F : Cellule des enfants de sexe féminin habitant dans la province de Eastern Cape et dont le chef de famille est une femme.
- F-NC-M : Cellule des enfants de sexe féminin habitant dans la province de Northern Cape et dont le chef de famille est un homme.
- F-NC-F : Cellule des enfants de sexe féminin habitant dans la province de Northern Cape et dont le chef de famille est une femme.
- F-FS-M : Cellule des enfants de sexe féminin habitant dans la province de Free State et dont le chef de famille est une femme.
- F-FS-F : Cellule des enfants de sexe féminin habitant dans la province de Free State et dont le chef de famille est une femme.
- F-KN-M : Cellule des enfants de sexe féminin habitant dans la province de KwaZulu/Natal et dont le chef de famille est un homme.
- F-KN-F : Cellule des enfants de sexe féminin habitant dans la province de KwaZulu/Natal et dont le chef de famille est une femme.
- F-NW-M : Cellule des enfants de sexe féminin habitant dans la province de North-West et dont le chef de famille est un homme.
- F-NW-F : Cellule des enfants de sexe féminin habitant dans la province de North-West et dont le chef de famille est une femme.
- F-G-M : Cellule des enfants de sexe féminin habitant dans la province de Gauteng et dont le chef de famille est un homme.
- F-G-F : Cellule des enfants de sexe féminin habitant dans la province de Gauteng et dont le chef de famille est une femme.
- F-Mp-M : Cellule des enfants de sexe féminin habitant dans la province de Mpumalanga et dont le chef de famille est un homme.
- F-Mp-F : Cellule des enfants de sexe féminin habitant dans la province de Mpumalanga et dont le chef de famille est une femme.

- F-L-M : Cellule des enfants de sexe féminin habitant dans la province de Limpopo et dont le chef de famille est une homme.
- F-L-F : Cellule des enfants de sexe féminin habitant dans la province de Limpopo et dont le chef de famille est une femme.

Tableau 7 : Statistiques descriptives des variables de pseudo-panel

	Nombre d'observations	Minimum	Maximum	Moyenne	Écart-type
Âge moyen des enfants par cellule	180	20.66	29.04	23.86	1.77
Nombre moyen d'années d'éducation des enfants par cellule	180	7.03	10.27	8.52	0.65
Âge moyen des chefs de famille par cellule	180	48.85	59.49	54.52	2.02
Nombre moyen d'années d'éducation des chefs de famille par cellule	180	2.43	7.56	4.71	1.13
Pourcentage moyen des enfants vivant dans le secteur rural par cellule	180	0.00	0.93	0.45	0.31
Niveau du capital humain moyen par district	180	3.21	6.82	4.72	1.04
Nombre moyen d'enfants dans une famille par cellule	180	2.67	4.90	3.74	0.46
Mesure d'inégalité moyen par cellule ¹⁹	180	0.52	1.31	0.89	0.22

¹⁹ La mesure du niveau d'inégalité retenue, rappelons-le, correspond au coefficient de variation du nombre d'années d'éducation des chefs de famille calculé par district de résidence.

Tableau 8: Résultats des estimations utilisant les données individuelles

	MCO (1)	Enfants de sexe masculin (2)			Enfants de sexe féminin (3)			MCO (4)	Enfants de sexe masculin (5)			Enfants de sexe féminin (6)		
		1e quartile	2e quartile	3e quartile	1e quartile	2e quartile	3e quartile		1er quartile	2e quartile	3e quartile	1er quartile	2e quartile	3e quartile
Nombres d'années d'éducation du chef de famille	0,21 *** (0,004)	0,27 *** (0,007)	0,22 *** (0,006)	0,13 *** (0,005)	0,24 *** (0,008)	0,20 *** (0,007)	0,12 *** (0,004)	0,211 *** (0,003)	0,267 *** (0,008)	0,218 *** (0,007)	0,126 *** (0,005)	0,244 *** (0,007)	0,201 *** (0,007)	0,120 *** (0,004)
Âge de l'enfant	-0,03 *** (0,002)	-0,03 *** (0,003)	0,01 *** (0,009)	0,06 *** (0,003)	-0,05 *** (0,003)	0,003	0,06 *** (0,003)	-0,028 *** (0,001)	-0,035 *** (0,003)	0,009 *** (0,003)	0,060 *** (0,003)	-0,046 *** (0,003)	0,003	0,062 *** (0,002)
L'enfant est de sexe féminin	0,46 *** (0,021)							0,457 *** (0,021)						
Nombre total d'enfants dans la famille	-0,01 * (0,006)	-0,02 (0,013)	-0,02 * (0,011)	-0,03 *** (0,008)	0,01 (0,013)	-0,001 (0,009)	-0,01 (0,007)	-0,011 * (0,006)	-0,015 (0,013)	-0,021 * (0,011)	-0,026 *** (0,008)	0,008 (0,013)	-0,002 (0,010)	-0,008 (0,007)
Niveau du capital humain moyen par district de résidence	0,21 *** (0,010)	0,24 *** (0,022)	0,19 *** (0,018)	0,13 *** (0,017)	0,25 *** (0,021)	0,17 *** (0,017)	0,08 *** (0,010)	0,171 *** (0,022)	0,154 *** (0,048)	0,203 *** (0,039)	0,170 *** (0,031)	0,158 *** (0,049)	0,197 *** (0,039)	0,098 *** (0,023)
Mesure du niveau d'inégalités par district de résidence								-0,189 * (0,098)	-0,464 * (0,238)	0,061 (0,198)	0,228 (0,144)	-0,447 ** (0,226)	0,112 (0,173)	0,093 (0,023)
Le chef de famille est une femme	-0,25 *** (0,035)	-0,31 *** (0,094)	-0,31 * (0,069)	-0,34 *** (0,064)	-0,24 *** (0,080)	-0,29 *** (0,059)	-0,25 *** (0,048)	-0,245 *** (0,031)	-0,313 *** (0,085)	-0,303 *** (0,079)	-0,348 *** (0,066)	-0,235 *** (0,083)	-0,287 *** (0,064)	-0,247 *** (0,052)
Sexe du chef de famille *														
Nombre d'années d'éducation du chef de famille	0,03 *** (0,005)	0,02 * (0,012)	0,02 *** (0,009)	0,02 *** (0,007)	0,05 *** (0,011)	0,04 *** (0,008)	0,02 *** (0,006)	0,027 *** (0,005)	0,023 * (0,012)	0,017 * (0,010)	0,024 *** (0,008)	0,050 *** (0,012)	0,039 *** (0,009)	0,022 *** (0,006)
L'enfant vit dans le secteur rural	-0,25 *** (0,021)	-0,23 *** (0,054)	-0,19 * (0,048)	-0,18 *** (0,035)	-0,27 *** (0,40)	-0,23 *** (0,038)	-0,11 *** (0,024)	-0,255 *** (0,022)	-0,226 *** (0,053)	-0,194 *** (0,038)	-0,189 *** (0,034)	-0,265 *** (0,049)	-0,233 *** (0,038)	-0,116 *** (0,022)
L'enfant vit dans la province de :														
Western Cape	-0,43 *** (0,064)	-0,77 *** (0,191)	-0,72 * (0,11)	-0,46 *** (0,114)	-0,29 * (0,17)	-0,37 *** (0,094)	-0,18 *** (0,056)	-0,442 *** (0,074)	-0,825 *** (0,178)	-0,708 *** (0,094)	-0,436 *** (0,107)	-0,330 ** (0,154)	-0,358 *** (0,092)	-0,161 ** (0,064)
Eastern Cape	-0,81 (0,040)	-1,31 *** (0,088)	-1,24 * (0,079)	-0,83 *** (0,061)	-0,72 *** (0,081)	-0,73 *** (0,069)	-0,45 *** (0,047)	-0,829 *** (0,044)	-1,334 *** (0,085)	-1,242 *** (0,076)	-0,805 *** (0,067)	-0,765 *** (0,085)	-0,714 *** (0,076)	-0,439 *** (0,045)
Northern Cape	-0,87 *** (0,089)	-1,08 *** (0,151)	-0,95 * (0,161)	-0,29 * (0,157)	-1,08 *** (0,19)	-0,94 *** (0,152)	-0,47 *** (0,127)	-0,887 *** (0,087)	-1,131 *** (0,168)	-0,946 *** (0,160)	-0,276 (0,180)	-1,150 *** (0,222)	-0,937 *** (0,174)	-0,464 *** (0,127)
Free State	-0,18 *** (0,047)	-0,28 ** (0,111)	-0,24 ** (0,100)	-0,03 (0,066)	-0,36 *** (0,095)	-0,22 *** (0,070)	-0,24 *** (0,049)	-0,212 *** (0,052)	-0,330 *** (0,099)	-0,234 ** (0,094)	-0,002 (0,070)	-0,444 *** (0,108)	-0,209 ** (0,082)	-0,230 *** (0,063)
KwaZulu / Natal	-0,17 *** (0,039)	-0,21 ** (0,098)	-0,17 * (0,087)	-0,02 (0,049)	-0,24 *** (0,077)	-0,09 (0,062)	-0,02 (0,041)	-0,187 *** (0,044)	-0,225 *** (0,083)	-0,163 ** (0,072)	-0,005 (0,058)	-0,272 *** (0,080)	-0,080 *** (0,072)	-0,016 (0,041)
North West	-0,41 *** (0,047)	-0,61 *** (0,105)	-0,43 *** (0,096)	-0,04 (0,070)	-0,50 *** (0,105)	-0,27 *** (0,065)	-0,09 ** (0,048)	-0,438 *** (0,050)	-0,637 *** (0,106)	-0,420 *** (0,086)	-0,019 (0,081)	-0,561 *** (0,104)	-0,258 *** (0,084)	-0,076 (0,059)

Mpumalanga	0,24 *** (0.051)	0,16 (0.120)	0,36 *** (0.096)	0,51 *** (0.078)	-0,03 (0.098)	0,15 * (0.084)	0,20 *** (0.051)	0,240 *** (0.052)	0,195 * (0.115)	0,361 *** (0.097)	0,512 *** (0.082)	0,022 (0.098)	0,162 * (0.093)	0,202 *** (0.058)
Limpopo	0,53 *** (0.049)	0,61 *** (0.113)	0,64 *** (0.096)	0,57 *** (0.065)	0,42 *** (0.096)	0,48 *** (0.084)	0,29 *** (0.052)	0,526 *** (0.051)	0,629 *** (0.099)	0,638 *** (0.084)	0,562 *** (0.063)	0,411 (0.100)	0,467 *** (0.098)	0,284 *** (0.048)
L'année de l'enquête est :														
1996	-0,16 *** (0.035)	-0,26 *** (0.076)	-0,17 ** (0.080)	-0,05 (0.050)	-0,22 *** (0.067)	-0,14 ** (0.058)	0,001 (0.041)	-0,155 *** (0.035)	-0,279 *** (0.082)	-0,164 ** (0.065)	-0,052 (0.049)	-0,209 *** (0.071)	-0,153 *** (0.054)	-0,004 (0.037)
1997	-0,30 *** (0.029)	-0,44 *** (0.068)	-0,43 *** (0.062)	-0,23 *** (0.044)	-0,25 *** (0.059)	-0,21 *** (0.038)	-0,12 *** (0.036)	-0,306 *** (0.030)	-0,453 *** (0.069)	-0,435 *** (0.059)	-0,229 *** (0.044)	-0,243 *** (0.071)	-0,212 *** (0.045)	-0,120 *** (0.030)
1998	-0,15 *** (0.034)	-0,35 *** (0.087)	-0,25 *** (0.066)	-0,16 *** (0.045)	-0,09 (0.072)	-0,04 (0.059)	-0,04 (0.039)	-0,154 *** (0.034)	-0,355 *** (0.078)	-0,251 *** (0.066)	-0,149 *** (0.047)	-0,080 (0.069)	-0,047 (0.052)	-0,070 ** (0.035)
1999	-0,09 *** (0.032)	-0,38 *** (0.075)	-0,32 *** (0.054)	-0,16 *** (0.047)	-0,22 *** (0.060)	-0,15 *** (0.056)	0,03 (0.040)	-0,089 *** (0.032)	-0,405 *** (0.085)	-0,319 *** (0.056)	-0,152 *** (0.048)	-0,216 *** (0.065)	-0,148 *** (0.052)	0,025 (0.038)
Constante	7,59 *** (0.086)	5,80 *** (204)	7,32 ** (0.164)	8,52 *** (0.130)	6,41 *** (0.166)	7,83 *** (0.12)	8,84 *** (0.086)	7,951 *** (0.086)	6,644 *** (0.462)	7,208 (0.374)	8,115 *** (0.307)	7,257 *** (0.455)	7,614 *** (0.375)	8,667 *** (0.232)
Nombre d'observations	83171	40233	40233	40233	42938	42938	42938	83158	40229	40229	40229	42229	42929	42929
R2 (Pseudo R2)	0,14	(0,096)	(0,066)	(0,046)	(0,091)	(0,063)	(0,044)	0,139	(0,096)	(0,066)	(0,046)	(0,091)	(0,063)	(0,044)

N.B :

- La variable dépendante est le nombre d'années d'éducation des enfants.
- Colonne (1) = Régression par MCO.
- Colonne (2) = Régressions par quartiles pour les enfants de sexe masculin.
- Colonne (3) = Régression par quartiles pour les enfants de sexe féminin.
- Colonne (4) = Régression MCO en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Colonne (5) = Régressions par quartiles pour les enfants de sexe masculin, en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Colonne (6) = Régression par quartiles pour les enfants de sexe féminin, en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Les valeurs entre parenthèses indiquent les écart-types robustes des estimateurs.
- *, **, *** indiquent que le coefficient correspondant est significatif à 1%, 5% et 10% respectivement.

Tableau 9 : Estimés des modèles logit et probit ordonnés

	Logit ordonné	Probit ordonné	Logit ordonné	Probit ordonné
Nombres d'années d'éducation du chef de famille	0.12 *** (0.002))	0.07 *** (0.001)	0.12 *** (0.002))	0.07 *** (0.001)
Âge de l'enfant	-0.001 (0.001)	-0.002 *** (0.001)	-0.001 (0.001)	-0.002 *** (0.001)
L'enfant est de sexe féminin	0.29 *** (0.01)	0.16 *** (0.007)	0.29 *** (0.01)	0.16 *** (0.007)
Nombre total d'enfants dans la famille	-0.01 ** (0.003)	-0.005 ** (0.002)	-0.01 ** (0.003)	-0.005 ** (0.002)
Niveau du capital humain	0.11 *** (0.01)	0.07 *** (0.003)	0.11 *** (0.01)	0.06 *** (0.007)
Niveau d'inégalité par district			0.06 (0.06)	0.01 (0.03)
Le chef de famille est une femme	-0.16 *** (0.02)	-0.10 *** (0.01)	-0.16 *** (0.02)	-0.10 *** (0.01)
sex du chef de famille * Nombre Années d'éducation du chef de famille	0.01 *** (0.003)	0.008 *** (0.002)	0.01 *** (0.003)	0.008 *** (0.002)
L'enfant vit dans le secteur rural	-0.15 *** (0.01)	-0.10 *** (0.007)	-0.15 *** (0.01)	-0.09 *** (0.007)
L'enfant vit dans la province de :				
-Western Cape	-0.31 *** (0.04)	-0.17 *** (0.02)	-0.30 *** (0.04)	-0.16 *** (0.02)
-Eastern Cape	-0.53 *** (0.02)	-0.29 *** (0.01)	-0.53 *** (0.02)	-0.29 *** (0.01)
-Northern Cape	-0.53 *** (0.05)	-0.30 *** (0.03)	-0.52 *** (0.05)	-0.30 *** (0.03)
-Free State	-0.17 *** (0.03)	-0.09 *** (0.02)	-0.16 *** (0.03)	-0.09 *** (0.02)
-KwaZulu / Natal	-0.09 *** (0.03)	-0.06 *** (0.01)	-0.09 *** (0.03)	-0.06 *** (0.01)
-North West	-0.24 *** (0.03)	-0.14 *** (0.02)	-0.23 *** (0.03)	-0.13 *** (0.02)
-Mpumalanga	0.13 *** (0.03)	0.08 *** (0.02)	0.13 *** (0.03)	0.08 *** (0.02)
-Limpopo	0.29 *** (0.03)	0.17 *** (0.02)	0.29 *** (0.03)	0.17 *** (0.02)
L'année de l'enquête est :				
-1996	-0.08 *** (0.02)	-0.05 *** (0.01)	-0.08 *** (0.02)	-0.05 *** (0.01)
-1997	-0.18 *** (0.02)	-0.13 *** (0.01)	-0.18 *** (0.02)	-0.13 *** (0.01)
-1998	-0.10 *** (0.02)	-0.07 *** (0.01)	-0.09 *** (0.02)	-0.07 *** (0.01)
-1999	-0.08 *** (0.02)	-0.03 *** (0.01)	-0.08 *** (0.02)	-0.03 *** (0.01)
Nombre d'observations	83171	83171	83158	83158
Pseudo R2	0.03	0.03	0.03	0.03

N.B :

- La variable dépendante est le nombre d'années d'éducation des enfants.
- Les valeurs entre parenthèses correspondent aux écart-types des coefficients estimés.
- *, **, *** indiquent que le coefficient correspondant est significatif à 1%, 5% et 10% respectivement.

Tableau 10: Effets marginaux des modèles logit et probit ordonnés (x1000)

	Éducation du chef de famille		Âge de l'enfant		Nombre d'enfants		Capital humain		Éducation d'un chef de famille femme	
Éducation de l'enfant	ologit	oprobit	ologit	oprobit	ologit	oprobit	ologit	oprobit	ologit	oprobit
0 an de scolarité	-8.62***	-10.23***	0.07	0.24***	0.57**	0.66**	-7.5***	-8.96***	-0.97***	-1.15***
1 an de scolarité	-0.48***	-0.47***	0.00	0.01***	0.03**	0.03**	-0.43***	-0.41***	-0.06***	-0.05***
2 ans de scolarité	-2.09***	-1.96***	0.02	0.05***	0.14**	0.13**	-1.83***	-1.71***	-0.24***	-0.22***
3 ans de scolarité	-2.63***	-2.36***	0.02	0.06***	0.17***	0.15**	-2.31***	-2.06***	-0.30***	-0.26***
4 ans de scolarité	-4.67***	-3.94***	0.04	0.09***	0.31**	0.25**	-4.09***	-3.45***	-0.53***	-0.44***
5 ans de scolarité	-4.95***	-3.95***	0.04	0.09***	0.33**	0.26**	-4.33***	-3.46***	-0.56***	-0.44***
6 ans de scolarité	-4.75***	-3.71***	0.04	0.09***	0.31**	0.24**	-4.15***	-3.25***	-0.53***	-0.42***
7 ans de scolarité	-2.92***	-2.54***	0.02	0.06***	0.16**	0.13**	-2.56***	-2.22***	-0.33***	-0.29***
8 ans de scolarité	1.36***	0.39***	-0.01	-0.00***	-0.09**	-0.03**	1.19***	0.35***	0.15***	0.04***
9 ans de scolarité	4.8***	3.28***	-0.04	-0.08***	-0.32**	-0.21**	4.26***	2.87***	0.55***	0.37***
10 ans de scolarité	7***	5.77***	-0.06	-0.14***	-0.46**	-0.37**	6.13***	5.06***	0.79***	0.65***
11 ans de scolarité	7.42***	7.34***	-0.06	-0.17***	-0.49**	-0.47**	6.5***	6.43***	0.84***	0.82***
12 ans de scolarité	9.46***	11.28***	-0.08	-0.27***	-0.63**	-0.73**	8.28***	9.88***	1.07***	1.27***
13 ans de scolarité	0.13***	0.15***	-0.00	-0.00***	-0.00**	-0.01**	0.11***	0.14***	0.01***	0.02***
14 ans de scolarité	0.69***	0.77***	-0.00	-0.02***	-0.05**	-0.05**	0.6***	0.67***	0.08***	0.09***
15 ans de scolarité	0.16***	0.15***	-0.00	-0.00***	-0.01**	-0.01**	0.14***	0.13***	0.02***	0.02***
16 ans de scolarité	0.02***	0.01***	-0.00	-0.00***	-0.00**	-0.00**	0.02***	0.01***	0.00***	0.00***

N.B :

- *, **, *** indiquent que l'effet marginal correspondant est significatif à 1%, 5% et 10%, respectivement.

Tableau 11: Résultats des estimations utilisant les données de pseudo-panel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nombres d'années d'éducation du chef de famille	0,29 *** (0,05)	0,27 *** (0,05)	0,25 *** (0,05)	0,24 *** (0,001)	0,24 *** (0,001)	0,26 *** (0,06)	0,23 *** (0,05)	0,20 *** (0,05)	0,23 *** (0,001)	0,23 *** (0,001)
Âge de l'enfant	-0,11 *** (0,03)	-0,11 *** (0,03)	-0,11 *** (0,03)	-0,03 *** (0,0003)	-0,03 *** (0,0004)	-0,10 *** (0,03)	-0,11 *** (0,03)	-0,11 *** (0,03)	-0,03 *** (0,0003)	-0,03 *** (0,0003)
L'enfant est de sexe féminin			0,43 *** (0,05)	0,44 *** (0,04)	0,44 *** (0,04)	0,44 *** (0,04)	0,44 *** (0,05)	0,44 *** (0,04)	0,44 *** (0,04)	0,44 *** (0,04)
Nombre total d'enfants dans la famille	0,05 (0,09)	0,04 (0,09)	0,03 (0,10)	-0,01 *** (0,001)	-0,01 *** (0,002)	0,09 (0,09)	0,09 (0,09)	0,09 (0,10)	-0,002 * (0,001)	-0,002 * (0,001)
Niveau d'inégalité par district de résidence						-0,79 (0,54)	-0,97 * (0,51)	-1,13 * (0,64)	-0,42 *** (0,01)	-0,42 *** (0,01)
Le chef de famille est de sexe féminin		0,12 (0,10)	0,10 (0,11)	-0,10 ** (0,04)	-0,10 ** (0,04)		0,12 (0,09)	0,12 (0,11)	-0,09 ** (0,04)	-0,09 ** (0,04)
L'enfant vit dans le secteur rural	-0,78 * (0,43)	-1,00 (0,39)	-1,13 *** (0,36)	-0,30 *** (0,001)	-0,30 *** (0,002)	-0,77 * (0,43)	-0,93 ** (0,39)	-1,05 *** (0,36)	-0,30 *** (0,005)	-0,30 *** (0,005)
L'enfant vit dans la province de :										
Western Cape		-0,67 (0,13)	-0,67 *** (0,11)	-0,49 *** (0,08)	-0,49 *** (0,08)		-0,73 *** (0,14)	-0,75 *** (0,11)	-0,52 *** (0,08)	-0,52 *** (0,08)
Eastern Cape		-0,64 ** (0,25)	-0,57 ** (0,27)	-1,01 *** (0,08)	-1,01 *** (0,08)		-0,52 * (0,27)	-0,44 (0,28)	-0,95 *** (0,09)	-0,95 *** (0,09)
Northern Cape		-1,17 *** (0,14)	-1,18 *** (0,11)	-1,21 *** (0,11)	-1,21 *** (0,11)		-0,96 *** (0,18)	-0,94 *** (0,17)	-1,11 *** (0,11)	-1,11 *** (0,11)
Free State		-0,37 ** (0,16)	-0,35 ** (0,15)	-0,48 *** (0,06)	-0,48 *** (0,06)		-0,22 (0,18)	-0,20 (0,19)	-0,41 *** (0,06)	-0,41 *** (0,06)
KwaZulu / Natal		-0,13 (0,27)	-0,05 (0,29)	-0,47 *** (0,05)	-0,47 *** (0,05)		0,05 (0,28)	0,13 (0,32)	-0,37 *** (0,05)	-0,37 *** (0,05)
North West		-0,33 (0,26)	-0,26 (0,25)	-0,73 *** (0,05)	-0,73 *** (0,05)		-0,14 (0,27)	-0,07 (0,29)	-0,64 *** (0,05)	-0,64 *** (0,05)
Mpumalanga		0,12 (0,26)	0,18 (0,30)	-0,18 *** (0,06)	-0,18 *** (0,06)		0,52 (0,33)	0,62 (0,41)	0,01 *** (0,07)	0,01 *** (0,07)
Limpopo		0,46 (0,31)	0,55 (0,36)	0,11 * (0,06)	0,11 * (0,06)		0,87 ** (0,38)	0,99 ** (0,46)	0,29 * (0,07)	0,29 * (0,07)
Constante	9,94 *** (0,98)	10,17 ** (0,81)	10,29 *** (0,75)	8,52 0,05	8,52 0,05	10,52 *** (1,06)	10,80 *** (0,87)	11,08 *** (0,83)	8,86 (0,05)	8,86 (0,05)
Nombre d'observations	180	180	180	180	180	180	180	180	180	180
R2 (Pseudo R2)	0,32	0,84	0,84			0,29	0,85	0,85		

N.B :

- La variable dépendante est le nombre moyen d'années d'éducation des enfants par cellule.
- Colonne (1) = Estimations avec effets fixes.
- Colonne (2) = Estimations avec effets aléatoires.
- Colonne (3) = Estimations sans effets spécifiques.
- Colonne (4) = Estimateur de Deaton.
- Colonne (5) = Estimateur de Verbeek et Nijman.
- Colonne (6) = Estimations avec effets fixes en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Colonne (7) = Estimations avec effets aléatoires en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Colonne (8) = Estimations sans effets spécifiques en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Colonne (9) = Estimateur de Deaton en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Colonne (10) = Estimateur de Verbeek et Nijman en ajoutant une mesure d'inégalité par district dans les variables explicatives du modèle.
- Les valeurs entre parenthèses indiquent les écart-types robustes des estimateurs pour les colonnes (1), (2) et (3) ou bien les écart-types estimés par *bootstrapping*, pour les colonnes (4) et (5).
- *, **, *** indiquent que le coefficient correspondant est significatif à 1%, 5% et 10% respectivement.

7. Références bibliographiques

- Atkinson, Anthony B. (1981), « On Intergenerational Income Mobility in Britain », *Journal of Post Keynesian Economics*, 3 : 194-218.
- _____, Alan K. Maynard et C.G. Trinder (1983), *Parents and Children : Incomes in Two Generations*, London : Heinemann Educational Books.
- Baltagi, Badi H. (2001), « *Econometric Analysis of Panel Data* », 2nd Edition, John Wiley & Sons, Ltd.
- Becker, Gary S. (1979), « An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility », *Journal of Political Economy*, 87(6) : 1153-89.
- Becker, Gary S. et Nigel Tomes (1986), « Human Capital and the Rise and the Fall of Families », *Journal of Labor Economics*, 4 : S1-S39.
- Behrman, Jere R. et Paul Taubman (1985), « Intergenerational Earnings Mobility in The United States : Some Estimates and a Test of Becker's Intergenerational Endowments Model », *The Review of Economics and Statistics*, 67 : 144-51.
- _____(1989), « Is Schooling Mostly in the Genes ? Nature-Nurture Decomposition Using Data Relatives », *Journal of Political Economy*, 97(6) : 1425-446.
- _____(1990), « The Intergenerational Correlation Between Children's Adult Earnings and Their Parents' Income : Results from the Michigan Panel Survey of Income Dynamics », *Review of Income and Wealth*, 36(2) : 115-27.
- Belzil, Christian et Jörgen Hansen (2002), « Unobserved Ability and Return to Schooling », *Econometrica*, 70(5) : 2075-91.

Björklund, Anders et Markus Jäntti (1997), « Intergenerational Income Mobility in Sweden Compared to the United States », *American Economic Review*, 87(5) : 1009-18.

Binder, Melissa et Christopher Woodruff (2002), « Inequality and Intergenerational Mobility in Schooling : The case of Mexico », *Economic Development and Cultural Change*, 50(2) : 249-68.

Birdsall, Nancy (1985), « Public Inputs and Child Schooling in Brazil », *Journal of Development Economics*, 18(1), 67-86.

Borjas, George J. (1995), « Ethnicity, Neighborhoods, and Human-Capital Externalities », *American Economic Review*, 85(3), p.365-90.

Bowles, Samuel et Herbert Gintis (2000), *The Inheritance of Economic Status : Education, Class and Genetics*, Amherst : University of Massachussets.

Bowles, Samuel et Herbert Gintis (2001), *Intergenerational Inequality*, Amherst : University of Massachusetts.

Browning, Martin; Deaton, Angus; et Margaret Irish (1985), « A Profitable Approach to Labor Supply and Commodity Demands over the Life Cycle », *Econometrica*, 53 : 503-43.

Burns, Justine (2000), *Inheriting the Future from the Past : Intergenerational Persistence in Educational Status in South Africa*, University of Massachusetts, Amherst.

Burns, Justine (2001), *Inheriting the Future : The Role of Family Background and Neighbourhood Characteristics in Child Schooling Outcomes in South Africa*, University of Massachusetts, Amherst.

Checchi, Daniel (2000), « Does Education Achievement Help to Explain Income Inequality ? », Working Paper n°208, The United Nations University World Institute for Development Economics Research.

_____, Ichino Andrea et Aldo Rustichini (1999), « More Equal but Less Mobile? Education Financing and Intergenerational Mobility in Italy and in the U.S. », *Journal of Public Economics*, 74 : 351-93.

Chiswick, Barry R. (1988), « Differences in Education and Earnings across Racial and Ethnic Groups : Tastes, Discrimination, and Investments in Child Quality », *Quarterly Journal of Economics*, 103(3) : 571-97.

Corak, Miles et Andrew Heisz (1999), « The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data », *Journal of Human Resources*, 34(3) : 504-33.

Corcoran, Mary, Gordon Roger, Laren Deborah et Gary Solon (1992), « The Association between Men's Economic Status and their Family and Community Origins », *Journal of Human Resources*, 27(4) : 575-601.

Cooper, Suzanne (1998), « A Positive Theory of Income Redistribution », *Journal of Economic Growth*, 3(2), 171-95.

_____, Steven Durlauf et Paul Johnson (1993), « On the Evolution of Economic Status across Generations », *American Statistical Association : Proceedings of the Business and Economic Statistics Section*, 50-8.

Couch, Kenneth A. et Thomas A. Dunn (1997), « Intergenerational Correlations in Labor Market Status : a Comparison of the United States and Germany », *Journal of Human Resources*, 32 (1) : 210-32.

Darity, William, Jr. et Jessica Gordon Nembhard (2000), « Cross-National Comparisons of Racial and Ethnic Economic Inequality », *American Economic Review(Papers and Proceedings)*, 90(2) : 308-11.

Datcher, Linda (1982), « Effects of Community and Family Background on Achievement », *The Review of Economics and Statistics*, 64 : 32-41.

Deaton, Angus (1985), « Panel Data from Times Series of Cross Sections », *Journal of Econometrics*, 30 : 109-26.

Dearden, Lorraine, Machin Stephen et Howard Reed (1997), « Intergenerational Mobility in Britain », *The Economic Journal*, 107 : 47-66.

Dessy, Sylvain E. et Désiré Vencatachellum (2003), « Explaining Cross-Country Differences in Policy Response to Child Labour », *Canadian Journal of Economics*, 36(1), à paraître.

Dickerson, Andy, Green Francis et Jorge S. Arbache (2001), *Trade Liberalization and the Returns to Education : A Pseudo-Panel Approach*, University of Kent.

Durlauf, Steven N. (1996), « A Theory of Persistent Income Inequality », *Journal of Economic Growth*, 1 : 75-96.

_____ (1997), *The Memberships Theory of Inequality : Ideas and Implications*, Madison : University of Wisconsin.

Fisher, Ronald A. (1918), « The Correlation Between Relatives on the Supposition of Mendelian Inheritance », *Trans. Royal Soc. Edinburgh*, 52(2), 399-433.

Freedman, David A. (1997), « Bootstrapping Regression Models », *Annals of Statistics*, 9, 1218-28.

Galor, Oded et Joseph Zeira (1993), « Income Distribution and Macroeconomics », *Review of Economics Studies*, 60(1) : 35-52.

Gang, Ira N. et Klaus F. Zimmerman (2000), « Is Child Like Parent ? Educational Attainment and Ethnic Origin », *Journal of Human Resources*, 34(3) : 551-69.

Gaviria, Alejandro (2002), « Intergenerational Mobility, Sibling Inequality and Borrowing Constraints », *Economics of Education Review*, 21 : 331-40.

Goldberger, Arthur S. (1989), « Economic and Mechanical Models of Intergenerational Transmission », *American Economic Review*, 79(3) : 504-13.

Gustaffsson, Björn (1989), « The Degree and Pattern of Income Immobility in Sweden », *The Review of Income and Wealth*, 40(1) : 67-86.

Haveman, Robert et Barbara Wolfe (1995), « The Determinants of Children's Attainments : A Review of Methods and Findings », *Journal of Economic Literature*, 33(4) : 1829-78.

Hsiao, Cheng (1985), « Benefits and Limitations of Panel Data », *Econometric Review*, 4 : 121-74.

Lucas, Robert (1988), « On the Mechanics of Economic Development », *Journal of Monetary Economics*, 22, 3-42.

MacKinnon, James G. (2002), « Bootstrap Inference in Econometrics », *Canadian Journal of Economics*, 35(4) : 615-45.

Michaud, Pierre-Carl et Désiré Vencatachellum (2003), « Human Capital Externalities in South Africa », *Economic Development and Cultural Change*, 51(3), à paraître.

Mincer, Jacob (1974), « *Schooling, Experience and Earnings* », Columbia University Press.

Moffitt, Robert (1993), « Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections », *Journal of Econometrics*, 59 : 99-123.

Moll, Peter G. (1993), « Industry Wage Differentials and Efficiency Wages. A Dissenting View with South Africa Evidence », *Journal of Development Economics*, 41 : 213-46.

Mulligan Casey B. (1997), *Parental Priorities and Economic Inequality*, Chicago : University of Chicago Press.

_____ (1999), « Galton versus the Human Capital Approach to Inheritance », *Journal of Political Economy*, 107(6) : S184-S224.

Mwabu, Germano et T.Paul Schultz (1996), « Education Returns Across Quantiles of the Wage Function : Alternative Explanations for Returns to Education in South Africa », *American Economic Review*, 86(2) : 335-39.

Peters, Elizabeth H. (1992), « Patterns of Intergenerational Mobility in Income and Earnings », *The Review of Economics and Statistics*, 74(3) : 456-66.

Ray, Debraj (1998), *Development Economics*, Princeton University Press.

Rosenzweig, Mark R. et T. Paul Schultz (1982), « Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution : Child Survival in Rural India », *American Economic Review*, 72(4), 808-15.

SADA (South African Data Archive), *October Household Survey, 1995, 1996, 1997, 1998*, Statistics South Africa (Stats SA).

Schultz, T. Paul (1984), « The Schooling and Health of Children of U.S. Immigrants and Natives », *Research in Population Economics*, Greenwich, Conn. : JAI Press, 251-88.

_____ (1993), « Returns to Women's Education », *Women's Education in Developing Countries : Barriers, Benefits and Policies*, Baltimore et London, Johns Hopkins University Press for the World Bank, 5 1-99.

Sewell, William H. et Robert M. Hauser (1975), *Education, Occupation, and Earnings : Achievement in the Early Career*, New York : Academic Press.

Sherer, George (2000), « Intergroup Economic Inequality in South Africa : The Post-Apartheid Era », *American Economic Review*, 90(2) : 317-21.

Solon, Gary R. (1989), « Biases in The Estimation of Intergenerational Earnings Correlations », *The Review of Economics and Statistics*, 71(1) : 172-74.

_____ (1992), « Intergenerational Income Mobility in The United States », *American Economic Review*, 82(3) : 393-408.

Soltow, Lee (1965), *Toward Income Equality in Norway*, Madison : University of Wisconsin Press.

Tansel, Aysit (2002), « Determinants of School Attainment of Boys and Girls in Turkey : Individual, Household and Community factors », *Economics of Education Review*, 21 : 455-70.

Thomas, Duncan (1996), « Education across Generations in South Africa », *American Economic Review*, 86(2) : 330-34.

Verbeek, Mano et Theo Nijman (1990), « Estimation of Time Dependent Parameters in Linear Models Using Cross Sections, Panels or Both », *Journal of Econometrics*, 46 : 333-46.

_____ (1992), « Can Cohort Data be Treated as Genuine Panel Data ? », *Empirical Economics*, 17 : 9-23.

_____ (1993), « Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-Sections », *Journal of Econometrics*, 59 : 125-36.

Zimmerman, David J. (1992), « Regression Toward Mediocrity in Economic Stature », *American Economic Review*, 82(3) : 409-29